Measure theory and probability

Alexander Grigoryan
University of Bielefeld

Lecture Notes, October 2007 - February 2008

Contents

1 Construction of measures

1.1 Introduction and examples

The main subject of this lecture course and the notion of *measure* (Maß). The rigorous definition of measure will be given later, but now we can recall the familiar from the elementary mathematics notions, which are all particular cases of measure:

1. *Length* of intervals in $\mathbb{R}$: if I is a bounded interval with the endpoints a, b (that is, I is one of the intervals (a,b), $[a,b]$, $[a,b)$, $(a,b]$) then its length is defined by

$$\ell(I) = |b-a|\,.$$

The useful property of the length is the *additivity*: if an interval I is a disjoint union of a finite family $\{I_k\}_{k=1}^n$ of intervals, that is, $I = \bigsqcup_k I_k$, then

$$\ell(I) = \sum_{k=1}^n \ell(I_k)\,.$$

Indeed, let $\{a_i\}_{i=0}^N$ be the set of all distinct endpoints of the intervals $I, I_1, ..., I_n$ enumerated in the increasing order. Then I has the endpoints a_0, a_N while each interval I_k has necessarily the endpoints a_i, a_{i+1} for some i (indeed, if the endpoints of I_k are a_i and a_j with $j > i+1$ then the point a_{i+1} is an interior point of I_k, which means that I_k must intersect with some other interval I_m). Conversely, any couple a_i, a_{i+1} of consecutive points are the end points of some interval I_k (indeed, the interval (a_i, a_{i+1}) must be covered by some interval I_k; since the endpoints of I_k are consecutive numbers in the sequence $\{a_j\}$, it follows that they are a_i and a_{i+1}). We conclude that

$$\ell(I) = a_N - a_0 = \sum_{i=0}^{N-1}(a_{i+1}-a_i) = \sum_{k=1}^n \ell(I_k)\,.$$

2. *Area* of domains in $\mathbb{R}^2$. The full notion of area will be constructed within the general measure theory later in this course. However, for rectangular domains the area is defined easily. A *rectangle* A in $\mathbb{R}^2$ is defined as the direct product of two intervals I, J from $\mathbb{R}$:

$$A = I \times J = \left\{(x,y) \in \mathbb{R}^2 : x \in I, y \in J\right\}.$$

Then set

$$\operatorname{area}(A) = \ell(I)\,\ell(J)\,.$$

We claim that the area is also additive: if a rectangle A is a disjoint union of a finite family of rectangles $A_1, ..., A_n$, that is, $A = \bigsqcup_k A_k$, then

$$\operatorname{area}(A) = \sum_{k=1}^n \operatorname{area}(A_k)\,.$$

For simplicity, let us restrict the consideration to the case when all sides of all rectangles are semi-open intervals of the form $[a,b)$. Consider first a particular case, when the rectangles $A_1, ..., A_k$ form a *regular tiling* of A; that is, let $A = I \times J$ where $I = \bigsqcup_i I_i$ and $J = \bigsqcup_j J_j$, and assume that all rectangles A_k have the form $I_i \times J_j$. Then

$$\operatorname{area}(A) = \ell(I)\,\ell(J) = \sum_i \ell(I_i) \sum_j \ell(J_j) = \sum_{i,j} \ell(I_i)\,\ell(J_j) = \sum_k \operatorname{area}(A_k).$$

Now consider the general case when A is an arbitrary disjoint union of rectangles A_k. Let $\{x_i\}$ be the set of all X-coordinates of the endpoints of the rectangles A_k put in the increasing order, and $\{y_j\}$ be similarly the set of all the Y-coordinates, also in the increasing order. Consider the rectangles

$$B_{ij} = [x_i, x_{i+1}) \times [y_j, y_{j+1}).$$

Then the family $\{B_{ij}\}_{i,j}$ forms a regular tiling of A and, by the first case,

$$\operatorname{area}(A) = \sum_{i,j} \operatorname{area}(B_{ij}).$$

On the other hand, each A_k is a disjoint union of some of B_{ij}, and, moreover, those B_{ij} that are subsets of A_k, form a regular tiling of A_k, which implies that

$$\operatorname{area}(A_k) = \sum_{B_{ij} \subset A_k} \operatorname{area}(B_{ij}).$$

Combining the previous two lines and using the fact that each B_{ij} is a subset of exactly one set A_k, we obtain

$$\sum_k \operatorname{area}(A_k) = \sum_k \sum_{B_{ij} \subset A_k} \operatorname{area}(B_{ij}) = \sum_{i,j} \operatorname{area}(B_{ij}) = \operatorname{area}(A).$$

3. *Volume* of domains in $\mathbb{R}^3$. The construction is similar to the area. Consider all *boxes* in $\mathbb{R}^3$, that is, the domains of the form $A = I \times J \times K$ where I, J, K are intervals in $\mathbb{R}$, and set

$$\operatorname{vol}(A) = \ell(I)\,\ell(J)\,\ell(K).$$

Then volume is also an additive functional, which is proved in a similar way. Later on, we will give the detailed proof of a similar statement in an abstract setting.

4. *Probability* is another example of an additive functional. In probability theory, one considers a set Ω of elementary events, and certain subsets of Ω are called *events* (Ereignisse). For each event $A \subset \Omega$, one assigns the probability, which is denoted by $\mathbb{P}(A)$ and which is a real number in $[0, 1]$. A reasonably defined probability must satisfy the additivity: if the event A is a disjoint union of a finite sequence of evens $A_1, ..., A_n$ then

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$

The fact that A_i and A_j are disjoint, when $i \neq j$, means that the events A_i and A_j cannot occur at the same time.

The common feature of all the above example is the following. We are given a non-empty set M (which in the above example was $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, Ω), a family S of its subsets (the families of intervals, rectangles, boxes, events), and a functional $\mu : S \to \mathbb{R}_+ := [0, +\infty)$ (length, area, volume, probability) with the following property: if $A \in S$ is a disjoint union of a finite family $\{A_k\}_{k=1}^n$ of sets from S then

$$\mu(A) = \sum_{k=1}^{n} \mu(A_k).$$

A functional μ with this property is called a *finitely additive measure.* Hence, length, area, volume, probability are all finitely additive measures.

1.2 σ-additive measures

As above, let M be an arbitrary non-empty set and S be a family of subsets of M.

Definition. A functional $\mu : S \to \mathbb{R}_+$ is called a *σ-additive measure* if whenever a set $A \in S$ is a disjoint union of an at most countable sequence $\{A_k\}_{k=1}^N$ (where N is either finite or $N = \infty$) then

$$\mu(A) = \sum_{k=1}^{N} \mu(A_k).$$

If $N = \infty$ then the above sum is understood as a series. If this property holds only for finite values of N then μ is a *finitely additive measure.*

Clearly, a σ-additive measure is also finitely additive (but the converse is not true). At first, the difference between finitely additive and σ-additive measures might look insignificant, but the σ-additivity provides much more possibilities for applications and is one of the central issues of the measure theory. On the other hand, it is normally more difficult to prove σ-additivity.

Theorem 1.1 *The length is a σ-additive measure on the family of all bounded intervals in $\mathbb{R}$.*

Before we prove this theorem, consider a simpler property.

Definition. A functional $\mu : S \to \mathbb{R}_+$ is called *σ-subadditive* if whenever $A \subset \bigcup_{k=1}^N A_k$ where A and all A_k are elements of S and N is either finite or infinite, then

$$\mu(A) \leq \sum_{k=1}^{N} \mu(A_k).$$

If this property holds only for finite values of N then μ is called *finitely subadditive.*

Lemma 1.2 *The length is σ-subadditive.*

Proof. Let I, $\{I_k\}_{k=1}^\infty$ be intervals such that $I \subset \bigcup_{k=1}^\infty I_k$ and let us prove that

$$\ell(I) \leq \sum_{k=1}^{\infty} \ell(I_k)$$

(the case $N<\infty$ follows from the case $N=\infty$ by adding the empty interval). Let us fix some $\varepsilon>0$ and choose a bounded *closed* interval $I'\subset I$ such that

$$\ell(I)\leq\ell(I')+\varepsilon.$$

For any k, choose an *open* interval $I'_k\supset I_k$ such that

$$\ell(I'_k)\leq\ell(I_k)+\frac{\varepsilon}{2^k}.$$

Then the bounded closed interval I' is covered by a sequence $\{I'_k\}$ of open intervals. By the Borel-Lebesgue lemma, there is a finite subfamily $\left\{I'_{k_j}\right\}_{j=1}^n$ that also covers I'. It follows from the finite additivity of length that it is finitely subadditive, that is,

$$\ell(I')\leq\sum_j\ell(I'_{k_j}),$$

(see Exercise 7), which implies that

$$\ell(I')\leq\sum_{k=1}^{\infty}\ell(I'_k).$$

This yields

$$l(I)\leq\varepsilon+\sum_{k=1}^{\infty}\left(\ell(I_k)+\frac{\varepsilon}{2^k}\right)=2\varepsilon+\sum_{k=1}^{\infty}\ell(I_k).$$

Since $\varepsilon>0$ is arbitrary, letting $\varepsilon\to0$ we finish the proof. ■

Proof of Theorem 1.1. We need to prove that if $I=\bigsqcup_{k=1}^{\infty}I_k$ then

$$\ell(I)=\sum_{k=1}^{\infty}\ell(I_k)$$

By Lemma 1.2, we have immediately

$$\ell(I)\leq\sum_{k=1}^{\infty}\ell(I_k)$$

so we are left to prove the opposite inequality. For any fixed $n\in\mathbb{N}$, we have

$$I\supset\bigsqcup_{k=1}^{n}I_k.$$

It follows from the finite additivity of length that

$$\ell(I)\geq\sum_{k=1}^{n}\ell(I_k)$$

(see Exercise 7). Letting $n\to\infty$, we obtain

$$\ell(I)\geq\sum_{k=1}^{\infty}\ell(I_k),$$

which finishes the proof. ■

1.3 An example of using probability theory

Probability theory deals with random events and their probabilities. A classical example of a random event is a *coin tossing.* The outcome of each tossing may be *heads* or *tails*: H or T. If the coin is fair then after N trials, H occurs approximately $N/2$ times, and so does T. It is natural to *believe* that if $N \to \infty$ then $\frac{\#H}{N} \to \frac{1}{2}$ so that one says that H occurs with probability $1/2$ and writes $\mathbb{P}(H) = 1/2$. In the same way $\mathbb{P}(T) = 1/2$. If a coin is biased (=not fair) then it could be the case that $\mathbb{P}(H)$ is different from $1/2$, say,

$$\mathbb{P}(H) = p \quad \text{and} \quad \mathbb{P}(T) = q := 1 - p. \tag{1.1}$$

We present here an amusing argument how random events H and T satisfying (1.1) can be used to prove the following purely algebraic inequality:

$$(1 - p^n)^m + (1 - q^m)^n \geq 1, \tag{1.2}$$

where $0 < p, q < 1$, $p + q = 1$, and n, m are positive integers. This inequality has also an algebraic proof which however is more complicated and less transparent than the probabilistic argument below.

Let us make *nm independent* tossing of the coin and write the outcomes in a $n \times m$ table putting in each cell H or T, for example, as below:

$$n\left\{\underbrace{\begin{array}{|c|c|c|c|c|}\hline H & T & T & H & T \\ \hline T & T & H & H & H \\ \hline H & H & T & H & T \\ \hline T & H & T & T & T \\ \hline\end{array}}_{m}\right.$$

Then, using the independence of the events, we obtain:

$$\begin{aligned} p^n &= \mathbb{P}\{\text{a given column contains only } H\} \\ 1 - p^n &= \mathbb{P}\{\text{ a given column contains at least one } T\} \end{aligned}$$

whence

$$(1 - p^n)^m = \mathbb{P}\{\text{any column contains at least one } T\}. \tag{1.3}$$

Similarly,

$$(1 - q^m)^n = \mathbb{P}\{\text{any row contains at least one } H\}. \tag{1.4}$$

Let us show that one of the events (1.3) and (1.4) will always take place which would imply that the sum of their probabilities is at least 1, and prove (1.2). Indeed, assume that the event (1.3) does not take place, that is, some column contains only H, say, as below:

$$\begin{array}{|c|c|c|c|c|}\hline & & H & & \\ \hline & & H & & \\ \hline & & H & & \\ \hline & & H & & \\ \hline\end{array}$$

Then one easily sees that H occurs in *any row* so that the event (1.4) takes place, which was to be proved.

Is this proof rigorous? It may leave impression of a rather empirical argument than a mathematical proof. The point is that we have used in this argument the *existence* of events with certain properties: firstly, H should have probability p where p a given number in $(0,1)$ and secondly, there must be enough independent events like that. Certainly, mathematics cannot rely on the existence of biased coins (or even fair coins!) so in order to make the above argument rigorous, one should have a mathematical notion of events and their probabilities, and prove the existence of the events with the required properties. This can and will be done using the measure theory.

1.4 Extension of measure from semi-ring to a ring

Let M be any non-empty set.

Definition. A family S of subsets of M is called a *semi-ring* if

- S contains $\emptyset$
- $A, B \in S \implies A \cap B \in S$
- $A, B \in S \implies A \setminus B$ is a disjoint union of a finite family sets from S.

Example. The family of all intervals in $\mathbb{R}$ is a semi-ring. Indeed, the intersection of two intervals is an interval, and the difference of two intervals is either an interval or the disjoint union of two intervals. In the same way, the family of all intervals of the form $[a, b)$ is a semi-ring.

The family of all rectangles in $\mathbb{R}^2$ is also a semi-ring (see Exercise 6).

Definition. A family S if subsets of M is called a *ring* if

- S contains $\emptyset$
- $A, B \in S \implies A \cup B \in S$ and $A \setminus B \in S$.

It follows that also the intersection $A \cap B$ belongs to S because

$$A \cap B = B \setminus (B \setminus A)$$

is also in S. Hence, a ring is closed under taking the set-theoretic operations $\cap, \cup, \setminus$. Also, it follows that a ring is a semi-ring.

Definition. A ring S is called a *σ-ring* if the union of any countable family $\{A_k\}_{k=1}^{\infty}$ of sets from S is also in S.

It follows that the intersection $A = \bigcap_k A_k$ is also in S. Indeed, let B be any of the sets A_k so that $B \supset A$. Then

$$A = B \setminus (B \setminus A) = B \setminus \left(\bigcup_k (B \setminus A_k) \right) \in S.$$

Trivial examples of rings are $S = \{\emptyset\}$, $S = \{\emptyset, M\}$, or $S = 2^M$ – the family of all subsets of M. On the other hand, the set of the intervals in $\mathbb{R}$ is not a ring.

Observe that if $\{S_\alpha\}$ is a family of rings (σ-rings) then the intersection $\bigcap_\alpha S_\alpha$is also a ring (resp., σ-ring), which trivially follows from the definition.

Definition. Given a family S of subsets of M, denote by $R(S)$ the intersection of all rings containing S.

Note that at least one ring containing S always exists: 2^M. The ring $R(S)$ is hence the minimal ring containing S.

Theorem 1.3 *Let S be a semi-ring.*

(a) *The minimal ring $R(S)$ consists of all finite disjoint unions of sets from S.*

(b) *If μ is a finitely additive measure on S then μ extends uniquely to a finitely additive measure on $R(S)$.*

(c) *If μ is a σ-additive measure on S then the extension of μ to $R(S)$ is also σ-additive.*

For example, if S is the semi-ring of all intervals on $\mathbb{R}$ then the minimal ring $R(S)$ consists of finite disjoint unions of intervals. The notion of the length extends then to all such sets and becomes a measure there. Clearly, if a set $A \in R(S)$ is a disjoint union of the intervals $\{I_k\}_{k=1}^n$ then $\ell(A) = \sum_{k=1}^n \ell(I_k)$. This formula follows from the fact that the extension of ℓ to $R(S)$ is a measure. On the other hand, this formula can be used to explicitly define the extension of ℓ.

Proof. (a) Let S' be the family of sets that consists of all finite disjoint unions of sets from S, that is,

$$S' = \left\{ \bigsqcup_{k=1}^n A_k : A_k \in S,\ n \in \mathbb{N} \right\}.$$

We need to show that $S' = R(S)$. It suffices to prove that S' is a ring. Indeed, if we know that already then $S' \supset R(S)$ because the ring S' contains S and $R(S)$ is the minimal ring containing S. On the other hand, $S' \subset R(S)$, because $R(S)$ being a ring contains all finite unions of elements of S and, in particular, all elements of S'.

The proof of the fact that S' is a ring will be split into steps. If A and B are elements of S then we can write $A = \bigsqcup_{k=1}^n A_k$ and $B = \bigsqcup_{l=1}^m B_l$, where $A_k, B_l \in S$.

Step 1. If $A, B \in S'$ and A, B are disjoint then $A \sqcup B \in S'$. Indeed, $A \sqcup B$ is the disjoint union of all the sets A_k and B_l so that $A \sqcup B \in S'$.

Step 2. If $A, B \in S'$ then $A \cap B \in S'$. Indeed, we have

$$A \cap B = \bigsqcup_{k,l} (A_k \cap B_l).$$

Since $A_k \cap B_l \in S$ by the definition of a semi-ring, we conclude that $A \cap B \in S'$.

Step 3. If $A, B \in S'$ then $A \setminus B \in S'$. Indeed, since

$$A \setminus B = \bigsqcup_k (A_k \setminus B),$$

it suffices to show that $A_k \setminus B \in S'$ (and then use Step 1). Next, we have

$$A_k \setminus B = A_k \setminus \bigsqcup_l B_l = \bigcap_l (A_k \setminus B_l).$$

By Step 2, it suffices to prove that $A_k \setminus B_l \in S'$. Indeed, since $A_k, B_l \in S$, by the definition of a semi-ring we conclude that $A_k \setminus B_l$ is a finite disjoint union of elements of S, whence $A_k \setminus B_l \in S'$.

Step 4. If $A, B \in S'$ then $A \cup B \in S'$. Indeed, we have

$$A \cup B = (A \setminus B) \bigsqcup (B \setminus A) \bigsqcup (A \cap B) .$$

By Steps 2 and 3, the sets $A \setminus B$, $B \setminus A$, $A \cap B$ are in S', and by Step 1 we conclude that their disjoint union is also in S'.

By Steps 3 and 4, we conclude that S' is a ring, which finishes the proof.

(b) Now let μ be a finitely additive measure on S. The extension to $S' = R(S)$ is unique: if $A \in R(S)$ and $A = \bigsqcup_{k=1}^{n} A_k$ where $A_k \in S$ then necessarily

$$\mu(A) = \sum_{k=1}^{n} \mu(A_k) . \tag{1.5}$$

Now let us prove the existence of μ on S'. For any set $A \in S'$ as above, let us define μ by (1.5) and prove that μ is indeed finitely additive. First of all, let us show that $\mu(A)$ does not depend on the choice of the splitting of $A = \bigsqcup_{k=1}^{n} A_k$. Let us have another splitting $A = \bigsqcup_l B_l$ where $B_l \in S$. Then

$$A_k = \bigsqcup_l A_k \cap B_l$$

and since $A_k \cap B_l \in S$ and μ is finitely additive on S, we obtain

$$\mu(A_k) = \sum_l \mu(A_k \cap B_l) .$$

Summing up on k, we obtain

$$\sum_k \mu(A_k) = \sum_k \sum_l \mu(A_k \cap B_l) .$$

Similarly, we have

$$\sum_l \mu(B_l) = \sum_l \sum_k (A_k \cap B_l) ,$$

whence $\sum_k \mu(A_k) = \sum_k \mu(B_k)$.

Finally, let us prove the finite additivity of the measure (1.5). Let A, B be two disjoint sets from $R(S)$ and assume that $A = \bigsqcup_k A_k$ and $B = \bigsqcup_l B_l$, where $A_k, B_l \in S$. Then $A \sqcup B$ is the disjoint union of all the sets A_k and B_l whence by the previous argument

$$\mu(A \sqcup B) = \sum_k \mu(A_k) + \sum_l \mu(B_l) = \mu(A) + \mu(B) .$$

If there is a finite family of disjoint sets $C_1, ..., C_n \in R(S)$ then using the fact that the unions of sets from $R(S)$ is also in $R(S)$, we obtain by induction in n that

$$\mu\left(\bigsqcup_k C_k\right) = \sum_k \mu(C_k) .$$

(c) Let $A = \bigsqcup_{l=1}^{\infty} B_l$ where $A, B_l \in R(S)$. We need to prove that

$$\mu(A) = \sum_{l=1}^{\infty} \mu(B_l) . \tag{1.6}$$

Represent the given sets in the form $A = \bigsqcup_k A_k$ and $B_l = \bigsqcup_m B_{lm}$ where the summations in k and m are finite and the sets A_k and B_{lm} belong to S. Set also

$$C_{klm} = A_k \cap B_{lm}$$

and observe that $C_{klm} \in S$. Also, we have

$$A_k = A_k \cap A = A_k \cap \bigsqcup_{l,m} B_{lm} = \bigsqcup_{l,m} (A_k \cap B_{lm}) = \bigsqcup_{l,m} C_{klm}$$

and

$$B_{lm} = B_{lm} \cap A = B_{lm} \cap \bigsqcup_k A_k = \bigsqcup_k (A_k \cap B_{lm}) = \bigsqcup_k C_{klm}.$$

By the σ-additivity of μ on S, we obtain

$$\mu(A_k) = \sum_{l,m} \mu(C_{klm})$$

and

$$\mu(B_{lm}) = \sum_k \mu(C_{klm}).$$

It follows that

$$\sum_k \mu(A_k) = \sum_{k,l.m} \mu(C_{klm}) = \sum_{l,m} \mu(B_{lm}).$$

On the other hand, we have by definition of μ on $R(S)$ that

$$\mu(A) = \sum_k \mu(A_k)$$

and

$$\mu(B_l) = \sum_m \mu(B_{lm})$$

whence

$$\sum_l \mu(B_l) = \sum_{l,m} \mu(B_{lm}).$$

Combining the above lines, we obtain

$$\mu(A) = \sum_k \mu(A_k) = \sum_{l,m} \mu(B_{lm}) = \sum_l \mu(B_l),$$

which finishes the proof. ∎

1.5 Extension of measure to a σ-algebra

1.5.1 σ-rings and σ-algebras

Recall that a σ-ring is a ring that is closed under taking countable unions (and intersections). For a σ-additive measure, a σ-ring would be a natural domain. **Our main aim in this section is to extend a σ-additive measure μ from a ring R to a σ-ring.**

So far we have only trivial examples of σ-rings: $S = \{\emptyset\}$ and $S = 2^M$. Implicit examples of σ-rings can be obtained using the following approach.

Definition. For any family S of subsets of M, denote by $\Sigma(S)$ the intersection of all σ-rings containing S.

At least one σ-ring containing S exists always: 2^M. Clearly, the intersection of any family of σ-rings is again a σ-ring. Hence, $\Sigma(S)$ is the minimal σ-ring containing S.

Most of examples of rings that we have seen so far were not σ-rings. For example, the ring R that consists of all finite disjoint unions of intervals is not σ-ring because it does not contain the countable union of intervals

$$\bigcup_{k=1}^{\infty} (k, k+1/2).$$

In general, it would be good to have an explicit description of $\Sigma(R)$, similarly to the description of $R(S)$ in Theorem 1.3. One might conjecture that $\Sigma(R)$ consists of disjoint countable unions of sets from R. However, this is not true. Let R be again the ring of finite disjoint unions of intervals. Consider the following construction of the *Cantor set* C: it is obtained from $[0,1]$ by removing the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$ then removing the middle third of the remaining intervals, etc:

$$C = [0,1] \setminus \left(\frac{1}{3}, \frac{2}{3}\right) \setminus \left(\frac{1}{9}, \frac{2}{9}\right) \setminus \left(\frac{7}{9}, \frac{8}{9}\right) \setminus \left(\frac{1}{27}, \frac{2}{27}\right) \setminus \left(\frac{7}{27}, \frac{8}{27}\right) \setminus \left(\frac{19}{27}, \frac{20}{27}\right) \setminus \left(\frac{25}{27}, \frac{26}{27}\right) \setminus \ldots$$

Since C is obtained from intervals by using countable union of interval and $\setminus$, we have $C \in \Sigma(R)$. However, the Cantor set is uncountable and contains no intervals expect for single points (see Exercise 13). Hence, C cannot be represented as a countable union of intervals.

The constructive procedure of obtaining the σ-ring $\Sigma(R)$ from a ring R is as follows. Denote by R_σ the family of all countable unions of elements from R. Clearly, $R \subset R_\sigma \subset \Sigma(R)$. Then denote by $R_{\sigma\delta}$ the family of all countable intersections from R_σ so that

$$R \subset R_\sigma \subset R_{\sigma\delta} \subset \Sigma(R).$$

Define similarly $R_{\sigma\delta\sigma}$, $R_{\sigma\delta\sigma\delta}$, etc. We obtain an increasing sequence of families of sets, all being subfamilies of $\Sigma(R)$. One might hope that their union is $\Sigma(R)$ but in general this is not true either. In order to exhaust $\Sigma(R)$ by this procedure, one has to apply it uncountable many times, using the transfinite induction. This is, however, beyond the scope of this course.

As a consequence of this discussion, we see that a simple procedure used in Theorem 1.3 for extension of a measure from a semi-ring to a ring, is not going to work in the present setting and, hence, the procedure of extension is more complicated.

At the first step of extension a measure to a σ-ring, we assume that the full set M is also an element of R and, hence, $\mu(M) < \infty$. Consider the following example: let M be an interval in $\mathbb{R}$, R consists of all bounded subintervals of M and μ is the length of an interval. If M is bounded, then $M \in R$. However, if M is unbounded (for example, $M = \mathbb{R}$) then M does belong to R while M clearly belongs to $\Sigma(R)$ since M can be represented as a countable union of bounded intervals. It is also clear that for any reasonable extension of length, $\mu(M)$ in this case must be ∞.

The assumption $M \in R$ and, hence, $\mu(M) < \infty$, simplifies the situation, while the opposite case will be treated later on.

Definition. A ring containing M is called an *algebra*. A σ-ring in M that contains M is called a σ-*algebra*.

1.5.2 Outer measure

Henceforth, assume that R is an algebra. It follows from the definition of an algebra that if $A \in R$ then $A^c := M \setminus A \in R$. Also, let μ be a σ-additive measure on R. For any set $A \subset M$, define its *outer measure* $\mu^*(A)$ by

$$\mu^*(A) = \inf\left\{\sum_{k=1}^{\infty} \mu(A_k) : \ A_k \in R \text{ and } A \subset \bigcup_{k=1}^{\infty} A_l\right\}. \tag{1.7}$$

In other words, we consider all coverings $\{A_k\}_{k=1}^{\infty}$ of A by a sequence of sets from the algebra R and define $\mu^*(A)$ as the infimum of the sum of all $\mu(A_k)$, taken over all such coverings.

It follows from (1.7) that μ^* is monotone in the following sense: if $A \subset B$ then $\mu^*(A) \leq \mu^*(B)$. Indeed, any sequence $\{A_k\} \subset R$ that covers B, will cover also A, which implies that the infimum in (1.7) in the case of the set A is taken over a larger family, than in the case of B, whence $\mu^*(A) \leq \mu^*(B)$ follows.

Let us emphasize that $\mu^*(A)$ is defined on all subsets A of M, but μ^* is not necessarily a measure on 2^M. Eventually, we will construct a σ-algebra containing R where μ^* will be a σ-additive measure. This will be preceded by a sequence of Lemmas.

Lemma 1.4 *For any set $A \subset M$, $\mu^*(A) < \infty$. If in addition $A \in R$ then $\mu^*(A) = \mu(A)$.*

Proof. Note that $\emptyset \in R$ and $\mu(\emptyset) = 0$ because

$$\mu(\emptyset) = \mu(\emptyset \sqcup \emptyset) = \mu(\emptyset) + \mu(\emptyset).$$

For any set $A \subset M$, consider a covering $\{A_k\} = \{M, \emptyset, \emptyset, ...\}$ of A. Since $M, \emptyset \in R$, it follows from (1.7) that

$$\mu^*(A) \leq \mu(M) + \mu(\emptyset) + \mu(\emptyset) + ... = \mu(M) < \infty.$$

Assume now $A \in R$. Considering a covering $\{A_k\} = \{A, \emptyset, \emptyset, ...\}$ and using that $A, \emptyset \in R$, we obtain in the same way that

$$\mu^*(A) \leq \mu(A) + \mu(\emptyset) + \mu(\emptyset) + ... = \mu(A). \tag{1.8}$$

On the other hand, for any sequence $\{A_k\}$ as in (1.7) we have by the σ-subadditivity of μ (see Exercise 6) that

$$\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k).$$

Taking inf over all such sequences $\{A_k\}$, we obtain

$$\mu(A) \leq \mu^*(A),$$

which together with (1.8) yields $\mu^*(A) = \mu(A)$. ∎

Lemma 1.5 *The outer measure μ^* is σ-subadditive on 2^M.*

Proof. We need to prove that if

$$A \subset \bigcup_{k=1}^{\infty} A_k \tag{1.9}$$

where A and A_k are subsets of M then

$$\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k). \tag{1.10}$$

By the definition of μ^*, for any set A_k and for any $\varepsilon > 0$ there exists a sequence $\{A_{kn}\}_{n=1}^{\infty}$ of sets from R such that

$$A_k \subset \bigcup_{n=1}^{\infty} A_{kn} \tag{1.11}$$

and

$$\mu^*(A_k) \geq \sum_{n=1}^{\infty} \mu(A_{kn}) - \frac{\varepsilon}{2^k}.$$

Adding up these inequalities for all k, we obtain

$$\sum_{k=1}^{\infty} \mu^*(A_k) \geq \sum_{k,n=1}^{\infty} \mu(A_{kn}) - \varepsilon. \tag{1.12}$$

On the other hand, by (1.9) and (1.11), we obtain that

$$A \subset \bigcup_{k,n=1}^{\infty} A_{kn}.$$

Since $A_{kn} \in R$, we obtain by (1.7)

$$\mu^*(A) \leq \sum_{k,n=1}^{\infty} \mu(A_{kn}).$$

Comparing with (1.12), we obtain

$$\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k) + \varepsilon.$$

Since this inequality is true for any $\varepsilon > 0$, it follows that it is also true for $\varepsilon = 0$, which finishes the proof. ∎

1.5.3 Symmetric difference

Definition. The *symmetric difference* of two sets $A, B \subset M$ is the set

$$A \vartriangle B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

Clearly, $A \vartriangle B = B \vartriangle A$. Also, $x \in A \vartriangle B$ if and only if x belongs to exactly one of the sets A, B, that is, either $x \in A$ and $x \notin B$ or $x \notin A$ and $x \in B$.

Lemma 1.6 *(a) For arbitrary sets $A_1, A_2, B_1, B_2 \subset M$,*

$$(A_1 \circ A_2) \vartriangle (B_1 \circ B_2) \subset (A_1 \vartriangle B_1) \cup (A_2 \vartriangle B_2), \tag{1.13}$$

where $\circ$ denotes any of the operations $\cup$, $\cap$, $\setminus$.

(b) If μ^* is an outer measure on M then

$$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \vartriangle B), \tag{1.14}$$

for arbitrary sets $A, B \subset M$.

Proof. (a) Set

$$C = (A_1 \vartriangle B_1) \cup (A_2 \vartriangle B_2)$$

and verify (1.13) in the case when $\circ$ is $\cup$, that is,

$$(A_1 \cup A_2) \vartriangle (B_1 \cup B_2) \subset C. \tag{1.15}$$

By definition, $x \in (A_1 \cup A_2) \vartriangle (B_1 \cup B_2)$ if and only if $x \in A_1 \cup A_2$ and $x \notin B_1 \cup B_2$ or conversely $x \in B_1 \cup B_2$ and $x \notin A_1 \cup A_2$. Since these two cases are symmetric, it suffices to consider the first case. Then either $x \in A_1$ or $x \in A_2$ while $x \notin B_1$ and $x \notin B_2$. If $x \in A_1$ then

$$x \in A_1 \setminus B_1 \subset A_1 \vartriangle B_1 \subset C$$

that is, $x \in C$. In the same way one treats the case $x \in A_2$, which proves (1.15).

For the case when $\circ$ is $\cap$, we need to prove

$$(A_1 \cap A_2) \vartriangle (B_1 \cap B_2) \subset C. \tag{1.16}$$

Observe that $x \in (A_1 \cap A_2) \vartriangle (B_1 \cap B_2)$ means that $x \in A_1 \cap A_2$ and $x \notin B_1 \cap B_2$ or conversely. Again, it suffices to consider the first case. Then $x \in A_1$ and $x \in A_2$ while either $x \notin B_1$ or $x \notin B_2$. If $x \notin B_1$ then $x \in A_1 \setminus B_1 \subset C$, and if $x \notin B_2$ then $x \in A_2 \setminus B_2 \subset C$.

For the case when $\circ$ is $\setminus$, we need to prove

$$(A_1 \setminus A_2) \vartriangle (B_1 \setminus B_2) \subset C. \tag{1.17}$$

Observe that $x \in (A_1 \setminus A_2) \vartriangle (B_1 \setminus B_2)$ means that $x \in A_1 \setminus A_2$ and $x \notin B_1 \setminus B_2$ or conversely. Consider the first case, when $x \in A_1$, $x \notin A_2$ and either $x \notin B_1$ or $x \in B_2$. If $x \notin B_1$ then combining with $x \in A_1$ we obtain $x \in A_1 \setminus B_1 \subset C$. If $x \in B_2$ then combining with $x \notin A_2$, we obtain

$$x \in B_2 \setminus A_2 \subset A_2 \vartriangle B_2 \subset C,$$

which finishes the proof.

(b) Note that

$$A \subset B \cup (A \setminus B) \subset B \cup (A \vartriangle B)$$

whence by the subadditivity of μ^* (Lemma 1.5)

$$\mu^*(A) \leq \mu^*(B) + \mu^*(A \vartriangle B),$$

whence

$$\mu^*(A) - \mu^*(B) \leq \mu^*(A \vartriangle B).$$

Switching A and B, we obtain a similar estimate

$$\mu^*(B) - \mu^*(A) \leq \mu^*(A \vartriangle B),$$

whence (1.14) follows. $\blacksquare$

Remark The only property of μ^* used here was the finite subadditivity. So, inequality (1.14) holds for any finitely subadditive functional.

1.5.4 Measurable sets

We continue considering the case when R is an algebra on M and μ is a σ-additive measure on R. Recall that the outer measure μ^* is defined by (1.7).

Definition. A set $A \subset M$ is called *measurable* (with respect to the algebra R and the measure μ) if, for any $\varepsilon > 0$, there exists $B \in R$ such that

$$\mu^*(A \vartriangle B) < \varepsilon. \tag{1.18}$$

In other words, set A is measurable if it can be approximated by sets from R arbitrarily closely, in the sense of (1.18).

Now we can state one of the main theorems in this course.

Theorem 1.7 (Carathéodory's extension theorem) *Let R be an algebra on a set M and μ be a σ-additive measure on R. Denote by $\mathcal{M}$ the family of all measurable subsets of M. Then the following is true.*

(a) $\mathcal{M}$ is a σ-algebra containing R.

(b) The restriction of μ^* on $\mathcal{M}$ is a σ-additive measure (that extends measure μ from R to $\mathcal{M}$).

(c) If $\widetilde{\mu}$ is a σ-additive measure defined on a σ-algebra Σ such that

$$R \subset \Sigma \subset \mathcal{M},$$

then $\widetilde{\mu} = \mu^*$ on Σ.

Hence, parts (a) and (b) of this theorem ensure that a σ-additive measure μ can be extended from the algebra R to the σ-algebra of all measurable sets $\mathcal{M}$. Moreover, applying (c) with $\Sigma = \mathcal{M}$, we see that this extension is unique.

Since the minimal σ-algebra $\Sigma(R)$ is contained in $\mathcal{M}$, it follows that measure μ can be extended from R to $\Sigma(R)$. Applying (c) with $\Sigma = \Sigma(R)$, we obtain that this extension is also unique.

Proof. We split the proof into a series of claims.

Claim 1 *The family $\mathcal{M}$ of all measurable sets is an algebra containing R.*

If $A \in R$ then A is measurable because

$$\mu^*(A \vartriangle A) = \mu^*(\emptyset) = \mu(\emptyset) = 0$$

where $\mu^*(\emptyset) = \mu(\emptyset)$ by Lemma 1.4. Hence, $R \subset \mathcal{M}$. In particular, also the entire space M is a measurable set.

In order to verify that $\mathcal{M}$ is an algebra, it suffices to show that if $A_1, A_2 \in \mathcal{M}$ then also $A_1 \cup A_2$ and $A_1 \setminus A_2$ are measurable. Let us prove this for $A = A_1 \cup A_2$. By definition, for any $\varepsilon > 0$ there are sets $B_1, B_2 \in R$ such that

$$\mu^* (A_1 \vartriangle B_1) < \varepsilon \text{ and } \mu^* (A_2 \vartriangle B_2) < \varepsilon. \tag{1.19}$$

Setting $B = B_1 \cup B_2 \in R$, we obtain by Lemma 1.6,

$$A \vartriangle B \subset (A_1 \vartriangle B_1) \cup (A_2 \vartriangle B_2)$$

and by the subadditivity of μ^* (Lemma 1.5)

$$\mu^* (A \vartriangle B) \le \mu^* (A_1 \vartriangle B_1) + \mu^* (A_2 \vartriangle B_2) < 2\varepsilon. \tag{1.20}$$

Since $\varepsilon > 0$ is arbitrary and $B \in R$ we obtain that A satisfies the definition of a measurable set.

The fact that $A_1 \setminus A_2 \in \mathcal{M}$ is proved in the same way.

Claim 2 *μ^* is σ-additive on $\mathcal{M}$.*

Since $\mathcal{M}$ is an algebra and μ^* is σ-subadditive by Lemma 1.5, it suffices to prove that μ^* is finitely additive on $\mathcal{M}$ (see Exercise 9).

Let us prove that, for any two disjoint measurable sets A_1 and A_2, we have

$$\mu^* (A) = \mu^* (A_1) + \mu^* (A_2)$$

where $A = A_1 \bigsqcup A_2$. By Lemma 1.5, we have the inequality

$$\mu^* (A) \le \mu^* (A_1) + \mu^* (A_2)$$

so that we are left to prove the opposite inequality

$$\mu^* (A) \ge \mu^* (A_1) + \mu^* (A_2).$$

As in the previous step, for any $\varepsilon > 0$ there are sets $B_1, B_2 \in R$ such that (1.19) holds. Set $B = B_1 \cup B_2 \in R$ and apply Lemma 1.6, which says that

$$|\mu^* (A) - \mu^* (B)| \le \mu^* (A \vartriangle B) < 2\varepsilon,$$

where in the last inequality we have used (1.20). In particular, we have

$$\mu^* (A) \ge \mu^* (B) - 2\varepsilon. \tag{1.21}$$

On the other hand, since $B \in R$, we have by Lemma 1.4 and the additivity of μ on R, that

$$\mu^* (B) = \mu (B) = \mu (B_1 \cup B_2) = \mu (B_1) + \mu (B_2) - \mu (B_1 \cap B_2). \tag{1.22}$$

Next, we will estimate here $\mu (B_i)$ from below via $\mu^* (A_i)$, and show that $\mu (B_1 \cap B_2)$ is small enough. Indeed, using (1.19) and Lemma 1.6, we obtain, for any $i = 1, 2$,

$$|\mu^* (A_i) - \mu^* (B_i)| \le \mu^* (A_i \vartriangle B_i) < \varepsilon$$

whence

$$\mu (B_1) \ge \mu^* (A_1) - \varepsilon \text{ and } \mu (B_2) \ge \mu^* (A_2) - \varepsilon. \tag{1.23}$$

On the other hand, by Lemma 1.6 and using $A_1 \cap A_2 = \emptyset$, we obtain

$$B_1 \cap B_2 = (A_1 \cap A_2) \vartriangle (B_1 \cap B_2) \subset (A_1 \vartriangle B_1) \cup (A_2 \vartriangle B_2)$$

whence by (1.20)

$$\mu (B_1 \cap B_2) = \mu^* (B_1 \cap B_2) \leq \mu^* (A_1 \vartriangle B_1) + \mu^* (A_2 \vartriangle B_2) < 2\varepsilon. \tag{1.24}$$

It follows from (1.21)–(1.24) that

$$\mu^* (A) \geq (\mu^* (A_1) - \varepsilon) + (\mu^* (A_2) - \varepsilon) - 2\varepsilon - 2\varepsilon = \mu^* (A_1) + \mu^* (A_2) - 6\varepsilon.$$

Letting $\varepsilon \to 0$, we finish the proof.

Claim 3 *$\mathcal{M}$ is σ-algebra.*

Assume that $\{A_n\}_{n=1}^\infty$ is a sequence of measurable sets and prove that $A := \bigcup_{n=1}^\infty A_n$ is also measurable. Note that

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2 \setminus A_1) ... = \bigsqcup_{n=1}^\infty \widetilde{A}_n,$$

where

$$\widetilde{A}_n := A_n \setminus A_{n-1} \setminus ... \setminus A_1 \in \mathcal{M}$$

(here we use the fact that $\mathcal{M}$ is an algebra – see Claim 1). Therefore, renaming $\widetilde{A}_n$ to A_n, we see that it suffices to treat the case of a disjoint union: given $A = \bigsqcup_{n=1}^\infty A_n$ where $A_n \in \mathcal{M}$, prove that $A \in \mathcal{M}$.

Note that, for any fixed N,

$$\mu^* (A) \geq \mu^* \left(\bigsqcup_{n=1}^N A_n \right) = \sum_{n=1}^N \mu^* (A_n),$$

where we have used the monotonicity of μ^* and the additivity of μ^* on $\mathcal{M}$ (Claim 2). This implies by $N \to \infty$ that

$$\sum_{n=1}^\infty \mu^* (A_n) \leq \mu^* (A) < \infty$$

(cf. Lemma 1.4), so that the series $\sum_{n=1}^\infty \mu^* (A_n)$ converges. In particular, for any $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\sum_{n=N+1}^\infty \mu^* (A_n) < \varepsilon.$$

Setting

$$A' = \bigsqcup_{n=1}^N A_n \text{ and } A'' = \bigsqcup_{n=N+1}^\infty A_n,$$

we obtain by the σ-subadditivity of μ^*

$$\mu^* (A'') \leq \sum_{n=N+1}^\infty \mu^* (A_n) < \varepsilon.$$

By Claim 1, the set A' is measurable as a finite union of measurable sets. Hence, there is $B \in R$ such that

$$\mu^* (A' \vartriangle B) < \varepsilon.$$

Since $A = A' \cup A''$, we have

$$A \vartriangle B \subset (A' \vartriangle B) \cup A''. \tag{1.25}$$

Indeed, $x \in A \vartriangle B$ means that $x \in A$ and $x \notin B$ or $x \notin A$ and $x \in B$. In the first case, we have $x \in A'$ or $x \in A''$. If $x \in A'$ then together with $x \notin B$ it gives

$$x \in A' \vartriangle B \subset (A' \vartriangle B) \cup A''.$$

If $x \in A''$ then the inclusion is obvious. In the second case, we have $x \notin A'$ which together with $x \in B$ implies $x \in A' \vartriangle B$, which finishes the proof of (1.25).

It follows from (1.25) that

$$\mu^* (A \vartriangle B) \le \mu^* (A' \vartriangle B) + \mu^* (A'') < 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary and $B \in R$, we conclude that $A \in \mathcal{M}$.

Claim 4 *Let Σ be a σ-algebra such that*

$$R \subset \Sigma \subset \mathcal{M}$$

and let $\widetilde{\mu}$ be a σ-additive measure on Σ such that $\widetilde{\mu} = \mu$ on R. Then $\widetilde{\mu} = \mu^*$ on Σ.

We need to prove that $\widetilde{\mu}(A) = \mu^*(A)$ for any $A \in \Sigma$ By the definition of μ^*, we have

$$\mu^* (A) = \inf \left\{ \sum_{n=1}^{\infty} \mu (A_n) : A_n \in R \text{ and } A \subset \bigcup_{n=1}^{\infty} A_n \right\}.$$

Applying the σ-subadditivity of $\widetilde{\mu}$ (see Exercise 7), we obtain

$$\widetilde{\mu}(A) \le \sum_{n=1}^{\infty} \widetilde{\mu}(A_n) = \sum_{n=1}^{\infty} \mu (A_n).$$

Taking inf over all such sequences $\{A_n\}$, we obtain

$$\widetilde{\mu}(A) \le \mu^* (A). \tag{1.26}$$

On the other hand, since A is measurable, for any $\varepsilon > 0$ there is $B \in R$ such that

$$\mu^* (A \vartriangle B) < \varepsilon. \tag{1.27}$$

By Lemma 1.6(b),

$$|\mu^* (A) - \mu^* (B)| \le \mu^* (A \vartriangle B) < \varepsilon. \tag{1.28}$$

Note that the proof of Lemma 1.6(b) uses only subadditivity of μ^*. Since $\widetilde{\mu}$ is also subadditive, we obtain that

$$|\widetilde{\mu}(A) - \widetilde{\mu}(B)| \le \widetilde{\mu}(A \vartriangle B) \le \mu^* (A \vartriangle B) < \varepsilon,$$

where we have also used (1.26) and (1.27) . Combining with (1.28) and using $\widetilde{\mu}(B) = \mu(B) = \mu^*(B)$, we obtain

$$|\widetilde{\mu}(A) - \mu^* (A)| \le |\mu^* (A) - \mu^* (B)| + |\widetilde{\mu}(A) - \widetilde{\mu}(B)| < 2\varepsilon.$$

Letting $\varepsilon \to 0$, we conclude that $\widetilde{\mu}(A) = \mu^*(A)$,which finishes the proof. ∎

1.6 σ-finite measures

Recall Theorem 1.7: any σ-additive measure on an algebra R can be uniquely extended to the σ-algebra of all measurable sets (and to the minimal σ-algebra containing R).

Theorem 1.7 has a certain restriction for applications: it requires that the full set M belongs to the initial ring R (so that R is an algebra). For example, if $M = \mathbb{R}$ and R is the ring of the bounded intervals then this condition is not satisfied. Moreover, for any reasonable extension of the notion of length, the length of the entire line $\mathbb{R}$ must be ∞. This suggest that we should extended the notion of a measure to include also the values of ∞.

So far, we have not yet defined the term "measure" itself using "finitely additive measures" and "σ-additive measures". Now we define the notion of a measure as follows.

Definition. Let M be a non-empty set and S be a family of subsets of M. A functional $\mu : S \to [0, +\infty]$ is called a *measure* if, for all sets $A, A_k \in S$ such that $A = \bigsqcup_{k=1}^{N} A_k$ (where N is either finite or infinite), we have

$$\mu(A) = \sum_{k=1}^{N} \mu(A_k).$$

Hence, a measure is always σ-additive. The difference with the previous definition of "σ-additive measure" is that we allow for measure to take value $+\infty$. In the summation, we use the convention that $(+\infty) + x = +\infty$ for any $x \in [0, +\infty]$.

Example. Let S be the family of **all** intervals on $\mathbb{R}$ (including the unbounded intervals) and define the length of any interval $I \subset \mathbb{R}$ with the endpoints a, b by $\ell(I) = |b - a|$ where a and b may take values from $[-\infty, +\infty]$. Clearly, for any unbounded interval I we have $\ell(I) = +\infty$. We claim that ℓ is a measure on S in the above sense.

Indeed, let $I = \bigsqcup_{k=1}^{N} I_k$ where $I, I_k \in S$ and N is either finite or infinite. We need to prove that

$$\ell(I) = \sum_{k=1}^{N} \ell(I_k) \tag{1.29}$$

If I is bounded then all intervals I_k are bounded, and (1.29) follows from Theorem 1.1. If one of the intervals I_k is unbounded then also I is unbounded, and (1.29) is true because the both sides are $+\infty$. Finally, consider the case when I is unbounded while all I_k are bounded. Choose any bounded closed interval $I' \subset I$. Since $I' \subset \bigcup_{k=1}^{\infty} I_k$, by the σ-subadditivity of the length on the ring of bounded intervals (Lemma 1.2), we obtain

$$\ell(I') \leq \sum_{k=1}^{N} \ell(I_k).$$

By the choose of I', the length $\ell(I')$ can be made arbitrarily large, whence it follows that

$$\sum_{k=1}^{N} \ell(I_k) = +\infty = \ell(I).$$

Of course, allowing the value $+\infty$ for measures, we receive trivial examples of measures as follows: just set $\mu(A) = +\infty$ for any $A \subset M$. The following definition is to ensure the existence of plenty of sets with finite measures.

Definition. A measure μ on a set S is called *finite* if $M \in S$ and $\mu(M) < \infty$. A measure μ is called *σ-finite* if there is a sequence $\{B_k\}_{k=1}^{\infty}$ of sets from S such that $\mu(B_k) < \infty$ and $M = \bigcup_{k=1}^{\infty} B_k$.

Of course, any finite measure is σ-finite. The measure in the setting of Theorem 1.7 is finite. The length defined on the family of all intervals in $\mathbb{R}$, is σ-finite but not finite.

Let R be a ring in a set M. Our next goal is to extend a σ-finite measure μ from R to a σ-algebra. For any $B \in R$, define the family R_B of subsets of B by

$$R_B = R \cap 2^B = \{A \subset B : A \in R\}.$$

Observe that R_B is an algebra in B (indeed, R_B is a ring as the intersection of two rings, and also $B \in R_B$). If $\mu(B) < \infty$ then μ is a finite measure on R_B and by Theorem 1.7, μ extends uniquely to a finite measure μ_B on the σ-algebra $\mathcal{M}_B$ of measurable subsets of B. Our purpose now is to construct a σ-algebra $\mathcal{M}$ of measurable sets in M and extend measure μ to a measure on $\mathcal{M}$.

Let $\{B_k\}_{k=1}^{\infty}$ be a sequence of sets from R such that $M = \bigcup_{k=1}^{\infty} B_k$ and $\mu(B_k) < \infty$ (such sequence exists by the σ-finiteness of μ). Replacing the sets B_k by the differences

$$B_1, B_2 \setminus B_1, B_3 \setminus B_1 \setminus B_2, ...,$$

we can assume that all B_k are disjoint, that is, $M = \bigsqcup_{k=1}^{\infty} B_k$. In what follows, we fix such a sequence $\{B_k\}$.

Definition. A set $A \in M$ is called measurable if $A \cap B_k \in \mathcal{M}_{B_k}$ for any k. For any measurable set A, set

$$\mu_M(A) = \sum_k \mu_{B_k}(A \cap B_k). \tag{1.30}$$

Theorem 1.8 *Let μ be a σ-finite measure on a ring R and $\mathcal{M}$ be the family of all measurable sets defined above. Then the following is true.*

(a) $\mathcal{M}$ is a σ-algebra containing R.

(b) The functional μ_M defined by (1.30) *is a measure on $\mathcal{M}$ that extends measure μ on R.*

(c) If $\widetilde{\mu}$ is a measure defined on a σ-algebra Σ such that

$$R \subset \Sigma \subset \mathcal{M},$$

then $\widetilde{\mu} = \mu_M$ on Σ.

Proof. (a) If $A \in R$ then $A \cap B_k \in R$ because $B_k \in R$. Therefore, $A \cap B_k \in R_{B_k}$ whence $A \cap B_k \in \mathcal{M}_{B_k}$ and $A \in \mathcal{M}$. Hence, $R \subset \mathcal{M}$.

Let us show that if $A = \bigcup_{n=1}^{N} A_n$ where $A_n \in \mathcal{M}$ then also $A \in \mathcal{M}$ (where N can be ∞). Indeed, we have

$$A \cap B_k = \bigcup_n (A_n \cap B_k) \in \mathcal{M}_{B_k}$$

because $A_n \cap B_k \in \mathcal{M}_{B_k}$ and $\mathcal{M}_{B_k}$ is a σ-algebra. Therefore, $A \in \mathcal{M}$. In the same way, if $A', A'' \in \mathcal{M}$ then the difference $A = A' \setminus A''$ belongs to $\mathcal{M}$ because

$$A \cap B_k = (A' \cap B_k) \setminus (A'' \cap B_k) \in \mathcal{M}_{B_k}.$$

Finally, $M \in \mathcal{M}$ because

$$M \cap B_k = B_k \in \mathcal{M}_{B_k}.$$

Hence, $\mathcal{M}$ satisfies the definition of a σ-algebra.

(b) If $A \in R$ then $A \cap B_k \in R_{B_k}$ whence $\mu_{B_k}(A \cap B_k) = \mu(A \cap B_k)$. Since

$$A = \bigsqcup_k (A \cap B_k),$$

and μ is a measure on R, we obtain

$$\mu(A) = \sum_k \mu(A \cap B_k) = \sum_k \mu_{B_k}(A \cap B_k).$$

Comparing with (1.30), we obtain $\mu_M(A) = \mu(A)$. Hence, μ_M on R coincides with μ.

Let us show that μ_M is a measure. Let $A = \bigsqcup_{n=1}^N A_n$ where $A_n \in \mathcal{M}$ and N is either finite or infinite. We need to prove that

$$\mu_M(A) = \sum_n \mu_M(A_n).$$

Indeed, we have

$$\begin{aligned}
\sum_n \mu_M(A_n) &= \sum_n \sum_k \mu_{B_k}(A_n \cap B_k) \\
&= \sum_k \sum_n \mu_{B_k}(A_n \cap B_k) \\
&= \sum_k \mu_{B_k}\left(\bigsqcup_n (A_n \cap B_k)\right) \\
&= \sum_k \mu_{B_k}(A \cap B_k) \\
&= \mu_M(A),
\end{aligned}$$

where we have used the fact that

$$A \cap B_k = \bigsqcup_n (A_n \cap B_k)$$

and the σ-additivity of measure μ_{B_k}.

(c) Let $\widetilde{\mu}$ be another measure defined on a σ-algebra Σ such that $R \subset \Sigma \subset \mathcal{M}$ and $\widetilde{\mu} = \mu$ on R. Let us prove that $\widetilde{\mu}(A) = \mu_M(A)$ for any $A \in \Sigma$. Observing that $\widetilde{\mu}$ and μ coincide on R_{B_k}, we obtain by Theorem 1.7 that $\widetilde{\mu}$ and μ_{B_k} coincide on $\Sigma_{B_k} := \Sigma \cap 2^{B_k}$. Then for any $A \in \Sigma$, we have $A = \bigsqcup_k (A \cap B_k)$ and $A \cap B_k \in \Sigma_{B_k}$ whence

$$\widetilde{\mu}(A) = \sum_k \widetilde{\mu}(A \cap B_k) = \sum_k \mu_{B_k}(A \cap B_k) = \sum_k \mu_M(A),$$

which finishes the proof. ∎

Remark. The definition of a measurable set uses the decomposition $M = \bigsqcup_k B_k$ where B_k are sets from R with finite measure. It seems to depend on the choice of B_k but in fact it does not. Here is an equivalent definition: a set $A \subset M$ is called measurable if $A \cap B \in R_B$ for any $B \in R$. For the proof see Exercise 19.

1.7 Null sets

Let R be a ring on a set M and μ be a finite measure on R (that is, μ is a σ-additive functional on R and $\mu(M) < \infty$). Let μ^* be the outer measure as defined by (1.7).

Definition. A set $A \subset M$ is called a *null set* (or a set of measure 0) if $\mu^*(A) = 0$. The family of all null sets is denoted by $\mathcal{N}$.

Using the definition (1.7) of the outer measure, one can restate the condition $\mu^*(A) = 0$ as follows: for any $\varepsilon > 0$ there exists a sequence $\{A_k\}_{k=1}^{\infty} \subset R$ such that

$$A \subset \bigcup_k A_k \quad \text{and} \quad \sum_k \mu(A_k) < \varepsilon.$$

If the ring R is the minimal ring containing a semi-ring S (that is, $R = R(S)$) then the sequence $\{A_k\}$ in the above definition can be taken from S. Indeed, if $A_k \subset R$ then, by Theorem 1.3, A_k is a disjoint finite union of some sets $\{A_{kn}\}$ from S where n varies in a finite set. It follows that

$$\mu(A_k) = \sum_n \mu(A_{kn})$$

and

$$\sum_k \mu(A_k) = \sum_k \sum_n \mu(A_{kn}).$$

Since the double sequence $\{A_{kn}\}$ covers A, the sequence $\{A_k\} \subset R$ can be replaced by the double sequence $\{A_{kn}\} \subset S$.

Example. Let S be the semi-ring of intervals in $\mathbb{R}$ and μ be the length. It is easy to see that a single point set $\{x\}$ is a null set. Indeed, for any $\varepsilon > 0$ there is an interval I covering x and such that $\ell(I) < \varepsilon$. Moreover, we claim that any countable set $A \subset \mathbb{R}$ is a null set. Indeed, if $A = \{x_k\}_{k=1}^{\infty}$ then cover x_k by an interval I_k of length $< \frac{\varepsilon}{2^k}$ so that the sequence $\{I_k\}$ of intervals covers A and $\sum_k \ell(I_k) < \varepsilon$. Hence, A is a null set. For example, the set $\mathbb{Q}$ of all rationals is a null set. The Cantor set from Exercise 13 is an example of an uncountable set of measure 0.

In the same way, any countable subset of $\mathbb{R}^2$ is a null set (with respect to the area).

Lemma 1.9 *(a) Any subset of a null set is a null set.*

(b) The family $\mathcal{N}$ of all null sets is a σ-ring.

(c) Any null set is measurable, that is, $\mathcal{N} \subset \mathcal{M}$.

Proof. (a) It follows from the monotonicity of μ^* that $B \subset A$ implies $\mu^*(B) \leq \mu^*(A)$. Hence, if $\mu^*(A) = 0$ then also $\mu^*(B) = 0$.

(b) If $A, B \in \mathcal{N}$ then $A \setminus B \in \mathcal{N}$ by part (a), because $A \setminus B \subset A$. Let $A = \bigcup_{n=1}^{N} A_n$ where $A_n \in \mathcal{N}$ and N is finite or infinite. Then, by the σ-subadditivity of μ^* (Lemma 1.5), we have

$$\mu^*(A) \leq \sum_n \mu^*(A_n) = 0$$

whence $A \in \mathcal{N}$.

(c) We need to show that if $A \in \mathcal{N}$ then, for any $\varepsilon > 0$ there is $B \in R$ such that $\mu^*(A \triangle B) < \varepsilon$. Indeed, just take $B = \emptyset \in R$. Then $\mu^*(A \triangle B) = \mu^*(A) = 0 < \varepsilon$, which was to be proved. ∎

Let now μ be a σ-finite measure on a ring R. Then we have $M = \bigsqcup_{k=1}^{\infty} B_k$ where $B_k \in R$ and $\mu(B_k) < \infty$.

Definition. In the case of a σ-finite measure, a set $A \subset M$ is called a null set if $A \cap B_k$ is a null set in each B_k.

Lemma 1.9 easily extends to σ-finite measures: one has just to apply the corresponding part of Lemma 1.9 to each B_k.

Let $\Sigma = \Sigma(R)$ be the minimal σ-algebra containing a ring R so that

$$R \subset \Sigma \subset \mathcal{M}.$$

The relation between two σ-algebras Σ and $\mathcal{M}$ is given by the following theorem.

Theorem 1.10 *Let μ be a σ-finite measure on a ring R. Then $A \in \mathcal{M}$ if and only if there is $B \in \Sigma$ such that $A \vartriangle B \in \mathcal{N}$, that is, $\mu^*(A \vartriangle B) = 0$.*

This can be equivalently stated as follows:

- $A \in \mathcal{M}$ if and only if there is $B \in \Sigma$ and $N \in \mathcal{N}$ such that $A = B \vartriangle N$.
- $A \in \mathcal{M}$ if and only if there is $N \in \mathcal{N}$ such that $A \vartriangle N \in \Sigma$.

Indeed, Theorem 1.10 says that for any $A \in \mathcal{M}$ there is $B \in \Sigma$ and $N \in \mathcal{N}$ such that $A \vartriangle B = N$. By the properties of the symmetric difference, the latter is equivalent to each of the following identities: $A = B \vartriangle N$ and $B = A \vartriangle N$ (see Exercise 15), which settle the claim.

Proof of Theorem 1.10. We prove the statement in the case when measure μ is finite. The case of a σ-finite measure follows then straightforwardly.

By the definition of a measurable set, for any $n \in \mathbb{N}$ there is $B_n \in R$ such that

$$\mu^*(A \vartriangle B_n) < 2^{-n}.$$

The set $B \in \Sigma$, which is to be found, will be constructed as a sort of limit of B_n as $n \to \infty$. For that, set

$$C_n = B_n \cup B_{n+1} \cup \ldots = \bigcup_{k=n}^{\infty} B_k$$

so that $\{C_n\}_{n=1}^{\infty}$ is a decreasing sequence, and define B by

$$B = \bigcap_{n=1}^{\infty} C_n.$$

Clearly, $B \in \Sigma$ since B is obtain from $B_n \in R$ by countable unions and intersections. Let us show that

$$\mu^*(A \vartriangle B) = 0.$$

We have

$$A \vartriangle C_n = A \vartriangle \bigcup_{k=n}^{\infty} B_k \subset \bigcup_{k=n}^{\infty} (A \vartriangle B_k)$$

(see Exercise 15) whence by the σ-subadditivity of μ^* (Lemma 1.5)

$$\mu^*(A \vartriangle C_n) \le \sum_{k=n}^{\infty} \mu^*(A \vartriangle B_k) < \sum_{k=n}^{\infty} 2^{-k} = 2^{1-n}.$$

Since the sequence $\{C_n\}_{n=1}^{\infty}$ is decreasing, the intersection of all sets C_n does not change if we omit finitely many terms. Hence, we have for any $N \in \mathbb{N}$

$$B = \bigcap_{n=N}^{\infty} C_n.$$

Hence, using again Exercise 15, we have

$$A \vartriangle B = A \vartriangle \bigcap_{n=N}^{\infty} C_n \subset \bigcup_{n=N}^{\infty} (A \vartriangle C_n)$$

whence

$$\mu^* (A \vartriangle B) \leq \sum_{n=N}^{\infty} \mu^* (A \vartriangle C_n) \leq \sum_{n=N}^{\infty} 2^{1-n} = 2^{2-N}.$$

Since N is arbitrary here, letting $N \to \infty$ we obtain $\mu^* (A \vartriangle B) = 0$.

Conversely, let us show that if $A \vartriangle B$ is null set for some $B \in \Sigma$ then A is measurable. Indeed, the set $N = A \vartriangle B$ is a null set and, by Lemma 1.9, is measurable. Since $A = B \vartriangle N$ and both B and N are measurable, we conclude that A is also measurable. ■

1.8 Lebesgue measure in $\mathbb{R}^n$

Now we apply the above theory of extension of measures to $\mathbb{R}^n$. For that, we need the notion of the *product measure.*

1.8.1 Product measure

Let M_1 and M_2 be two non-empty sets, and S_1, S_2 be families of subsets of M_1 and M_2, respectively. Consider the set

$$M = M_1 \times M_2 := \{(x, y) : x \in M_1, y \in M_2\}$$

and the family S of subsets of M, defined

$$S = S_1 \times S_2 := \{A \times B : A \in S_1, B \in S_2\}.$$

For example, if $M_1 = M_2 = \mathbb{R}$ then $M = \mathbb{R}^2$. If S_1 and S_2 are families of all intervals in $\mathbb{R}$ then S consists of all rectangles in $\mathbb{R}^2$.

Claim. *If S_1 and S_2 are semi-rings then S is also a semi-ring (see Exercise 6).*

In the sequel, let S_1 and S_2 be semi-rings.

Let μ_1 be a finitely additive measure on the semi-ring S_1 and μ_2 be a finitely additive measure on the semi-ring S_2. Define the *product measure* $\mu = \mu_1 \times \mu_2$ on S as follows: if $A \in S_1$ and $B \in S_2$ then set

$$\mu (A \times B) = \mu_1 (A) \mu_2 (B).$$

Claim. *μ is also a finitely additive measure on S (see By Exercise 20).*

Remark. It is true that if μ_1 and μ_2 are σ-additive then μ is σ-additive as well, but the proof in the full generality is rather hard and will be given later in the course.

By induction, one defines the product of n semi-rings and the product of n measures for any $n \in \mathbb{N}$.

1.8.2 Construction of measure in $\mathbb{R}^n$.

Let S_1 be the semi-ring of all bounded intervals in $\mathbb{R}$ and define S_n (being the family of subsets of $\mathbb{R}^n = \mathbb{R} \times ... \times \mathbb{R}$) as the product of n copies of S_1:

$$S_n = S_1 \times S_1 \times ... \times S_1.$$

That is, any set $A \in S_n$ has the form

$$A = I_1 \times I_2 \times ... \times I_n \tag{1.31}$$

where I_k are bounded intervals in $\mathbb{R}$. In other words, S_n consists of bounded *boxes*. Clearly, S_n is a semi-ring as the product of semi-rings. Define the product measure λ_n on S_n by

$$\lambda_n = \ell \times ... \times \ell$$

where ℓ is the length on S_1. That is, for the set A from (1.31),

$$\lambda_n (A) = \ell (I_1) ....\ell (I_n) .$$

Then λ_n is a finitely additive measure on the semi-ring S_n in $\mathbb{R}^n$.

Lemma 1.11 *λ_n is a σ-additive measure on S_n.*

Proof. We use the same approach as in the proof of σ-additivity of ℓ (Theorem 1.1), which was based on the finite additivity of ℓ and on the compactness of a closed bounded interval. Since λ_n is finitely additive, in order to prove that it is also σ-additive, it suffices to prove that λ_n is *regular* in the following sense: for any $A \in S_n$ and any $\varepsilon > 0$, there exist a closed box $K \in S_n$ and an open box $U \in S_n$ such that

$$K \subset A \subset U \text{ and } \lambda_n (U) \leq \lambda_n (K) + \varepsilon$$

(see Exercise 17). Indeed, if A is as in (1.31) then, for any $\delta > 0$ and for any index j, there is a closed interval K_j and an open interval U_j such that

$$K_j \subset I_j \subset U_j \text{ and } \ell (U_j) \leq \ell (K_j) + \delta.$$

Set $K = K_1 \times ... \times K_n$ and $U = U_1 \times ... \times U_n$ so that K is closed, U is open, and

$$K \subset A \subset U.$$

It also follows that

$$\lambda_n (U) = \ell (U_1) ...\ell (U_n) \leq (\ell (K_1) + \delta) ... (\ell (K_n) + \delta) < \ell (K_1) ...\ell (K_n) + \varepsilon = \lambda_n (K) + \varepsilon,$$

provided $\delta = \delta (\varepsilon)$ is chosen small enough. Hence, λ_n is regular, which finishes the proof. ∎

Note that measure λ_n is also σ-finite. Indeed, let us cover $\mathbb{R}$ by a sequence $\{I_k\}_{k=1}^{\infty}$ of bounded intervals. Then all possible products $I_{k_1} \times I_{k_2} \times ... \times I_{k_n}$ forms a covering of $\mathbb{R}^n$ by a sequence of bounded boxes. Hence, λ_n is a σ-finite measure on S_n.

By Theorem 1.3, λ_n can be uniquely extended as a σ-additive measure to the minimal ring $R_n = R(S_n)$, which consists of finite union of bounded boxes. Denote this extension also by λ_n. Then λ_n is a σ-finite measure on R_n. By Theorem 1.8, λ_n extends uniquely to the σ-algebra $\mathcal{M}_n$ of all measurable sets in $\mathbb{R}^n$. Denote this extension also by λ_n. Hence, λ_n is a measure on the σ-algebra $\mathcal{M}_n$ that contains R_n.

Definition. Measure λ_n on $\mathcal{M}_n$ is called the *Lebesgue measure* in $\mathbb{R}^n$. The measurable sets in $\mathbb{R}^n$ are also called *Lebesgue measurable.*

In particular, the measure λ_2 in $\mathbb{R}^2$ is called *area* and the measure λ_3 in $\mathbb{R}^3$ is called *volume.* Also, λ_n for any $n \geq 1$ is frequently referred to as an *n-dimensional volume.*

Definition. The minimal σ-algebra $\Sigma(R_n)$ containing R_n is called the *Borel* σ-algebra and is denoted by $\mathcal{B}_n$ (or by $\mathcal{B}(\mathbb{R}^n)$). The sets from $\mathcal{B}_n$ are called the *Borel sets* (or Borel measurable sets).

Hence, we have the inclusions $S_n \subset R_n \subset \mathcal{B}_n \subset \mathcal{M}_n$.

Example. Let us show that any open subset U of $\mathbb{R}^n$ is a Borel set. Indeed, for any $x \in U$ there is an open box B_x such that $x \in B_x \subset U$. Clearly, $B_x \in R_n$ and $U = \bigcup_{x\in U} B_x$. However, this does not immediately imply that U is Borel since the union is uncountable. To fix this, observe that B_x can be chosen so that all the coordinates of B_x (that is, all the endpoints of the intervals forming B_x) are rationals. The family of all possible boxes in $\mathbb{R}^n$ with rational coordinates is countable. For every such box, we can see if it occurs in the family $\{B_x\}_{x\in U}$ or not. Taking only those boxes that occur in this family, we obtain an at most countable family of boxes, whose union is also U. It follows that U is a Borel set as a countable union of Borel sets.

Since closed sets are complements of open sets, it follows that also closed sets are Borel sets. Then we obtain that countable intersections of open sets and countable unions of closed sets are also Borel, etc. Any set, that can be obtained from open and closed sets by applying countably many unions, intersections, subtractions, is again a Borel set.

Example. Any non-empty open set U has a positive Lebesgue measure, because U contains a non-empty box B and $\lambda(B) > 0$. As an example, let us show that the hyperplane $A = \{x_n = 0\}$ of $\mathbb{R}^n$ has measure 0. It suffices to prove that the intersection $A \cap B$ is a null set for any bounded box B in $\mathbb{R}^n$. Indeed, set

$$B_0 = A \cap B = \{x \in B : x_n = 0\}$$

and note that B_0 is a bounded box in $\mathbb{R}^{n-1}$. Choose $\varepsilon > 0$ and set

$$B_\varepsilon = B_0 \times (-\varepsilon, \varepsilon) = \{x \in B : |x_n| < \varepsilon\}$$

so that B_ε is a box in $\mathbb{R}^n$ and $B_0 \subset B_\varepsilon$. Clearly,

$$\lambda_n(B_\varepsilon) = \lambda_{n-1}(B_0)\,\ell(-\varepsilon, \varepsilon) = 2\varepsilon\lambda_{n-1}(B_0).$$

Since ε can be made arbitrarily small, we conclude that B_0 is a null set.

Remark. Recall that $\mathcal{B}_n \subset \mathcal{M}_n$ and, by Theorem 1.10, any set from $\mathcal{M}_n$ is the symmetric difference of a set from $\mathcal{B}_n$ and a null set. An interesting question is whether the family $\mathcal{M}_n$ of Lebesgue measurable sets is actually larger than the family $\mathcal{B}_n$ of Borel sets. The answer is yes. Although it is difficult to construct an explicit example of a set in $\mathcal{M}_n \setminus \mathcal{B}_n$

the fact that $\mathcal{M}_n \setminus \mathcal{B}_n$ is non-empty can be used comparing the cardinal numbers $|\mathcal{M}_n|$ and $|\mathcal{B}_n|$. Indeed, it is possible to show that

$$|\mathcal{B}_n| = |\mathbb{R}| < \left|2^{\mathbb{R}}\right| = |\mathcal{M}_n|\,, \tag{1.32}$$

which of course implies that $\mathcal{M}_n \setminus \mathcal{B}_n$ is non-empty. Let us explain (not prove) why (1.32) is true. As we have seen, any open set is the union of a countable sequence of boxes with rational coordinates. Since the cardinality of such sequences is $|\mathbb{R}|$, it follows that the family of all open sets has the cardinality $|\mathbb{R}|$. Then the cardinality of the countable intersections of open sets amounts to that of countable sequences of reals, which is again $|\mathbb{R}|$. Continuing this way and using the transfinite induction, one can show that the cardinality of the family of the Borel sets is $|\mathbb{R}|$.

To show that $|\mathcal{M}_n| = \left|2^{\mathbb{R}}\right|$, we use the fact that any subset of a null set is also a null set and, hence, is measurable. Assume first that $n \geq 2$ and let A be a hyperplane from the previous example. Since $|A| = |\mathbb{R}|$ and any subset of A belongs to $\mathcal{M}_n$, we obtain that $|\mathcal{M}_n| \geq \left|2^A\right| = \left|2^{\mathbb{R}}\right|$, which finishes the proof. If $n = 1$ then the example with a hyperplane does not work, but one can choose A to be the Cantor set (see Exercise 13), which has the cardinality $|\mathbb{R}|$ and measure 0. Then the same argument works.

1.9 Probability spaces

Probability theory can be considered as a branch of a measure theory where one uses specific notation and terminology, and considers specific questions. We start with the probabilistic notation and terminology.

Definition. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- Ω is a non-empty set, which is called the *sample space*.
- $\mathcal{F}$ is a σ-algebra of subsets of Ω, whose elements are called *events*.
- $\mathbb{P}$ is a *probability measure* on $\mathcal{F}$, that is, $\mathbb{P}$ is a measure on $\mathcal{F}$ and $\mathbb{P}(\Omega) = 1$ (in particular, $\mathbb{P}$ is a finite measure). For any event $A \in \mathcal{F}$, $\mathbb{P}(A)$ is called the *probability* of A.

Since $\mathcal{F}$ is an algebra, that is, $\Omega \in \mathcal{F}$, the operation of taking complement is defined in $\mathcal{F}$, that is, if $A \in \mathcal{F}$ then the complement $A^c := \Omega \setminus A$ is also in $\mathcal{F}$. The event A^c is opposite to A and

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega \setminus A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A)\,.$$

Example. 1. Let $\Omega = \{1, 2, ..., N\}$ be a finite set, and $\mathcal{F}$ is the set of all subsets of Ω. Given N non-negative numbers p_i such that $\sum_{i=1}^N p_i = 1$, define $\mathbb{P}$ by

$$\mathbb{P}(A) = \sum_{i \in A} p_i. \tag{1.33}$$

The condition (1.33) ensures that $\mathbb{P}(\Omega) = 1$.

For example, in the simplest case $N = 2$, $\mathcal{F}$ consists of the sets $\emptyset, \{1\}, \{2\}, \{1,2\}$. Given two non-negative numbers p and q such that $p+q = 1$, set $\mathbb{P}(\{1\}) = p$, $\mathbb{P}(\{2\}) = q$, while $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\{1,2\}) = 1$.

This example can be generalized to the case when Ω is a countable set, say, $\Omega = \mathbb{N}$. Given a sequence $\{p_i\}_{i=1}^{\infty}$ of non-negative numbers p_i such that $\sum_{i=1}^{\infty} p_i = 1$, define for any set $A \subset \Omega$ its probability by (1.33). Measure $\mathbb{P}$ constructed by means of (1.33) is called a *discrete* probability measure, and the corresponding space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *discrete* probability space.

2. Let $\Omega = [0,1]$, $\mathcal{F}$ be the set of all Lebesgue measurable subsets of $[0,1]$ and $\mathbb{P}$ be the Lebesgue measure λ_1 restricted to $[0,1]$. Clearly, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

1.10 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition. Two events A and B are called *independent* (unabhängig) if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

More generally, let $\{A_i\}$ be a family of events parametrized by an index i. Then the family $\{A_i\}$ is called *independent* (or one says that the events A_i are independent) if, for any finite set of distinct indices $i_1, i_2, ..., i_k$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})...\mathbb{P}(A_{i_k}). \tag{1.34}$$

For example, three events A, B, C are independent if

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B), \ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \ \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Example. In any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, any couple A, Ω is independent, for any event $A \in \mathcal{F}$, which follows from

$$\mathbb{P}(A \cap \Omega) = \mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(\Omega).$$

Similarly, the couple $A, \emptyset$ is independent. If the couple A, A is independent then $\mathbb{P}(A) = 0$ or 1, which follows from

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2.$$

Example. Let Ω be a unit square $[0,1]^2$, $\mathcal{F}$ consist of all measurable sets in Ω and $\mathbb{P} = \lambda_2$. Let I and J be two intervals in $[0,1]$, and consider the events $A = I \times [0,1]$ and $B = [0,1] \times J$. We claim that the events A and B are independent. Indeed, $A \cap B$ is the rectangle $I \times J$, and

$$\mathbb{P}(A \cap B) = \lambda_2(I \times J) = \ell(I)\,\ell(J).$$

Since

$$\mathbb{P}(A) = \ell(I) \quad \text{and} \quad \mathbb{P}(B) = \ell(J),$$

we conclude

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

We may choose I and J with the lengths p and q, respectively, for any prescribed couple $p, q \in (0,1)$. Hence, this example shows how to construct two independent event with given probabilities p, q.

In the same way, one can construct n independent events (with prescribed probabilities) in the probability space $\Omega = [0,1]^n$ where $\mathcal{F}$ is the family of all measurable sets in Ω and $\mathbb{P}$ is the n-dimensional Lebesgue measure λ_n. Indeed, let $I_1, ..., I_n$ be intervals in $[0,1]$ of the length $p_1, ..., p_n$. Consider the events

$$A_k = [0,1] \times ... \times I_k \times ... \times [0,1],$$

where all terms in the direct product are $[0,1]$ except for the k-th term I_k. Then A_k is a box in $\mathbb{R}^n$ and

$$\mathbb{P}(A_k) = \lambda_n(A_k) = \ell(I_k) = p_k.$$

For any sequence of distinct indices $i_1, i_2, ..., i_k$, the intersection

$$A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}$$

is a direct product of some intervals $[0,1]$ with the intervals $I_{i_1}, ..., I_{i_k}$ so that

$$\mathbb{P}(A_{i_1} \cap ... \cap A_{i_k}) = p_{i_1} ... p_{i_k} = \mathbb{P}(A_{i_1}) ... \mathbb{P}(A_{i_n}).$$

Hence, the sequence $\{A_k\}_{k=1}^n$ is independent.

It is natural to expect that independence is preserved by certain operations on events. For example, let A, B, C, D be independent events and let us ask whether the following couples of events are independent:

1. $A \cap B$ and $C \cap D$
2. $A \cup B$ and $C \cup D$
3. $E = (A \cap B) \cup (C \setminus A)$ and D.

It is easy to show that $A \cap B$ and $C \cap D$ are independent:

$$\begin{aligned}\mathbb{P}((A \cap B) \cap (C \cap D)) &= \mathbb{P}(A \cap B \cap C \cap D) \\ &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\mathbb{P}(D) = \mathbb{P}(A \cap B)\mathbb{P}(C \cap D).\end{aligned}$$

It is less obvious how to prove that $A \cup B$ and $C \cup D$ are independent. This will follow from the following more general statement.

Lemma 1.12 *Let $\mathcal{A} = \{A_i\}$ be an independent family of events. Suppose that a family $\mathcal{A}'$ of events is obtained from $\mathcal{A}$ by one of the following procedures:*

1. *Adding to $\mathcal{A}$ one of the events $\emptyset$ or Ω.*
2. *Replacing two events $A, B \in \mathcal{A}$ by one event $A \cap B$,*
3. *Replacing an event $A \in \mathcal{A}$ by its complement A^c.*

4. Replacing two events $A, B \in \mathcal{A}$ by one event $A \cup B$.

5. Replacing two events $A, B \in \mathcal{A}$ by one event $A \setminus B$.

Then the family $\mathcal{A}'$ is independent.

Applying the operation 4 twice, we obtain that if A, B, C, D are independent then $A \cup B$ and $C \cup D$ are independent. However, this lemma still does not answer why E and D are independent.

Proof. Each of the above procedures removes from $\mathcal{A}$ some of the events and adds a new event, say N. Denote by $\mathcal{A}''$ the family that remains after the removal, so that $\mathcal{A}'$ is obtained from $\mathcal{A}''$ by adding N. Clearly, removing events does not change the independence, so that $\mathcal{A}''$ is independent. Hence, in order to prove that $\mathcal{A}'$ is independent, it suffices to show that, for any events $A_1, A_2, ..., A_k$ from $\mathcal{A}''$ with distinct indices,

$$\mathbb{P}(N \cap A_1 \cap ... \cap A_k) = \mathbb{P}(N)\mathbb{P}(A_1)...\mathbb{P}(A_k). \tag{1.35}$$

Case 1. $N = \emptyset$ or Ω. The both sides of (1.35) vanish if $N = \emptyset$. If $N = \Omega$ then it can be removed from both sides of (1.35), so (1.35) follows from the independence of $A_1, A_2, ..., A_k$.

Case 2. $N = A \cap B$. We have

$$\begin{aligned}\mathbb{P}((A \cap B) \cap A_1 \cap A_2 \cap ... \cap A_k) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k) \\ &= \mathbb{P}(A \cap B)\mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k),\end{aligned}$$

which proves (1.35).

Case 3. $N = A^c$. Using the identity

$$A^c \cap B = B \setminus A = B \setminus (A \cap B)$$

and its consequence

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B),$$

we obtain, for $B = A_1 \cap A_2 \cap ... \cap A_k$, that

$$\begin{aligned}\mathbb{P}(A^c \cap A_1 \cap A_2 \cap ... \cap A_k) &= \mathbb{P}(A_1 \cap A_2 \cap ... \cap A_k) - \mathbb{P}(A \cap A_1 \cap A_2 \cap ... \cap A_k) \\ &= \mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k) - \mathbb{P}(A)\mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k) \\ &= (1 - \mathbb{P}(A))\,\mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k) \\ &= \mathbb{P}(A^c)\mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_k)\end{aligned}$$

Case 4. $N = A \cup B$. By the identity

$$A \cup B = (A^c \cap B^c)^c$$

this case amounts to 2 and 3.

Case 5. $N = A \setminus B$. By the identity

$$A \setminus B = A \cap B^c$$

this case amounts to 2 and 3. ■

Example. Using Lemma 1.12, we can justify the probabilistic argument, introduced in Section 1.3 in order to prove the inequality

$$(1-p^n)^m + (1-q^m)^n \geq 1, \tag{1.36}$$

where $p, q \in [0,1]$, $p+q=1$, and $n, m \in \mathbb{N}$. Indeed, for that argument we need nm independent events, each with the given probability p. As we have seen in an example above, there exists an arbitrarily long finite sequence of independent events, and with arbitrarily prescribed probabilities. So, choose nm independent events each with probability p and denote them by A_{ij} where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, so that they can be arranged in a $n \times m$ matrix $\{A_{ij}\}$. For any index j (which is a column index), consider an event

$$C_j = \bigcap_{i=1}^{n} A_{ij},$$

that is, the intersection of all events A_{ij} in the column j. Since A_{ij} are independent, we obtain

$$\mathbb{P}(C_j) = p^n \text{ and } \mathbb{P}(C_j^c) = 1 - p^n.$$

By Lemma 1.12, the events $\{C_j\}$ are independent and, hence, $\{C_j^c\}$ are also independent. Setting

$$C = \bigcap_{j=1}^{m} C_j^c$$

we obtain

$$\mathbb{P}(C) = (1-p^n)^m.$$

Similarly, considering the events

$$R_i = \bigcap_{j=1}^{m} A_{ij}^c$$

and

$$R = \bigcap_{i=1}^{n} R_i^c,$$

we obtain in the same way that

$$\mathbb{P}(R) = (1-q^m)^n.$$

Finally, we claim that

$$C \cup R = \Omega, \tag{1.37}$$

which would imply that

$$\mathbb{P}(C) + \mathbb{P}(R) \geq 1,$$

that is, (1.36).

To prove (1.37), observe that, by the definitions of R and C,

$$C = \bigcap_{j=1}^{m} C_j^c = \bigcap_{j=1}^{m} \left(\bigcap_{i=1}^{n} A_{ij} \right)^c = \bigcap_{j=1}^{m} \bigcup_{i=1}^{n} A_{ij}^c$$

and

$$R^c = \left(\bigcap_{i=1}^{n} R_i^c \right)^c = \bigcup_{i=1}^{n} R_i = \bigcup_{i=1}^{n} \bigcap_{j=1}^{m} A_{ij}^c.$$

Denoting by ω an arbitrary element of Ω, we see that if $\omega \in R^c$ then there is an index i_0 such that $\omega \in A_{i_0 j}^c$ for all j. This implies that $\omega \in \bigcup_{i=1}^n A_{ij}^c$ for any j, whence $\omega \in C$. Hence, $R^c \subset C$, which is equivalent to (1.37).

Let us give an alternative proof of (1.37) which is a rigorous version of the argument with the coin tossing. The result of nm trials with coin tossing was a sequence of letters H, T (heads and tails). Now we make a sequence of digits $1, 0$ instead as follows: for any $\omega \in \Omega$, set

$$M_{ij}(\omega) = \begin{cases} 1, & \omega \in A_{ij}, \\ 0, & \omega \notin A_{ij}. \end{cases}$$

Hence, a point $\omega \in \Omega$ represents the sequence of nm trials, and the results of the trials are registered in a *random* $n \times m$ matrix $M = \{M_{ij}\}$ (the word "random" simply means that M is a function of ω). For example, if

$$M(\omega) = \left(\begin{array}{|c|c|c|} \hline 1 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline \end{array}\right)$$

then $\omega \in A_{11}$, $\omega \in A_{13}$, $\omega \in A_{23}$ but $\omega \notin A_{12}$, $\omega \notin A_{21}$, $\omega \notin A_{22}$.

The fact that $\omega \in C_j$ means that all the entries in the column j of the matrix $M(\omega)$ are 1; $\omega \in C_j^c$ means that there is an entry 0 in the column j; $\omega \in C$ means that there is 0 in every column. In the same way, $\omega \in R$ means that there is 1 in every row. The desired identity $C \cup R = \Omega$ means that either there is 0 in every column or there is 1 in every row. Indeed, if the first event does not occur, that is, there is a column without 0, then this column contains only 1, for example, as here:

$$M(\omega) = \left(\begin{array}{|c|c|c|c|c|} \hline & & 1 & & \\ \hline & & 1 & & \\ \hline & & 1 & & \\ \hline \end{array}\right).$$

However, this means that there is 1 in every row, which proves that $C \cup R = \Omega$.

Although Lemma 1.12 was useful in the above argument, it is still not enough for other applications. For example, it does not imply that $E = (A \cap B) \cup (C \setminus A)$ and D are independent (if A, B, C, D are independent), since A is involved twice in the formula defining E. There is a general theorem which allows to handle all such cases. Before we state it, let us generalize the notion of independence as follows.

Definition. Let $\{\mathcal{A}_i\}$ be a sequence of families of events. We say that $\{\mathcal{A}_i\}$ is independent if any sequence $\{A_i\}$ such that $A_i \in \mathcal{A}_i$, is independent.

Example. Let $\{A_i\}$ be an independent sequence of events and consider the families

$$\mathcal{A}_i = \{\emptyset, A_i, A_i^c, \Omega\}.$$

Then the sequence $\{\mathcal{A}_i\}$ is independent, which follows from Lemma 1.12.

For any family $\mathcal{A}$ of subsets of Ω, denote by $R(\mathcal{A})$ the minimal algebra containing $\mathcal{A}$ and by $\Sigma(\mathcal{A})$ – the minimal σ-algebra containing $\mathcal{A}$ (we used to denote by $R(\mathcal{A})$ and $\Sigma(\mathcal{A})$ respectively the minimal ring and σ-ring containing $\mathcal{A}$, but in the present context we switch to algebras).

Theorem 1.13 *Suppose that $\{A_{ij}\}$ is an independent sequence of events parametrized by two indices i and j. Denote by $\mathcal{A}_i$ the family of all events A_{ij} with fixed i and arbitrary j. Then the sequence of algebras $\{R(\mathcal{A}_i)\}$ is independent. Moreover, also the sequence of σ-algebras $\{\Sigma(\mathcal{A}_i)\}$ is independent.*

For example, if the sequence $\{A_{ij}\}$ is represented by a matrix

$$\begin{pmatrix} A_{11} & A_{12} & ... \\ A_{21} & A_{22} & ... \\ ... & ... & ... \end{pmatrix}$$

then $\mathcal{A}_i$ is the family of all events in the row i, and the claim is that the algebras (resp., σ-algebras) generated by different rows, are independent. Clearly, the same applies to the columns.

Let us apply this theorem to the above example of $E = (A \cap B) \cup (C \setminus A)$ and D. Indeed, in the matrix

$$\begin{pmatrix} A & B & C \\ D & \Omega & \Omega \end{pmatrix}$$

all events are independent. Therefore, $R(A, B, C)$ and $R(D)$ are independent, and $E = (A \cap B) \cup (C \setminus A)$ and D are independent because $E \in R(A, B, C)$.

The proof of Theorem 1.13 is largely based on a powerful theorem of Dynkin, which explains how a given family $\mathcal{A}$ of subsets can be completed into algebra or σ-algebra. Before we state it, let us introduce the following notation.

Let Ω be any non-empty set (not necessarily a probability space) and $\mathcal{A}$ be a family of subsets of Ω. Let $*$ be an operation on subsets of Ω (such that union, intersection, etc). Then denote by $\mathcal{A}^*$ the minimal extension of $\mathcal{A}$, which is closed under the operation $*$. More precisely, consider all families of subsets of Ω, which contain $\mathcal{A}$ and which are closed under $*$ (the latter means that applying $*$ to the sets from such a family, we obtain again a set from that family). For example, the family 2^Ω of all subsets of Ω satisfies all these conditions. Then taking intersection of all such families, we obtain $\mathcal{A}^*$. If $*$ is an operation over a finite number of elements then $\mathcal{A}^*$ can be obtained from $\mathcal{A}$ by applying all possible finite sequences of operations $*$.

The operations to which we apply this notion are the following:

1. Intersection "$\cap$" that is $A, B \mapsto A \cap B$
2. The *monotone difference* "$-$" defined as follows: if $A \supset B$ then $A - B = A \setminus B$.
3. The *monotone limit* lim defined on monotone sequences $\{A_n\}_{n=1}^\infty$ as follows: if the sequence is increasing that is $A_n \subset A_{n+1}$ then

$$\lim A_n = \bigcup_{n=1}^\infty A_n$$

and if A_n is decreasing that is $A_n \supset A_{n+1}$ then

$$\lim A_n = \bigcap_{n=1}^\infty A_n.$$

Using the above notation, $\mathcal{A}^-$ is the minimal extension of $\mathcal{A}$ using the monotone difference, and $\mathcal{A}^{\lim}$ is the minimal extension of $\mathcal{A}$ using the monotone limit.

Theorem 1.14 *(*Theorem of Dynkin*) Let Ω be an arbitrary non-empty set and $\mathcal{A}$ be a family of subsets of Ω.*

(a) *If $\mathcal{A}$ contains Ω and is closed under $\cap$ then*

$$R(\mathcal{A}) = \mathcal{A}^-.$$

(b) *If $\mathcal{A}$ is algebra of subsets of Ω then*

$$\Sigma(\mathcal{A}) = \mathcal{A}^{\lim}.$$

As a consequence we see that if $\mathcal{A}$ is any family of subsets containing Ω then it can be extended to the minimal algebra $R(\mathcal{A})$ as follows: firstly, extend it using intersections so that the resulting family $A^\cap$ is closed under intersections; secondly, extend $A^\cap$ to the algebra $R(\mathcal{A}^\cap) = R(\mathcal{A})$ using the monotone difference (which requires part (a) of Theorem 1.14). Hence, we obtain the identity

$$R(\mathcal{A}) = (\mathcal{A}^\cap)^- . \tag{1.38}$$

Similarly, applying further part (b), we obtain

$$\Sigma(\mathcal{A}) = \left((\mathcal{A}^\cap)^-\right)^{\lim} . \tag{1.39}$$

The most non-trivial part is (1.38). Indeed, it says that any set that can be obtained from sets of $\mathcal{A}$ by a finite number of operations $\cap, \cup, \setminus$, can also be obtained by first applying a finite number of $\cap$ and then applying finite number of "$-$". This is not quite obvious even for the simplest case

$$\mathcal{A} = \{\Omega, A, B\}.$$

Indeed, (1.38) implies that the union $A \cup B$ can be obtained from Ω, A, B by applying first $\cap$ and then "$-$". However, Theorem 1.14 does not say how exactly one can do that. The answer in this particular case is

$$A \cup B = \Omega - (\Omega - A - (B - (A \cap B))).$$

Proof of Theorem 1.14. Let us prove the first part of the theorem. Assuming that $\mathcal{A}$ contains Ω and is closed under $\cap$, let us show that $\mathcal{A}^-$ is algebra, which will settle the claim. Indeed, as an algebra, $\mathcal{A}^-$ must contain $R(\mathcal{A})$ since $R(\mathcal{A})$ is the smallest algebra extension of $\mathcal{A}$. On the other hand, $R(\mathcal{A})$ is closed under "$-$" and contains $\mathcal{A}$; then it contains also $\mathcal{A}^-$, whence $R(\mathcal{A}) = \mathcal{A}^-$.

To show that $\mathcal{A}^-$ is an algebra, we need to verify that $\mathcal{A}^-$ contains $\emptyset, \Omega$ and is closed under $\cap, \cup, \setminus$. Obviously, $\Omega \in \mathcal{A}^-$ and $\emptyset = \Omega - \Omega \in \mathcal{A}^-$. Also, if $A \in \mathcal{A}^-$ then also $A^c \in \mathcal{A}^c$ because $A^c = \Omega - A$. It remains to show that $\mathcal{A}^-$ is closed under intersections, that is, if $A, B \in \mathcal{A}^-$ then $A \cap B \in \mathcal{A}^-$ (this will imply that also $A \cup B = (A^c \cap B^c)^c \in \mathcal{A}^-$ and $A \setminus B = A - (A \cap B) \in \mathcal{A}^-$).

Given a set $A \in \mathcal{A}^-$, call a set $B \subset \Omega$ *suitable* for A if $A \cap B \in \mathcal{A}^-$. Denote the family of suitable sets by S, that is,

$$S = \left\{ B \subset \Omega : A \cap B \in \mathcal{A}^- \right\}.$$

We need to prove that $S \supset \mathcal{A}^-$. Assume first that $A \in \mathcal{A}$. Then S contains $\mathcal{A}$ because $\mathcal{A}$ is closed under intersections. Let us verify that S is closed under monotone difference. Indeed, if $B_1 \supset B_2$ are suitable sets then

$$A \cap (B_1 - B_2) = (A \cap B_1) - (A \cap B_2) \in \mathcal{A}^-,$$

whence $B_1 \cap B_2 \in S$. Hence, the family S of all suitable sets contains $\mathcal{A}$ and is closed under "$-$", which implies that S contains $\mathcal{A}^-$ (because $\mathcal{A}^-$ is the minimal family with these properties). Hence, we have proved that

$$A \cap B \in \mathcal{A}^- \tag{1.40}$$

whenever $A \in \mathcal{A}$ and $B \in \mathcal{A}^-$. Switching A and B, we see that (1.40) holds also if $A \in \mathcal{A}^-$ and $B \in \mathcal{A}$.

Now we prove (1.40) for all $A, B \in \mathcal{A}^-$. Consider again the family S of suitable sets for A. As we have just shown, $S \supset \mathcal{A}$. Since S is closed under "$-$", we conclude that $S \supset \mathcal{A}^-$, which finishes the proof.

The second statement about σ-algebras can be obtained similarly. Using the method of suitable sets, one proves that $\mathcal{A}^{\lim}$ is an algebra, and the rest follows from the following observation.

Lemma 1.15 *If an algebra $\mathcal{A}$ is closed under monotone limits then it is a σ-algebra (conversely, any σ-algebra is an algebra and is closed under* $\lim$*).*

Indeed, it suffices to prove that if $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Indeed, setting $B_n = \bigcup_{i=1}^{n} A_i$, we have the identity

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n = \lim B_n.$$

Since $B_n \in \mathcal{A}$, it follows that also $\lim B_n \in \mathcal{A}$, which finishes the proof. ∎

Theorem 1.13 will be deduced from the following more general theorem.

Theorem 1.16 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{\mathcal{A}_i\}$ be a sequence of families of events such that each family $\mathcal{A}_i$ contain Ω and is closed under $\cap$. If the sequence $\{\mathcal{A}_i\}$ is independent then the sequence of the algebras $\{R(\mathcal{A}_i)\}$ is also independent. Moreover, the sequence of σ-algebras $\{\Sigma(\mathcal{A}_i)\}$ is independent, too.*

Proof of Theorem 1.16. In order to check the independence of families, one needs to test finite sequences of events chosen from those families. Hence, it suffice to restrict to the case when the number of families $\mathcal{A}_i$ is finite. So, assume that i runs over $1, 2, ..., n$. It suffices to show that the sequence $\{R(\mathcal{A}_1), \mathcal{A}_2, ..., \mathcal{A}_n\}$ is independent. If we know that then we can by induction replace $\mathcal{A}_2$ by $R(\mathcal{A}_2)$ etc. To show the independence of this sequence, we need to take arbitrary events $A_1 \in R(\mathcal{A}_1)$, $A_2 \in \mathcal{A}_2, ..., A_n \in \mathcal{A}_n$ and prove that

$$\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_n). \tag{1.41}$$

Indeed, by the definition of the independent events, one needs to check this property also for subsequences $\{A_{i_k}\}$ but this amounts to the full sequence $\{A_i\}$ by setting the missing events to be Ω.

Denote for simplicity $A = A_1$ and $B = A_2 \cap A_3 \cap ... \cap A_n$. Since $\{A_2, A_3, ..., A_n\}$ are independent, (1.41) amounts to

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{1.42}$$

In other words, we are left to prove that A and B are independent, for any $A \in R(\mathcal{A}_1)$ and B being an intersection of events from $\mathcal{A}_2, \mathcal{A}_3, ..., \mathcal{A}_n$.

Fix such B and call an event A *suitable* for B if (1.42) holds. We need to show that all events from $R(\mathcal{A}_1)$ are suitable. Observe that all events from $\mathcal{A}_1$ are suitable. Let us prove that the family of suitable sets is closed under the monotone difference "$-$". Indeed, if A and A' are suitable and $A \supset A'$ then

$$\begin{aligned}\mathbb{P}((A - A') \cap B) &= \mathbb{P}(A \cap B) - \mathbb{P}(A' \cap B)\\ &= \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A')\mathbb{P}(B)\\ &= \mathbb{P}(A - A')\mathbb{P}(B).\end{aligned}$$

Hence, we conclude that the family of all suitable sets contains the extension of $\mathcal{A}_1$ by "$-$", that is $\mathcal{A}_1^-$. By Theorem 1.14, $\mathcal{A}_1^- = R(\mathcal{A}_1)$. Hence, all events from $R(\mathcal{A}_1)$ are suitable, which was to be proved.

The independence of $\{\Sigma(\mathcal{A}_i)\}$ is treated in the same way by verifying that the equality (1.42) is preserved by monotone limits. ■

Proof of Theorem 1.13. Adding to each family $\mathcal{A}_i$ the event Ω does not change the independence, so we may assume $\Omega \in \mathcal{A}_i$. Also, if we extend $\mathcal{A}_i$ to $\mathcal{A}_i^\cap$ then $\{\mathcal{A}_i^\cap\}$ are also independent. Indeed, each event $B_i \in \mathcal{A}_i^\cap$ is an intersection of a finite number of events from $\mathcal{A}_i$, that is, has the form

$$B_i = A_{ij_1} \cap A_{ij_2} \cap ... \cap A_{ij_k}.$$

Hence, the sequence $\{B_i\}$ can be obtained by replacing in the double sequence $\{A_{ij}\}$ some elements by their intersections (and throwing away the rest), and the independence of $\{B_i\}$ follows by Lemma 1.12 from the independence of $\{A_{ij}\}$ across all i and j.

Therefore, the sequence $\{\mathcal{A}_i^\cap\}$ satisfies the hypotheses of Theorem 1.16, and the sequence $\{R(\mathcal{A}_i^\cap)\}$ (and $\{\Sigma(\mathcal{A}_i^\cap)\}$) is independent. Since $R(\mathcal{A}_i^\cap) = R(\mathcal{A}_i)$ and $\Sigma(\mathcal{A}_i^\cap) = \Sigma(\mathcal{A}_i)$, we obtain the claim. ■

2 Integration

In this Chapter, we define the general notion of the Lebesgue integral. Given a set M, a σ-algebra $\mathcal{M}$ and a measure μ on $\mathcal{M}$, we should be able to define the integral

$$\int_M f \, d\mu$$

of any function f on M of an appropriate class. Recall that if M is a bounded closed interval $[a,b] \subset \mathbb{R}$ then the integral $\int_a^b f(x)\,dx$ is defined for Riemann integrable function, in particular, for continuous functions on $[a,b]$, and is obtained as the limit of the Riemann integral sums

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

where $\{x_i\}_{i=0}^n$ is a partition of the interval $[a,b]$ and $\{\xi_i\}_{i=1}^n$ is a sequence of tags, that is, $\xi_i \in [x_{i-1}, x_i]$. One can try also in the general case arrange a partition of the set M into smaller sets E_1, E_2,... and define the integral sums

$$\sum_i f(\xi_i)\,\mu(E_i)$$

where $\xi_i \in E_i$. There are two problems here: in what sense to understand the limit of the integral sums and how to describe the class of functions for which the limit exists. Clearly, the partition $\{E_i\}$ should be chosen so that the values of f in E_i do not change much and can be approximated by $f(\xi_i)$. In the case of the Riemann integrals, this is achieved by choosing E_i to be small intervals and by using the continuity of f. In the general case, there is another approach, due to Lebesgue, whose main idea is to choose E_i depending of f, as follows:

$$E_i = \{x \in M : c_{i-1} < f(x) \leq c_i\}$$

where $\{c_i\}$ is an increasing sequence of reals. The values of f in E_i are all inside the interval $(c_{i-1}, c_i]$ so that they are all close to $f(\xi_i)$ provided the sequence $\{c_i\}$ is fine enough. This choice of E_i allows to avoid the use of the continuity of f but raised another question: in order to use $\mu(E_i)$, sets E_i must lie in the domain of the measure μ. For this reason, it is important to have measure defined on possibly larger domain. On the other hand, we will have to restrict functions f to those for which the sets of the form $\{a < f(x) \leq b\}$ are in the domain of μ. Functions with this property are called measurable. So, we first give a precise definition and discuss the properties of measurable functions.

2.1 Measurable functions

Let M be an arbitrary non-empty set and $\mathcal{M}$ be a σ-algebra of subsets of M.

Definition. We say that a set $A \subset M$ is measurable if $A \in \mathcal{M}$. We say that a function $f : M \to \mathbb{R}$ is measurable if, for any $c \in \mathbb{R}$, the set $\{x \in M : f(x) \leq c\}$ is measurable, that is, belongs to $\mathcal{M}$.

Of course, the measurability of sets and functions depends on the choice of the σ-algebra $\mathcal{M}$. For example, in $\mathbb{R}^n$ we distinguish Lebesgue measurable sets and functions when $\mathcal{M} = \mathcal{M}_n$ (they are frequently called simply "measurable"), and the Borel measurable sets and functions when $\mathcal{M} = \mathcal{B}_n$ (they are normally called "Borel sets" and "Borel functions" avoiding the word "measurable").

The measurability of a function f can be also restated as follows. Since

$$\{f(x) \le c\} = f^{-1}(-\infty, c],$$

we can say that a function f is measurable if, for any $c \in \mathbb{R}$, the set $f^{-1}(-\infty, c]$ is measurable. Let us refer to the intervals of the form $(-\infty, c]$ as *special intervals*. Then we can say that a mapping $f : M \to \mathbb{R}$ is measurable if the preimage of any special interval is a measurable set.

Example. Let A be an arbitrary subset of M. Define the *indicator function* 1_A on M by

$$1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

We claim that the set A is measurable if and only if the function 1_A is measurable. Indeed, the set $\{f(x) \le c\}$ can be described as follows:

$$\{f(x) \le c\} = \begin{cases} \emptyset, & c < 0, \\ A^c, & 0 \le c < 1, \\ M, & c \ge 1. \end{cases}$$

The sets $\emptyset$ and M are always measurable, and A^c is measurable if and only if A is measurable, whence the claim follows.

Example. Let $M = \mathbb{R}^n$ and $\mathcal{M} = \mathcal{B}_n$. Let $f(x)$ be a continuous function from $\mathbb{R}^n$ to $\mathbb{R}$. Then the set $f^{-1}(-\infty, c]$ is a closed subset of $\mathbb{R}^n$ as the preimage of a closed set $(-\infty, c]$ in $\mathbb{R}$. Since the closed sets are Borel, we conclude that any continuous function f is a Borel function.

Definition. A mapping $f : M \to \mathbb{R}^n$ is called *measurable* if, for all $c_1, c_2, ..., c_n \in \mathbb{R}$, the set

$$\{x \in M : f_1(x) \le c_1, f_2(x) \le c_2, ..., f_n(x) \le c_n\}$$

is measurable. Here f_k is the k-th component of f.

In other words, consider an infinite box in $\mathbb{R}^n$ of the form:

$$B = (-\infty, c_1] \times (-\infty, c_2] \times ... \times (-\infty, c_n],$$

and call it a *special box*. Then a mapping $f : M \to \mathbb{R}^n$ is measurable if, for any special box B, the preimage $f^{-1}(B)$ is a measurable subset of M.

Lemma 2.1 *If a mapping $f : M \to \mathbb{R}^n$ is measurable then, for any Borel set $A \subset \mathbb{R}^n$, the preimage $f^{-1}(A)$ is a measurable set.*

Proof. Let $\mathcal{A}$ be the family of all sets $A \subset \mathbb{R}^n$ such that $f^{-1}(A)$ is measurable. By hypothesis, $\mathcal{A}$ contains all special boxes. Let us prove that $\mathcal{A}$ is a σ-algebra (this follows also from Exercise 4 since in the notation of that exercise $\mathcal{A} = f(\mathcal{M})$). If $A, B \in \mathcal{A}$ then $f^{-1}(A)$ and $f^{-1}(B)$ are measurable whence

$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B) \in \mathcal{M},$$

whence $A \setminus B \in \mathcal{A}$. Also, if $\{A_k\}_{k=1}^N$ is a finite or countable sequence of sets from $\mathcal{A}$ then $f^{-1}(A_k)$ is measurable for all k whence

$$f^{-1}\left(\bigcup_{k=1}^N A_k\right) = \bigcup_{k=1}^N f^{-1}(A_k) \in \mathcal{M},$$

which implies that $\bigcup_{k=1}^N A_k \in \mathcal{A}$. Finally, $\mathcal{A}$ contains $\mathbb{R}$ because $\mathbb{R}$ is the countable union of the intervals $(-\infty, n]$ where $n \in \mathbb{N}$, and $\mathcal{A}$ contains $\emptyset = \mathbb{R} \setminus \mathbb{R}$.

Hence, $\mathcal{A}$ is a σ-algebra containing all special boxes. It remains to show that $\mathcal{A}$ contains all the boxes in $\mathbb{R}^n$, which will imply that $\mathcal{A}$ contains all Borel sets in $\mathbb{R}^n$. In fact, it suffices to show that any box in $\mathbb{R}^n$ can be obtained from special boxes by a countable sequence of set-theoretic operations.

Assume first $n = 1$ and consider different types of intervals. If $A = (-\infty, a]$ then $A \in \mathcal{A}$ by hypothesis.

Let $A = (a, b]$ where $a < b$. Then $A = (-\infty, b] \setminus (-\infty, a]$, which proves that A belongs to $\mathcal{A}$ as the difference of two special intervals.

Let $A = (a, b)$ where $a < b$ and $a, b \in \overline{\mathbb{R}}$. Consider a strictly increasing sequence $\{b_k\}_{k=1}^\infty$ such that $b_k \to b$ as $k \to \infty$. Then the intervals $(a, b_k]$ belong to $\mathcal{A}$ by the previous argument, and the obvious identity

$$A = (a, b) = \bigcup_{k=1}^\infty (a, b_k]$$

implies that $A \in \mathcal{A}$.

Let $A = [a, b)$. Consider a strictly increasing sequence $\{a_k\}_{k=1}^\infty$ such that $a_k \to a$ as $k \to \infty$. Then $(a_k, b) \in \mathcal{A}$ by the previous argument, and the set

$$A = [a, b) = \bigcap_{k=1}^\infty (a_k, b)$$

is also in $\mathcal{A}$.

Finally, let $A = [a, b]$. Observing that $A = \mathbb{R} \setminus (-\infty, a) \setminus (b, +\infty)$ where all the terms belong to $\mathcal{A}$, we conclude that $A \in \mathcal{A}$.

Consider now that general case $n > 1$. We are given that $\mathcal{A}$ contains all boxes of the form

$$B = I_1 \times I_2 \times ... \times I_n$$

where I_k are special intervals, and we need to prove that $\mathcal{A}$ contains all boxes of this form with arbitrary intervals I_k. If I_1 is an arbitrary interval and $I_2, ..., I_n$ are special intervals then one shows that $B \in$A using the same argument as in the case $n = 1$ since I_1 can be obtained from the special intervals by a countable sequence of set-theoretic operations, and the same sequence of operations can be applied to the product $I_1 \times I_2 \times ... \times I_n$. Now let I_1 and I_2 be arbitrary intervals and $I_3, ..., I_n$ be special. We know that if I_2 is special then $B \in \mathcal{A}$. Obtaining an arbitrary interval I_2 from special intervals by a countable sequence of operations, we obtain that $B \in \mathcal{A}$ also for arbitrary I_1 and I_2. Continuing the same way, we obtain that $B \in \mathcal{A}$ if I_1, I_2, I_3 are arbitrary intervals while $I_4, ..., I_n$ are special, etc. Finally, we allow all the intervals $I_1, ..., I_n$ to be arbitrary. ■

Example. If $f : M \to \mathbb{R}$ is a measurable function then the set

$$\{x \in M : f(x) \text{ is irrational}\}$$

is measurable, because this set coincides with $f^{-1}(\mathbb{Q}^c)$, and $\mathbb{Q}^c$ is Borel since $\mathbb{Q}$ is Borel as a countable set.

Theorem 2.2 *Let $f_1, ..., f_n$ be measurable functions from M to $\mathbb{R}$ and let Φ be a Borel function from $\mathbb{R}^n$ to $\mathbb{R}$. Then the function*

$$F = \Phi(f_1, ..., f_n)$$

is measurable.

In other words, the composition of a Borel function with measurable functions is measurable. Note that the composition of two measurable functions may be not measurable.

Example. It follows from Theorem 2.2 that if f_1 and f_2 are two measurable functions on M then their sum $f_1 + f_2$ is also measurable. Indeed, consider the function $\Phi(x_1, x_2) = x_1 + x_2$ in $\mathbb{R}^2$, which is continuous and, hence, is Borel. Then $f_1 + f_2 = \Phi(f_1, f_2)$ and this function is measurable by Theorem 2.2. A direct proof by definition may be difficult: the fact that the set $\{f_1 + f_2 \leq c\}$ is measurable, is not immediately clear how to reduce this set to the measurable sets $\{f_1 \leq a\}$ and $\{f_1 \leq b\}$.

In the same way, the functions $f_1 f_2$, f_1/f_2 (provided $f_2 \neq 0$) are measurable. Also, the functions $\max(f_1, f_2)$ and $\min(f_1, f_2)$ are measurable, etc.

Proof of Theorem 2.2. Consider the mapping $f : M \to \mathbb{R}^n$ whose components are f_k. This mapping is measurable because for any $c \in \mathbb{R}^n$, the set

$$\{x \in M : f_1(x) \leq c_1, ..., f_n(x) \leq c_n\} = \{f_1(x) \leq c_1\} \cap \{f_2(x) \leq c_2\} \cap ... \cap \{f_n(x) \leq c_n\}$$

is measurable as the intersection of measurable sets. Let us show that $F^{-1}(I)$ is a measurable set for any special interval I, which will prove that F is measurable. Indeed, since $F(x) = \Phi(f(x))$, we obtain that

$$\begin{aligned} F^{-1}(I) &= \{x \in M : F(x) \in I\} \\ &= \{x \in M : \Phi(f(x)) \in I\} \\ &= \left\{x \in M : f(x) \in \Phi^{-1}(I)\right\} \\ &= f^{-1}\left(\Phi^{-1}(I)\right). \end{aligned}$$

Since $\Phi^{-1}(I)$ is a Borel set, we obtain by Lemma 2.1 that $f^{-1}(\Phi^{-1}(I))$ is measurable, which proves that $F^{-1}(I)$ is measurable. ■

Example. If $A_1, A_2, ..., A_n$ is a finite sequence of measurable sets then the function

$$f = c_1 1_{A_1} + c_1 1_{A_2} + ... + c_n 1_{A_n}$$

is measurable (where c_i are constants). Indeed, each of the function 1_{A_i} is measurable, whence the claim following upon application of Theorem 2.2 with the function

$$\Phi(x) = c_1 x_1 + ... + c_n x_n.$$

2.2 Sequences of measurable functions

As before, let M be a non-empty set and $\mathcal{M}$ be a σ-algebra on M.

Definition. We say that a sequence $\{f_n\}_{n=1}^{\infty}$ of functions on M converges to a function f *pointwise* and write $f_n \to f$ if $f_n(x) \to f(x)$ as $n \to \infty$ for any $x \in M$.

Theorem 2.3 *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions that converges pointwise to a function f. Then f is measurable, too.*

Proof. Fix some real c. Using the definition of a limit and the hypothesis that $f_n(x) \to f(x)$ as $n \to \infty$, we obtain that the inequality $f(x) \leq c$ is equivalent to the following condition: for any $k \in \mathbb{N}$ there is $m \in \mathbb{N}$ such that, for all $n \geq m$,

$$f_n(x) < c + \frac{1}{k}.$$

This can be written in the form of set-theoretic inclusion as follows:

$$\{f(x) \leq c\} = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \left\{ f_n(x) < c + \frac{1}{k} \right\}.$$

(Indeed, any logical condition "for any ..." transforms to the intersection of the corresponding sets, and the condition "there is ..." transforms to the union.)

Since the set $\{f_n < c + 1/k\}$ is measurable and the measurability is preserved by countable unions and intersections, we conclude that the set $\{f(x) \leq c\}$ is measurable, which finishes the proof. ■

Corollary. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of Lebesgue measurable (or Borel) functions on $\mathbb{R}^n$ that converges pointwise to a function f. Then f is Lebesgue measurable (resp., Borel) as well.*

Proof. Indeed, this is a particular case of Theorem 2.3 with $\mathcal{M} = \mathcal{M}_n$ (for Lebesgue measurable functions) and with $\mathcal{M} = \mathcal{B}_n$ (for Borel functions). ■

Example. Let us give an alternative proof of the fact that any continuous function f on $\mathbb{R}$ is Borel. Indeed, consider first a function of the form

$$g(x) = \sum_{k=1}^{\infty} c_k 1_{I_k}(x)$$

where $\{I_k\}$ is a disjoint sequence of intervals. Then

$$g(x) = \lim_{N\to\infty} \sum_{k=1}^{N} c_k 1_{I_k}(x)$$

and, hence, $g(x)$ is Borel as the limit of Borel functions. Now fix some $n \in \mathbb{N}$ and consider the following sequence of intervals:

$$I_k = [\frac{k}{n}, \frac{k+1}{n}) \quad \text{where} \quad k \in \mathbb{Z},$$

so that $\mathbb{R} = \bigsqcup_{k \in \mathbb{Z}} I_k$. Also, consider the function

$$g_n(x) = \sum_{k \in \mathbb{Z}} f\left(\frac{k}{n}\right) 1_{I_k}(x).$$

By the continuity of f, for any $x \in \mathbb{R}$, we have $g_n(x) \to f(x)$ as $n \to \infty$. Hence, we conclude that f is Borel as the pointwise limit of Borel functions.

So far we have only assumed that $\mathcal{M}$ is a σ-algebra of subsets of M. Now assume that there is also a measure μ on $\mathcal{M}$.

Definition. We say that a measure μ is *complete* if any subset of a set of measure 0 belongs to $\mathcal{M}$. That is, if $A \in \mathcal{M}$ and $\mu(A) = 0$ then every subset A' of A is also in $\mathcal{M}$ (and, hence, $\mu(A') = 0$).

As we know, if μ is a σ-finite measure initially defined on some ring R then by the Carathéodory extension theorems (Theorems 1.7 and 1.8), μ can be extended to a measure on a σ-algebra $\mathcal{M}$ of measurable sets, which contains all null sets. It follows that if $\mu(A) = 0$ then A is a null set and, by Theorem 1.10, any subset of A is also a null set and, hence, is measurable. Therefore, any measure μ that is constructed by the Carathéodory extension theorems, is automatically compete. In particular, the Lebesgue measure λ_n with the domain $\mathcal{M}_n$ is complete.

In general, a measure does not have to be complete, It is possible to show that the Lebesgue measure λ_n restricted to $\mathcal{B}_n$ is no longer complete (this means, that if$A$ is a Borel set of measure 0 in $\mathbb{R}^n$ then not necessarily any subset of it is Borel). On the other hand, every measure can be completed by adding to its domain the null sets – see Exercise 36.

Assume in the sequence that μ is a complete measure defined on a σ-algebra $\mathcal{M}$ of subsets of M. Then we use the term a *null set* as synonymous for a set of measure 0.

Definition. We say that two functions $f, g : M \to \mathbb{R}$ are equal *almost everywhere* and write $f = g$ a.e. if the set $\{x \in M : f(x) \neq g(x)\}$ has measure 0. In other words, there is a set N of measure 0 such that $f = g$ on $M \setminus N$.

More generally, the term "almost everywhere" is used to indicate that some property holds for all points $x \in M \setminus N$ where $\mu(N) = 0$.

Claim 1. *The relation* $f = g$ a.e.*is an equivalence relation.*

Proof. We need to prove three properties that characterize the equivalence relations:

1. $f = f$ a.e.. Indeed, the set $\{f(x) \neq f(x)\}$ is empty and, hence, is a null set.

2. $f = g$ a.e.is equivalent to $g = f$ a.e., which is obvious.

3. $f = g$ a.e.and $g = h$ a.e. imply $f = h$ a.e. Indeed, we have

$$\{f \neq h\} \subset \{f \neq g\} \cup \{h \neq g\}$$

whence we conclude that the set $\{f(x) \neq g(x)\}$ is a null set as a subset of the union of two null sets.

∎

Claim 2 *If f is measurable and $g = f$* a.e. *then g is also measurable.*

Proof. Fix some real c and prove that the set $\{g \leq c\}$ is measurable. Observe that

$$N := \{f \leq c\} \vartriangle \{g \leq c\} \subset \{f \neq g\}.$$

Indeed, $x \in N$ if x belongs to exactly to one of the sets $\{f \leq c\}$ and $\{g \leq c\}$. For example, x belongs to the first one and does not belong to the second one, then $f(x) \leq c$ and $g(x) > c$ whence $f(x) \neq g(x)$. Since $\{f \neq g\}$ is a null set, the set N is also a null set. Then we have

$$\{g \leq c\} = \{f \leq c\} \vartriangle N,$$

which implies that $\{g \leq c\}$ is measurable. ■

Definition. We say that a sequence of functions f_n on M converges to a function f *almost everywhere* and write $f_n \to f$ a.e. if the set $\{x \in M : f_n(x) \not\to f(x)\}$ is a null set. In other words, there is a set N of measure 0 such that $f_n \to f$ pointwise on $M \setminus N$.

Theorem 2.4 *If $\{f_n\}$ is a sequence of measurable functions and $f_n \to f$* a.e. *then f is also a measurable function.*

Proof. Consider the set

$$N = \{x \in M : f_n(x) \not\to f(x)\},$$

which has measure 0. Redefine $f_n(x)$ for $x \in N$ by setting $f_n(x) = 0$. Since we have changed f_n on a null set, the new function f_n is also measurable. Then the new sequence $\{f_n(x)\}$ converges for all $x \in M$, because $f_n(x) \to f(x)$ for all $x \in M \setminus N$ by hypothesis, and $f_n(x) \to 0$ for all $x \in N$ by construction. By Theorem 2.3, the limit function is measurable, and since f is equal to the limit function almost everywhere, f is also measurable. ■

We say a sequence of functions $\{f_n\}$ on a set M converges to a function f *uniformly* on M and write $f_n \rightrightarrows f$ on M if

$$\sup_{x \in M} |f_n(x) - f(x)| \to 0 \text{ as } n \to \infty.$$

We have obviously the following relations between the convergences:

$$f \rightrightarrows f \implies f_n \to f \text{ pointwise} \implies f_n \to f \text{ a.e.} \tag{2.1}$$

In general the converse for the both implications is not true. For example, let us show that the pointwise convergence does not imply the uniform convergence if the set M is infinite. Indeed, let $\{x_k\}_{k=1}^{\infty}$ be a sequence of distinct points in M and set $f_k = 1_{\{x_k\}}$. Then, for any point $x \in M$, $f_k(x) = 0$ for large enough k, which implies that $f_k(x) \to 0$ pointwise. On the other hand, $\sup |f_k| = 1$ so that $f_k \not\rightrightarrows 0$.

For the second example, let $\{f_k\}$ be any sequence of functions that converges pointwise to f. Define $\widetilde{f}$ as an arbitrary modification of f on a non-empty set of measure 0. Then still $f \to \widetilde{f}$ a.e. while f does not converge to $\widetilde{f}$ pointwise.

Surprisingly enough, the convergence a.e. still implies the uniform convergence but on a smaller set, as is stated in the following theorem.

Theorem 2.5 (Theorem of Egorov) *Let μ be a complete finite measure on a σ-algebra $\mathcal{M}$ on a set M. Let $\{f_n\}$ be a sequence of measurable functions and assume that $f_n \to f$* a.e. *on M. Then, for any $\varepsilon > 0$, there is a set $M_\varepsilon \subset M$ such that:*

1. *$\mu(M \setminus M_\varepsilon) < \varepsilon$*
2. *$f_n \rightrightarrows f$ on M_ε.*

In other words, by removing a set $M \setminus M_\varepsilon$ is measure smaller than ε, one can achieve that on the remaining set M_ε the convergence is uniform.

Proof. The condition $f_n \rightrightarrows f$ on M_ε (where M_ε is yet to be defined) means that for any $m \in \mathbb{N}$ there is $n = n(m)$ such that for all $k \geq n$

$$\sup_{M_\varepsilon} |f_k - f| < \frac{1}{m}.$$

Hence, for any $x \in M_\varepsilon$,

$$\text{for any } m \geq 1 \text{ for any } k \geq n(m) \ \ |f_k(x) - f(x)| < \frac{1}{m},$$

which implies that

$$M_\varepsilon \subset \bigcap_{m=1}^{\infty} \bigcap_{k \geq n(m)} \left\{ x \in M : |f_k(x) - f(x)| < \frac{1}{m} \right\}.$$

Now we can define M_ε to be the right hand side of this relation, but first we need to define $n(m)$.

For any couple of positive integers n, m, consider the set

$$A_{m,n} = \left\{ x \in M : |f_k(x) - f(x)| < \frac{1}{m} \text{ for all } k \geq n \right\}.$$

This can also be written in the form

$$A_{m,n} = \bigcap_{k \geq n} \left\{ x \in M : |f_k(x) - f(x)| < \frac{1}{m} \right\}.$$

By Theorem 2.4, function f is measurable, by Theorem 2.2 the function $|f_n - f|$ is measurable, which implies that A_{mn} is measurable as a countable intersection of measurable sets.

Observe that, for any fixed m, the set $A_{m,n}$ increases with n. Set

$$A_m = \bigcup_{n=1}^{\infty} A_{m,n} = \lim_{n \to \infty} A_{m,n},$$

that is, A_m is the monotone limit of $A_{m,n}$. Then, by Exercise 9, we have

$$\mu(A_m) = \lim_{n \to \infty} \mu(A_{m,n})$$

whence

$$\lim_{n \to \infty} \mu(A_m \setminus A_{m,n}) = 0 \tag{2.2}$$

(to pass to (2.2), we use that $\mu(A_m) < \infty$, which is true by the hypothesis of the finiteness of measure μ).

Claim *For any m, we have*

$$\mu(M \setminus A_m) = 0. \tag{2.3}$$

Indeed, by definition,

$$A_m = \bigcup_{n=1}^{\infty} A_{m,n} = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \left\{ x \in M : |f_k(x) - f(x)| < \frac{1}{m} \right\},$$

which means that $x \in A_m$ if and only if there is n such that for all $k \geq n$,

$$|f_k(x) - f(x)| < \frac{1}{m}.$$

In particular, if $f_k(x) \to f(x)$ for some point x then this condition is satisfied so that this point x is in A_m. By hypothesis, $f_n(x) \to f(x)$ for almost all x, which implies that $\mu(M \setminus A_m) = 0$.

It follows from (2.2) and (2.3) that

$$\lim_{n \to \infty} \mu(M \setminus A_{m,n}) = 0,$$

which implies that there is $n = n(m)$ such that

$$\mu\left(M \setminus A_{m,n(m)}\right) < \frac{\varepsilon}{2^m}. \tag{2.4}$$

Set

$$M_\varepsilon = \bigcap_{m=1}^{\infty} A_{m,n(m)} = \bigcap_{m=1}^{\infty} \bigcap_{k \geq n(m)} \left\{ x \in M : |f_k(x) - f(x)| < \frac{1}{m} \right\} \tag{2.5}$$

and show that the set M_ε satisfies the required conditions.

1. *Proof of* $\mu(M \setminus M_\varepsilon) < \varepsilon$. Observe that by (2.5)

$$\begin{aligned} \mu(M \setminus M_\varepsilon) &= \mu\left(\left(\bigcap_{m=1}^{\infty} A_{m,n(m)} \right)^c \right) = \mu\left(\bigcup_{m=1}^{\infty} A_{n,m(n)}^c \right) \\ &\leq \sum_{m=1}^{\infty} \mu\left(M \setminus A_{n,m(n)}\right) \\ &< \sum_{m=1}^{\infty} \frac{\varepsilon}{2^m} = \varepsilon. \end{aligned}$$

Here we have used the subadditivity of measure and (2.4).

2. *Proof of $f_n \rightrightarrows f$ on M_ε.* Indeed, for any $x \in M$, we have by (2.5) that, for any $m \geq 1$ and any $k \geq n(m)$,

$$|f_k(x) - f(x)| < \frac{1}{m}.$$

Hence, taking sup in $x \in M_\varepsilon$, we obtain that also

$$\sup_{M_\varepsilon} |f_k - f| \leq \frac{1}{m},$$

which implies that

$$\sup_{M_\varepsilon} |f_k - f| \to 0$$

and, hence, $f_k \rightrightarrows f$ on M_ε. ∎

2.3 The Lebesgue integral for finite measures

Let M be an arbitrary set, $\mathcal{M}$ be a σ-algebra on M and μ be a complete measure on $\mathcal{M}$. We are going to define the notion of the integral $\int_M f d\mu$ for an appropriate class of functions f. We will first do this for a finite measure, that is, assuming that $\mu(M) < \infty$, and then extend to the σ-finite measure.

Hence, assume here that μ is finite.

2.3.1 Simple functions

Definition. A function $f : M \to \mathbb{R}$ is called *simple* if it is measurable and the set of its values is at most countable.

Let $\{a_k\}$ be the sequence of distinct values of a simple function f. Consider the sets

$$A_k = \{x \in M : f(x) = a_k\} \tag{2.6}$$

which are measurable, and observe that

$$M = \bigsqcup_k A_k. \tag{2.7}$$

Clearly, we have the identity

$$f(x) = \sum_k a_k 1_{A_k}(x) \tag{2.8}$$

for all $x \in M$. Note that any sequence $\{A_k\}$ of disjoint measurable sets such that (2.7) and any sequence of distinct reals $\{a_k\}$ determine by (2.8) a function $f(x)$ that satisfies also (2.6), which means that all simple functions have the form (2.8).

Definition. If $f \geq 0$ is a simple function then define the Lebesgue integral $\int_M f d\mu$ by

$$\int_M f d\mu := \sum_k a_k \mu(A_k). \tag{2.9}$$

The value in the right hand side of (2.9) is always defined as the sum of a non-negative series, and can be either a non-negative real number or infinity. Note also that in order to be able to define $\mu(A_k)$, sets A_k must be measurable, which is equivalent to the measurability of f.

For example, if $f \equiv C$ for some constant C then we can write $f = C1_M$ and by (2.9)

$$\int_M f d\mu = C\mu(M).$$

The expression $\int_M f d\mu$ has the full title "the integral of f over M against measure μ". The notation $\int_M f d\mu$ should be understood as a whole, since we do not define what $d\mu$ means. This notation is traditionally used and has certain advantages. A modern shorter notation for the integral is $\mu(f)$, which reflects the idea that measure μ induces a functional on functions, which is exactly the integral. We have defined so far this functional for simple functions and then will extend it to a more general class. However, first we prove some property of the integral of simple functions.

Lemma 2.6 *(a) Let $M = \bigsqcup_k B_k$ where $\{B_k\}$ is a finite or countable sequence of measurable sets. Define a function f by*

$$f = \sum_k b_k 1_{B_k}$$

where $\{b_k\}$ is a sequence of non-negative reals, not necessarily distinct. Then

$$\int_M f d\mu = \sum_k b_k \mu(B_k).$$

(b) If f is a non-negative simple function then, for any real $c \geq 0$, cf is also non-negative real and

$$\int_M cf d\mu = c \int_M f d\mu$$

(if $c = 0$ and $\int_M f d\mu = +\infty$ and then we use the convention $0 \cdot \infty = 0$).

(c) If f, g are non-negative simple functions then $f + g$ is also simple and

$$\int_M (f+g)\, d\mu = \int_M f d\mu + \int_M g d\mu.$$

(d) If f, g are simple functions and $0 \leq f \leq g$ then

$$\int_M f d\mu \leq \int_M g d\mu.$$

Proof. (a) Let $\{a_j\}$ be the sequence of all distinct values in $\{b_k\}$, that is, $\{a_j\}$ is the sequence of all distinct values of f. Set

$$A_j = \{x \in M : f(x) = a_j\}.$$

Then

$$A_j = \bigsqcup_{\{k: b_k = a_j\}} B_k$$

and

$$\mu(A_j) = \sum_{\{k: b_k = a_j\}} \mu(B_k),$$

whence

$$\int_M f d\mu = \sum_j a_j \mu(A_j) = \sum_j a_j \sum_{\{k: b_k = a_j\}} \mu(B_k) = \sum_j \sum_{\{k: b_k = a_j\}} b_k \mu(B_k) = \sum_k b_k \mu(B_k).$$

(b) Let $f = \sum_k a_k 1_{A_k}$ where $\bigsqcup_k A_k = M$. Then

$$cf = \sum_k c a_k 1_{A_k}$$

whence by (a)

$$\int_M cf d\mu = \sum_k c a_k \mu(A_k) = c \sum_k a_k \mu(A_k).$$

(c) Let $f = \sum_k a_k 1_{A_k}$ where $\bigsqcup_k A_k = M$ and $g = \sum_j b_j 1_{B_j}$ where $\bigsqcup_j B_j = M$. Then

$$M = \bigsqcup_{k,j} (A_k \cap B_j)$$

and on the set $A_k \cap B_j$ we have $f = a_k$ and $g = b_j$ so that $f + g = a_k + b_j$. Hence, $f + g$ is a simple function, and by part (a) we obtain

$$\int_M (f+g)\, d\mu = \sum_{k,j} (a_k + b_j)\, \mu (A_k \cap B_j)\,.$$

Also, applying the same to functions f and g, we have

$$\int_M f d\mu = \sum_{k,j} a_k \mu (A_k \cap B_j)$$

and

$$\int_M g d\mu = \sum_{k,j} b_j \mu (A_k \cap B_j)\,,$$

whence the claim follows.

(d) Clearly, $g - f$ is a non-negative simple functions so that by (c)

$$\int_M g d\mu = \int_M (g - f)\, d\mu + \int_M f d\mu \geq \int_M f d\mu.$$

■

2.3.2 Positive measurable functions

Definition. Let $f \geq 0$ be any measurable function on M. The Lebesgue integral of f is defined by

$$\int_M f d\mu = \lim_{n\to\infty} \int_M f_n d\mu$$

where $\{f_n\}$ is any sequence of non-negative simple functions such that $f_n \rightrightarrows f$ on M as $n \to \infty$.

To justify this definition, we prove the following statement.

Lemma 2.7 *For any non-negative measurable functions f, there is a sequence of non-negative simple functions $\{f_n\}$ such that $f_n \rightrightarrows f$ on M. Moreover, for any such sequence the limit*

$$\lim_{n\to\infty} \int_M f_n d\mu$$

exists and does not depend on the choice of the sequence $\{f_n\}$ as long as $f_n \rightrightarrows f$ on M.

Proof. Fix the index n and, for any non-negative integer k, consider the set

$$A_{k,n} = \left\{ x \in M : \frac{k}{n} \leq f(x) < \frac{k+1}{n} \right\}.$$

Clearly, $M = \bigsqcup_{k=0}^{\infty} A_{k,n}$. Define function f_n by

$$f_n = \sum_k \frac{k}{n} 1_{A_{k,n}},$$

that is, $f_n = \frac{k}{n}$ on $A_{k,n}$. Then f_n is a non-negative simple function and, on a set $A_{k,n}$, we have

$$0 \le f - f_n < \frac{k+1}{n} - \frac{k}{n} = \frac{1}{n}$$

so that

$$\sup_M |f - f_n| \le \frac{1}{n}.$$

It follows that $f_n \rightrightarrows f$ on M.

Let now $\{f_n\}$ be any sequence of non-negative simple functions such that $f_n \rightrightarrows f$. Let us show that $\lim_{n\to\infty} \int_M f_n d\mu$ exists. The condition $f_n \rightrightarrows f$ on M implies that

$$\sup_M |f_n - f_m| \to 0 \text{ as } n, m \to \infty.$$

Assume that n, m are so big that $C := \sup_M |f_n - f_m|$ is finite. Writing

$$f_m \le f_n + C$$

and noting that all the functions f_m, f_n, C are simple, we obtain by Lemma 2.6

$$\begin{aligned} \int_M f_m d\mu &\le \int_M f_n d\mu + \int_M C d\mu \\ &= \int_M f_n d\mu + \sup_M |f_n - f_m| \, \mu(M). \end{aligned}$$

If $\int_M f_m d\mu = +\infty$ for some m, then implies that that $\int_M f_n d\mu = +\infty$ for all large enough n, whence it follows that $\lim_{n\to\infty} \int_M f_n d\mu = +\infty$. If $\int_M f_m d\mu < \infty$ for all large enough m, then it follows that

$$\left| \int_M f_m d\mu - \int_M f_n d\mu \right| \le \sup_M |f_n - f_m| \, \mu(M)$$

which implies that the numerical sequence

$$\left\{ \int_M f_n d\mu \right\}$$

is Cauchy and, hence, has a limit.

Let now $\{f_n\}$ and $\{g_n\}$ be two sequences of non-negative simple functions such that $f_n \rightrightarrows f$ and $g_n \rightrightarrows f$. Let us show that

$$\lim_{n\to\infty} \int_M f_n d\mu = \lim_{n\to\infty} \int_M g_n d\mu. \tag{2.10}$$

Indeed, consider a mixed sequence $\{f_1, g_1, f_2, g_2, ...\}$. Obviously, this sequence converges uniformly to f. Hence, by the previous part of the proof, the sequence of integrals

$$\int_M f_1 d\mu, \ \int_M g_1 d\mu, \ \int_M f_2 d\mu, \ \int_M g_2 d\mu, ...$$

converges, which implies (2.10). ■

Hence, if f is a non-negative measurable function then the integral $\int_M f d\mu$ is well-defined and takes value in $[0,+\infty]$.

Theorem 2.8 (a) (Linearity of the integral). *If f is a non-negative measurable function and $c \geq 0$ is a real then*

$$\int_M cf d\mu = c\int_M f d\mu.$$

If f and g are two non-negative measurable functions, then

$$\int_M (f+g)\, d\mu = \int_M f d\mu + \int_M g d\mu.$$

(b) (Monotonicity of the integral) *If $f \leq g$ are non-negative measurable function then*

$$\int_M f d\mu \leq \int_M g d\mu.$$

Proof. (a) By Lemma 2.7, there are sequences $\{f_n\}$ and $\{g_n\}$ of non-negative simple functions such that $f_n \rightrightarrows f$ and $g_n \rightrightarrows g$ on M. Then $cf_n \rightrightarrows cf$ and by Lemma 2.6,

$$\int_M cf d\mu = \lim_{n\to\infty} \int_M cf_n d\mu = \lim_{n\to\infty} c\int_M f_n d\mu = c\int_M f d\mu.$$

Also, we have $f_n + g_n \rightrightarrows f+g$ and, by Lemma 2.6,

$$\int_M (f+g)\, d\mu = \lim_{n\to\infty} \int_M (f_n+g_n)\, d\mu = \lim_{n\to\infty} \left(\int_M f_n d\mu + \int_M g_n d\mu\right) = \int_M f d\mu + \int_M g d\mu.$$

(b) If $f \leq g$ then $g-f$ is a non-negative measurable functions, and $g = (g-f)+f$ whence by (a)

$$\int_M g d\mu = \int_M (g-f)\, d\mu + \int_M f d\mu \geq \int_M f d\mu.$$

■

Example. Let $M = [a,b]$ where $a < b$ and let $\mu = \lambda_1$ be the Lebesgue measure on $[a,b]$. Let $f \geq 0$ be a continuous function on $[a,b]$. Then f is measurable so that the Lebesgue integral $\int_{[a,b]} f d\mu$ is defined. Let us show that it coincides with the Riemann integral $\int_a^b f(x)\, dx$. Let $p = \{x_i\}_{i=0}^n$ be a partition of $[a,b]$ that is,

$$a = x_0 < x_1 < x_2 < ... < x_n = b.$$

The lower Darboux sum is defined by

$$S_*(f,p) = \sum_{i=1}^n m_i (x_i - x_{i-1}),$$

where

$$m_i = \inf_{[x_{i-1},x_i]} f.$$

By the property of the Riemann integral, we have

$$\int_a^b f(x)\,dx = \lim_{m(p)\to 0} S_*(f,p) \tag{2.11}$$

where $m(p) = \max_i |x_i - x_{i-1}|$ is the mesh of the partition.

Consider now a simple function F_p defined by

$$F_p = \sum_{i=1}^n m_i 1_{[x_{i-1},x_i)}.$$

By Lemma 2.6,

$$\int_{[a,b]} F_p d\mu = \sum_{i=1}^n m_i \mu([x_{i-1},x_i)) = S_*(f,p).$$

On the other hand, by the uniform continuity of function f, we have $F_p \rightrightarrows f$ as $m(p) \to 0$, which implies by the definition of the Lebesgue integral that

$$\int_{[a,b]} f d\mu = \lim_{m(p)\to 0} \int_{[a,b]} F_p d\mu = \lim_{m(p)\to 0} S_*(f,p).$$

Comparing with (2.11) we obtain the identity

$$\int_{[a,b]} f\,d\mu = \int_a^b f(x)\,dx.$$

Example. Consider on $[0,1]$ the Dirichlet function

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q}. \end{cases}$$

This function is not Riemann integrable because any upper Darboux sum is 1 and the lower Darboux sum is 0. But the function f is non-negative and simple since it can be represented in the form $f = 1_A$ where $A = \mathbb{Q} \cap [0,1]$ is a measurable set. Therefore, the Lebesgue integral $\int_{[0,1]} f\,d\mu$ is defined. Moreover, since A is a countable set, we have $\mu(A) = 0$ and, hence, $\int_{[0,1]} f\,d\mu = 0$.

2.3.3 Integrable functions

To define the integral of a signed function f on M, let us introduce the notation

$$f_+(x) = \begin{cases} f(x), & \text{if } f(x) \geq 0 \\ 0, & \text{if } f(x) < 0 \end{cases} \quad \text{and} \quad f_- = \begin{cases} 0, & \text{if } f(x) \geq 0 \\ -f(x), & \text{if } f(x) < 0 \end{cases}.$$

The function f_+ is called the *positive part* of f and f_- is called the *negative part* of f. Note that f_+ and f_- are non-negative functions,

$$f = f_+ - f_- \quad \text{and} \quad |f| = f_+ + f_-.$$

It follows that
$$f_+ = \frac{|f| + f}{2} \text{ and } f_- = \frac{|f| - f}{2}.$$
Also, if f is measurable then both f_+ and f_- are measurable.

Definition. A measurable function f is called (Lebesgue) *integrable* if
$$\int_M f_+ d\mu < \infty \text{ and } \int_M f_- d\mu < \infty.$$
For any integrable function, define its Lebesgue integral by
$$\int_M f d\mu := \int_M f_+ d\mu - \int_M f_- d\mu.$$
Note that the integral $\int_M f d\mu$ takes values in $(-\infty, +\infty)$.

In particular, if $f \geq 0$ then $f_+ = f$, $f_- = 0$ and f is integrable if and only if
$$\int_M f d\mu < \infty.$$

Lemma 2.9 (a) *If f is a measurable function then the following conditions are equivalent:*

1. *f is integrable,*
2. *f_+ and f_- are integrable,*
3. *$|f|$ is integrable.*

(b) *If f is integrable then*
$$\left| \int_M f d\mu \right| \leq \int_M |f| \, d\mu.$$

Proof. (a) The equivalence 1. $\Leftrightarrow$ 2. holds by definition. Since $|f| = f_+ + f_-$, it follows that $\int_M |f| \, d\mu < \infty$ if and only if $\int_M f_+ d\mu < \infty$ and $\int_M f_- d\mu < \infty$, that is, 2. $\Leftrightarrow$ 3.

(b) We have
$$\left| \int_M f d\mu \right| = \left| \int_M f_+ d\mu - \int_M f_- d\mu \right| \leq \int_M f_+ d\mu + \int_M f_- d\mu = \int_M |f| \, d\mu.$$
■

Example. Let us show that if f is a continuous function on an interval $[a, b]$ and μ is the Lebesgue measure on $[a, b]$ then f is Lebesgue integrable. Indeed, f_+ and $f\_$ are non-negative and continuous so that they are Lebesgue integrable by the previous Example. Hence, f is also Lebesgue integrable. Moreover, we have
$$\int_{[a,b]} f \, d\mu = \int_{[a,b]} f_+ \, d\mu - \int_{[a,b]} f_- \, d\mu = \int_a^b f_+ \, d\mu - \int_a^b f_- \, d\mu = \int_a^b f \, d\mu$$
so that the Riemann and Lebesgue integrals of f coincide.

Theorem 2.10 (a) (Linearity of integral) *If f is an integrable then, for any real c, cf is also integrable and*

$$\int_M cf d\mu = c\int_M f d\mu.$$

If f, g are integrable then $f+g$ is also integrable and

$$\int_M (f+g)\, d\mu = \int_M f d\mu + \int_M g d\mu. \tag{2.12}$$

(b) (Monotonicity of integral) *If f, g are integrable and $f \leq g$ then $\int_M f d\mu \leq \int_M g d\mu$.*

Proof. (a) If $c = 0$ then there is nothing to prove. Let $c > 0$. Then $(cf)_+ = cf_+$ and $(cf)_- = cf_-$ whence by Lemma 2.8

$$\int_M cf d\mu = \int_M cf_+ d\mu - \int_M cf_- d\mu = c\int_M f_+ d\mu - c\int_M f_- d\mu = c\int_M f d\mu.$$

If $c < 0$ then $(cf)_+ = |c|\, f_-$ and $(cf)_- = |c|\, f_+$ whence

$$\int_M cf d\mu = \int_M |c|\, f_- d\mu - \int_M |c|\, f_+ d\mu = -\,|c| \int_M f d\mu = c\int_M f d\mu.$$

Note that $(f+g)_+$ is not necessarily equal to $f_+ + g_+$ so that the previous simple argument does not work here. Using the triangle inequality

$$|f+g| \leq |f| + |g|,$$

we obtain

$$\int_M |f+g|\, d\mu \leq \int_M |f|\, d\mu + \int_M |g|\, d\mu < \infty,$$

which implies that the function $f+g$ is integrable.

To prove (2.12), observe that

$$f_+ + g_+ - f_- - g_- = f + g = (f+g)_+ - (f+g)_-$$

whence

$$f_+ + g_+ + (f+g)_- = (f+g)_+ + f_- + g_-.$$

Since these all are non-negative measurable (and even integrable) functions, we obtain by Theorem 2.8 that

$$\int_M f_+ d\mu + \int_M g_+ d\mu + \int_M (f+g)_-\, d\mu = \int_M (f+g)_+\, d\mu + \int_M f_- d\mu + \int_M g_- d\mu.$$

It follows that

$$\begin{aligned}
\int_M (f+g)\, d\mu &= \int_M (f+g)_+\, d\mu - \int_M (f+g)_-\, d\mu \\
&= \int_M f_+ d\mu + \int_M g_+ d\mu - \int_M f_- d\mu - \int_M g_- d\mu \\
&= \int_M f d\mu + \int_M g d\mu.
\end{aligned}$$

(b) Indeed, using the identity $g = (g - f) + f$ and that $g - f \geq 0$, we obtain by part (a)

$$\int_M g\, d\mu = \int_M (g - f)\, d\mu + \int_M f\, d\mu \geq \int_M f\, d\mu.$$

■

Example. Let us show that, for any integrable function f,

$$(\inf f)\, \mu(M) \leq \int_M f\, d\mu \leq (\sup f)\, \mu(M).$$

Indeed, consider a function $g(x) \equiv \sup f$ so that $f \leq g$ on M. Since g is a constant function, we have

$$\int_M f\, d\mu \leq \int_M g d\mu = (\sup f)\, \mu(M).$$

In the same way one proves the lower bound.

The following statement shows the connection of integration to the notion $f = g$ a.e.

Theorem 2.11 (a) *If* $f = 0$ a.e.*then* f *is integrable and* $\int_M f d\mu = 0$.
(b) *If* f *is integrable,* $f \geq 0$ a.e. *and* $\int_M f d\mu = 0$ *then* $f = 0$ a.e..

Proof. (a) Since the constant 0 function is measurable, the function f is also measurable. It is suffices to prove that f_+ and f_- are integrable and $\int_M f_+ d\mu = \int_M f_- d\mu = 0$. Note that $f_+ = 0$ a.e. and $f_- = 0$ a.e.. Hence, renaming f_+ or f_- to f, we can assume from the beginning that $f \geq 0$ and $f = 0$ a.e., and need to prove that $\int_M f\, d\mu = 0$. We have by definition

$$\int_M f d\mu = \lim_{n\to\infty} \int_M f_n d\mu$$

where f_n is a simple function defined by

$$f_n(x) = \sum_{k=0}^{\infty} \frac{k}{n} A_{k,n},$$

where

$$A_{k,n} = \left\{ x \in M : \frac{k}{n} \leq f(x) < \frac{k+1}{n} \right\}.$$

The set $A_{k,n}$ has measure 0 if $k > 0$ whence it follows that

$$\int_M f_n d\mu = \sum_{k=0}^{\infty} \frac{k}{n} \mu(A_{k,n}) = 0.$$

Hence, also $\int_M f d\mu = 0$.

(b) Since $f_- = 0$ a.e., we have by part (a) that $\int_M f_-\, d\mu = 0$ which implies that also $\int_M f_+\, d\mu = 0$. It suffices to prove that $f_+ = 0$ a.e.. Renaming f_+ to f, we can assume from the beginning that $f \geq 0$ on M, and need to prove that $\int_M f\, d\mu = 0$ implies $f = 0$ a.e.. Assume from the contrary that $f = 0$ a.e. is not true, that is, the set $\{f > 0\}$ has positive measure. For any $k \in \mathbb{N}$, set $A_k = \left\{ x \in M : f(x) > \frac{1}{k} \right\}$ and observe that

$$\{f > 0\} = \bigcup_{k=1}^{\infty} A_k.$$

It follows that one of the sets A_k must have a positive measure. Fix this k and consider a simple function

$$g(x) = \begin{cases} \frac{1}{k}, & x \in A_k \\ 0, & \text{otherwise,} \end{cases}$$

that is, $g = \frac{1}{k} 1_{A_k}$. It follows that g is measurable and $0 \leq g \leq f$. Hence,

$$\int_M f d\mu \geq \int_M g d\mu = \frac{1}{k}\mu(A_k) > 0,$$

which contradicts the hypothesis. ■

Corollary. *If g is integrable function and f is a function such that $f = g$* a.e. *then f is integrable and $\int_M f d\mu = \int_M g d\mu$.*

Proof. Consider the function $f - g$ that vanishes a.e.. By the previous theorem, $f - g$ is integrable and $\int_M (f - g)\, d\mu = 0$. Then the function $f = (f - g) + g$ is also integrable and

$$\int_M f d\mu = \int_M (f - g)\, d\mu + \int_M g d\mu = \int_M g d\mu,$$

which was to be proved. ■

2.4 Integration over subsets

If $A \subset M$ is a non-empty measurable subset of M and f is a measurable function on A then restricting measure μ to A, we obtain the notion of the Lebesgue integral of f over set A, which is denoted by

$$\int_A f\, d\mu.$$

If f is a measurable function on M then the integral of f over A is defined by

$$\int_A f\, d\mu = \int_A f|_A\, d\mu.$$

Claim *If f is either a non-negative measurable function on M or an integrable function on M then*

$$\int_A f d\mu = \int_M f 1_A\, d\mu. \tag{2.13}$$

Proof. Note that $f 1_A|_A = f|_A$ so that we can rename $f 1_A$ by f and, hence, assume in the sequel that $f = 0$ on $M \setminus A$. Then (2.13) amounts to

$$\int_A f\, d\mu = \int_M f\, d\mu. \tag{2.14}$$

Assume first that f is a simple non-negative function and represent it in the form

$$f = \sum_n b_k 1_{B_k}, \tag{2.15}$$

where the reals $\{b_k\}$ are distinct and the sets $\{B_k\}$ are disjoint. If for some k we have $b_k = 0$ then this value of index k can be removed from the sequence $\{b_k\}$ without violating

(2.15). Therefore, we can assume that $b_k \neq 0$. Then $f \neq 0$ on B_k, and it follows that $B_k \subset A$. Hence, considering the identity (2.15) on both sets A and M, we obtain

$$\int_A f\,d\mu = \sum_k b_k \mu(B_k) = \int_M f\,d\mu.$$

If f is an arbitrary non-negative measurable function then, approximating it by a sequence of simple function, we obtain the same result. Finally, if f is an integrable function then applying the previous claim to f_+ and f_-, we obtain again (2.14). ■

The identity (2.13) is frequently used as the definition of $\int_A f\,d\mu$. It has advantage that it allows to define this integral also for $A = \emptyset$. Indeed, in this case $f 1_A = 0$ and the integral is 0. Hence, we take by definition that also $\int_A f\,d\mu = 0$ when$A = \emptyset$ so that (2.13) remains true for empty A as well.

Theorem 2.12 *Let μ be a finite complete measure on a σ-algebra $\mathcal{M}$ on a set M. Fix a non-negative measurable f function on M and, for any non-empty set $A \in \mathcal{M}$, define a functional $\nu(A)$ by*

$$\nu(A) = \int_A f\,d\mu. \tag{2.16}$$

Then ν is a measure on $\mathcal{M}$.

Example. It follows that any non-negative continuous function f on an interval $(0,1)$ defines a new measure on this interval using (2.16). For example, if $f = \frac{1}{x}$ then, for any interval $[a,b] \subset (0,1)$, we have

$$\nu([a,b]) = \int_{[a,b]} f\,d\mu = \int_a^b \frac{1}{x}dx = \ln\frac{b}{a}.$$

This measure is not finite since $\nu(0,1) = \infty$.

Proof. Note that $\nu(A) \in [0,+\infty]$. We need to prove that ν is σ-additive, that is, if $\{A_n\}$ is a finite or countable sequence of disjoint measurable sets and $A = \bigsqcup_n A_n$ then

$$\int_A f\,d\mu = \sum_n \int_{A_n} f\,d\mu.$$

Assume first that the function f is simple, say, $f = \sum_k b_k 1_{B_k}$ for a sequence $\{B_k\}$ of disjoint sets. Then

$$1_A f = \sum_k b_k 1_{\{A \cap B_k\}}$$

whence

$$\int_A f\,d\mu = \sum_k b_k \mu(A \cap B_k)$$

and in the same way

$$\int_{A_n} f\,d\mu = \sum_k b_k \mu(A_n \cap B_k).$$

It follows that

$$
\begin{aligned}
\sum_n \int_{A_n} f\,d\mu &= \sum_n \sum_k b_k \mu\left(A_n \cap B_k\right) \\
&= \sum_k b_k \mu\left(A \cap B_k\right) \\
&= \int_A f\,d\mu,
\end{aligned}
$$

where we have used the σ-additivity of μ.

For an arbitrary non-negative measurable f, find a simple non-negative function g so that $0 \leq f - g \leq \varepsilon$, for a given $\varepsilon > 0$. Then we have

$$
\int_A g\,d\mu \leq \int_A f\,d\mu = \int_A g\,d\mu + \int_A (f-g)\ d\mu \leq \int_A g\,d\mu + \varepsilon\mu\left(A\right).
$$

In the same way,

$$
\int_{A_n} g\,d\mu \leq \int_{A_n} f d\mu \leq \int_{A_n} g d\mu + \varepsilon\mu\left(A_n\right).
$$

Adding up these inequalities and using the fact that by the previous argument

$$
\sum_n \int_{A_n} g\,d\mu = \int_A g\,d\mu,
$$

we obtain

$$
\begin{aligned}
\int_A g\,d\mu &\leq \sum_n \int_{A_n} f\,d\mu \\
&\leq \sum_n \int_{A_n} g d\mu + \varepsilon \sum_n \mu\left(A_n\right) \\
&= \int_A g d\mu + \varepsilon\mu\left(A\right).
\end{aligned}
$$

Hence, both $\int_A f\,d\mu$ and $\sum_n \int_{A_n} f d\mu$ belong to the interval $\left[\int_A g\,d\mu, \int_A g\,d\mu + \varepsilon\mu\left(A\right)\right]$, which implies that

$$
\left|\sum_n \int_{A_n} f d\mu - \int_A f d\mu\right| \leq \varepsilon\mu\left(A\right).
$$

Letting $\varepsilon \to 0$, we obtain the required identity. ■

Corollary. *Let f be a non-negative measurable function or a (signed) integrable function on M.*

(*a*) (σ-additivity of integral) *If $\{A_n\}$ is a sequence of disjoint measurable sets and $A = \bigsqcup_n A_n$ then*

$$
\int_A f\,d\mu = \sum_n \int_{A_n} f\,d\mu. \tag{2.17}
$$

(*b*) (Continuity of integral) *If $\{A_n\}$ is a monotone (increasing or decreasing) sequence of measurable sets and $A = \lim A_n$ then*

$$
\int_A f\,d\mu = \lim_{n\to\infty} \int_{A_n} f\,d\mu. \tag{2.18}
$$

Proof. If f is non-negative measurable then (a) is equivalent to Theorem 2.12, and (b) follows from (a) because the continuity of measure ν is equivalent to σ-additivity (see Exercise 9). Consider now the case when f is integrable, that is, both $\int_M f_+\,d\mu$ and $\int_M f_-\,d\mu$ are finite. Then, using the σ-additivity of the integral for f_+ and f_-, we obtain

$$\begin{aligned}\int_A f\,d\mu &= \int_A f_+\,d\mu - \int_A f_-\,d\mu \\ &= \sum_n \int_{A_n} f_+\,d\mu - \sum_n \int_{A_n} f_- d\mu \\ &= \sum_n \left(\int_{A_n} f_+\,d\mu - \int_{A_n} f_-\,d\mu\right) \\ &= \sum_n \int_{A_n} f\,d\mu,\end{aligned}$$

which proves (2.17). Finally, since (2.18) holds for the non-negative functions f_+ and f_-, it follows that (2.18) holds also for $f = f_+ - f_-$. ■

2.5 The Lebesgue integral for σ-finite measure

Let us now extend the notion of the Lebesgue integral from finite measures to σ-finite measures.

Let μ be a σ-finite measure on a σ-algebra $\mathcal{M}$ on a set M. By definition, there is a sequence $\{B_k\}_{k=1}^{\infty}$ of measurable sets in M such that $\mu(B_k) < \infty$ and $\bigsqcup_k B_k = M$. As before, denote by μ_{B_k} the restriction of measure μ to measurable subsets of B_k so that μ_{B_k} is a finite measure on B_k.

Definition. For any non-negative measurable function f on M, set

$$\int_M f d\mu := \sum_k \int_{B_k} f d\mu_{B_k}. \tag{2.19}$$

The function f is called integrable if $\int_M f d\mu < \infty$.

If f is a signed measurable function then f is called integrable if both f_+ and f_- are integrable. If f is integrable then set

$$\int_M f d\mu := \int_M f_+ d\mu - \int_M f_- d\mu.$$

Claim. *The definition of the integral in (2.19) is independent of the choice of the sequence $\{B_k\}$.*

Proof. Indeed, let $\{C_j\}$ be another sequence of disjoint measurable subsets of M such that $M = \bigsqcup_j C_j$. Using Theorem 2.12 for finite measures μ_{B_k} and μ_{C_j} (that is, the σ-additivity of the integrals against these measures), as well as the fact that in the intersection of $B_k \cap C_j$ measures μ_{B_k} and μ_{C_j} coincide, we obtain

$$\begin{aligned}
\sum_j \int_{C_j} f \, d\mu_{C_j} &= \sum_j \sum_k \int_{C_j \cap B_k} f \, d\mu_{C_j} \\
&= \sum_k \sum_j \int_{C_j \cap B_k} f \, d\mu_{B_k} \\
&= \sum_k \int_{B_k} f \, d\mu_k .
\end{aligned}$$

■

If A is a non-empty measurable subset of M then the restriction of measure μ on A is also σ-additive, since $A = \bigsqcup_k (A \cap B_k)$ and $\mu(A \cap B_k) < \infty$. Therefore, for any non-negative measurable function f on A, its integral over set A is defined by

$$\int_A f \, d\mu = \sum_k \int_{A \cap B_k} f \, d\mu_{A \cap B_k} .$$

If f is a non-negative measurable function on M then it follows that

$$\int_A f \, d\mu = \int_M f 1_A \, d\mu$$

(which follows from the same identity for finite measures).

Most of the properties of integrals considered above for finite measures, remain true for σ-finite measures: this includes linearity, monotonicity, and additivity properties. More precisely, Theorems 2.8, 2.10, 2.11, 2.12 and Lemma 2.9 are true also for σ-finite measures, and the proofs are straightforward.

For example, let us prove Theorem 2.12 for σ-additive measure: if f is a non-negative measurable function on M then then functional

$$\nu(A) = \int_A f \, d\mu$$

defines a measure on the σ-algebra $\mathcal{M}$. In fact, we need only to prove that ν is σ-additive, that is, if $\{A_n\}$ is a finite or countable sequence of disjoint measurable subsets of M and $A = \bigsqcup_n A_n$ then

$$\int_A f \, d\mu = \sum_n \int_{A_n} f \, d\mu .$$

Indeed, using the above sequence $\{B_k\}$, we obtain

$$\begin{aligned}
\sum_n \int_{A_n} f \, d\mu &= \sum_n \sum_k \int_{A_n \cap B_k} f \, d\mu_{A_n \cap B_k} \\
&= \sum_k \sum_n \int_{A_n \cap B_k} f \, d\mu_{B_k} \\
&= \sum_k \int_{A \cap B_k} f \, d\mu_{B_k} \\
&= \int_A f \, d\mu ,
\end{aligned}$$

where we have used that the integral of f against measure μ_{B_k} is σ-additive (which is true by Theorem 2.12 for finite measures).

Example. Any non-negative continuous function $f(x)$ on $\mathbb{R}$ gives rise to a σ-finite measure ν on $\mathbb{R}$ defined by

$$\nu(A) = \int_A f \, d\mu$$

where $\mu = \lambda_1$ is the Lebesgue measure and A is any Lebesgue measurable subset of $\mathbb{R}$. Indeed, the fact that ν is a measure follows from the version of Theorem 2.12 proved above, and the σ-finite is obvious because $\nu(I) < \infty$ for any bounded interval I. If F is the primitive of f, that is, $F' = f$, then, for any interval I with endpoints $a < b$, we have

$$\nu(I) = \int_{[a,b]} f \, d\mu = \int_a^b f(x) \, dx = F(b) - F(a).$$

Alternatively, measure ν can also be constructed as follows: first define the functional ν of intervals by

$$\nu(I) = F(b) - F(a),$$

prove that it is σ-additive, and then extend ν to measurable sets by the Carathéodory extension theorem.

For example, taking

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

we obtain a finite measure on $\mathbb{R}$; moreover, in this case

$$\nu(\mathbb{R}) = \int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = [\arctan x]_{-\infty}^{+\infty} = 1.$$

Hence, ν is a probability measure on $\mathbb{R}$.

2.6 Convergence theorems

Let μ be a complete measure on a σ-algebra $\mathcal{M}$ on a set M. Considering a sequence $\{f_k\}$ of integrable (or non-negative measurable) functions on M, we will be concerned with the following question: assuming that $\{f_k\}$ converges to a function f in some sense, say, pointwise or almost everywhere, when one can claim that

$$\int_M f_k \, d\mu \to \int_M f \, d\mu \quad ?$$

The following example shows that in general this is not the case.

Example. Consider on an interval $(0,1)$ a sequence of functions $f_k = k 1_{A_k}$ where $A_k = \left[\frac{1}{k}, \frac{2}{k}\right]$. Clearly, $f_k(x) \to 0$ as $k \to \infty$ for any point $x \in (0,1)$. On the other hand, for $\mu = \lambda_1$, we have

$$\int_{(0,1)} f_k \, d\mu = k\mu(A_k) = 1 \not\to 0.$$

For positive results, we start with a simple observation.

Lemma 2.13 *Let μ be a finite complete measure. If $\{f_k\}$ is a sequence of integrable (or non-negative measurable) functions and $f_k \rightrightarrows f$ on M then*

$$\int_M f_k \, d\mu \to \int_M f \, d\mu.$$

Proof. Indeed, we have

$$\begin{aligned}
\int_M f_k \, d\mu &= \int_M f \, d\mu + \int_M (f_k - f) \, d\mu \\
&\leq \int_M f \, d\mu + \int_M |f_k - f| \, d\mu \\
&\leq \int_M f \, d\mu + \sup |f_k - f| \, \mu(M).
\end{aligned}$$

In the same way,

$$\int_M f_k \, d\mu \geq \int_M f \, d\mu - \sup |f_k - f| \, \mu(M).$$

Since $\sup |f_k - f| \, \mu(M) \to 0$ as $k \to \infty$, we conclude that

$$\int_M f_k \, d\mu \to \int_M f \, d\mu.$$

∎

Next, we prove the major results about the integrals of convergent sequences.

Lemma 2.14 (Fatou's lemma) *Let μ be a σ-finite complete measure. Let $\{f_k\}$ be a sequence of non-negative measurable functions on M such that*

$$f_k \to f \text{ a.e.}$$

Assume that, for some constant C and all $k \geq 1$,

$$\int_M f_k \, d\mu \leq C.$$

Then also

$$\int_M f \, d\mu \leq C.$$

Proof. First we assume that measure μ is finite. By Theorem 2.5 (Egorov's theorem), for any $\varepsilon > 0$ there is a set $M_\varepsilon \subset M$ such that $\mu(M \setminus M_\varepsilon) \leq \varepsilon$ and $f_k \rightrightarrows f$ on M_ε. Set

$$A_n = M_1 \cup M_{1/2} \cup M_{1/3} ... \cup M_{1/n}.$$

Then $\mu(M \setminus A_n) \leq \frac{1}{n}$ and $f_n \rightrightarrows f$ on A_n (because $f_n \rightrightarrows f$ on any $M_{1/k}$). By construction, the sequence $\{A_n\}$ is increasing. Set

$$A = \lim_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

By the continuity of integral (Corollary to Theorem 2.12), we have

$$\int_A f\,d\mu = \lim_{n\to\infty}\int_{A_n} f\,d\mu.$$

On the other hand, since $f_k \rightrightarrows f$ on A_n, we have by Lemma 2.13

$$\int_{A_n} f\ d\mu = \lim_{k\to\infty}\int_{A_n} f_k\,d\mu.$$

By hypothesis,

$$\int_{A_n} f_k\,d\mu \leq \int_M f_k\,d\mu \leq C.$$

Passing the limits as $k\to\infty$ and as $n\to\infty$, we obtain

$$\int_A f\,d\mu \leq C.$$

Let us show that

$$\int_{A^c} f\,d\mu = 0, \tag{2.20}$$

which will finish the proof. Indeed,

$$A^c = M\setminus\bigcup_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} A_n^c.$$

Then, for any index n, we have

$$\mu(A^c) \leq \mu(A_n^c) \leq 1/n,$$

whence $\mu(A^c) = 0$. by Theorem 2.11, this implies (2.20) because $f = 0$ a.e. on A^c.

Let now μ be a σ-finite measure. Then there is a sequence of measurable sets $\{B_k\}$ such that $\mu(B_k) < \infty$ and $\bigcup_{k=1}^{\infty} B_k = M$. Setting

$$A_n = B_1 \cup B_2 \cup ... \cup B_n$$

we obtain a similar sequence $\{A_n\}$ but with the additional property that this sequence is increasing. Note that, for any non-negative measurable function f on M,

$$\int_M f\,d\mu = \lim_{n\to\infty}\int_{A_n} f\,d\mu.$$

The hypothesis

$$\int_M f_k\,d\mu \leq C$$

implies that

$$\int_{A_n} f_k\,d\mu \leq C,$$

for all k and n. Since measure μ on A_n is finite, applying the first part of the proof, we conclude that

$$\int_{A_n} f\,d\mu \leq C.$$

Letting $n\to\infty$, we finish the proof. ■

Theorem 2.15 (Monotone convergence theorem) *Let μ be a σ-finite compete measure. Let $\{f_k\}$ be an increasing sequence of non-negative measurable functions and assume that the limit function $f(x) = \lim_{k\to\infty} f_k(x)$ is finite* a.e.. *Then*

$$\int_M f\,d\mu = \lim_{k\to\infty}\int_M f_k\,d\mu.$$

Remark. The integral of f was defined for finite functions f. However, if f is finite a.e., that is, away from some set N of measure 0 then still $\int_M f\,d\mu$ makes sense as follows. Let us change the function f on the set N by assigning on N finite values. Then the integral $\int_M f\,d\mu$ makes sense and is independent of the values of f on N (Theorem 2.11).

Proof. Since $f \geq f_k$, we obtain

$$\int_M f\,d\mu \geq \int_M f_k\,d\mu$$

whence

$$\int_M f\,d\mu \geq \lim_{k\to\infty}\int_M f_k\,d\mu.$$

To prove the opposite inequality, set

$$C = \lim_{k\to\infty}\int_M f_k\,d\mu.$$

Since $\{f_k\}$ is an increasing sequence, we have for any k

$$\int_M f_k\,d\mu \leq C$$

whence by Lemma 2.14 (Fatou's lemma)

$$\int_M f\,d\mu \leq C,$$

which finishes the proof. ■

Theorem 2.16 (The second monotone convergence theorem) *Let μ be a σ-finite compete measure. Let $\{f_k\}$ be an increasing sequence of non-negative measurable functions. If there is a constant C such that, for all k,*

$$\int_M f_k\,d\mu \leq C \tag{2.21}$$

then the sequence $\{f_n(x)\}$ converges a.e. *on M, and for the limit function $f(x) = \lim_{k\to\infty} f_k(x)$ we have*

$$\int_M f\,d\mu = \lim_{k\to\infty}\int_M f_k\,d\mu. \tag{2.22}$$

Proof. Since the sequence $\{f_n(x)\}$ is increasing, it has the limit $f(x)$ for any $x \in M$. We need to prove that the limit $f(x)$ is finite a.e., that is, $\mu(\{f = \infty\}) = 0$ (then (2.22) follows from Theorem 2.15). It follows from (2.21) that, for any $t > 0$,

$$\mu\{f_k > t\} \leq \frac{1}{t}\int_M f_k \, d\mu \leq \frac{C}{t}. \tag{2.23}$$

Indeed, since

$$f_k \geq t1_{\{f_k > t\}},$$

it follows that

$$\int_M f_k \, d\mu \geq \int_M t1_{\{f_k>t\}} d\mu = t\mu\{f_k > t\},$$

which proves (2.23).

Next, since $f(x) = \lim_{k\to\infty} f_k(x)$, it follows that

$$\begin{aligned} \{f(x) > t\} &\subset \{f_k(x) > t \text{ for some } k\} \\ &= \bigcup_k \{f_k(x) > t\} \\ &= \lim_{k\to\infty} \{f_k(x) > t\}, \end{aligned}$$

where we use the fact that f_k monotone increases whence the sets $\{f_k > t\}$ also monotone increases with k. Hence,

$$\mu(f > t) = \lim_{k\to\infty} \mu\{f_k > t\}$$

whence by (2.23)

$$\mu(f > t) \leq \frac{C}{t}.$$

Letting $t \to \infty$, we obtain that $\mu(f = \infty) = 0$ which finishes the proof. ∎

Theorem 2.17 (Dominated convergence theorem) *Let μ be a complete σ-finite measure. Let $\{f_k\}$ be a sequence of measurable functions such that $f_k \to f$* a.e. *Assume that there is a non-negative integrable function g such that $|f_k| \leq g$* a.e. *for all k. Then f_k and f are integrable and*

$$\int_M f \, d\mu = \lim_{k\to\infty} \int_M f_k \, d\mu. \tag{2.24}$$

Proof. By removing from M a set of measure 0, we can assume that $f_k \to f$ pointwise and $|f_k| \leq g$ on M. The latter implies that

$$\int_M |f_k| \, d\mu < \infty,$$

that is, f_k is integrable. whence the integrability of f_k follows. Since also $|f| \leq g$, the function f is integrable in the same way. We will prove that, in fact,

$$\int_M |f_k - f| \, d\mu \to 0 \text{ as } k \to \infty,$$

which will imply that

$$\left|\int_M f_k d\mu - \int_M f \, d\mu\right| = \left|\int_M (f_k - f) \, d\mu\right| \leq \int_M |f_k - f| \, d\mu \to 0$$

which proves (2.24).

Observe that $0 \leq |f_k - f| \leq 2g$ and $|f_k - f| \to 0$ pointwise. Hence, renaming $|f_k - f|$ by f_k and $2g$ by g, we reduce the problem to the following: given that

$$0 \leq f_k \leq g \text{ and } f_k \to 0 \text{ pointwise,}$$

prove that

$$\int_M f_k \, d\mu \to 0 \text{ as } k \to \infty.$$

Consider the function

$$\widetilde{f_n} = \sup \{f_n, f_{n+1}, ...\} = \sup_{k \geq n} \{f_k\} = \lim_{m \to \infty} \sup_{n \leq k \leq m} \{f_k\},$$

which is measurable as the limit of measurable functions. Clearly, we have $0 \leq \widetilde{f_n} \leq g$ and the sequence $\left\{\widetilde{f_n}\right\}$ is decreasing. Let us show that $\widetilde{f_n} \to 0$ pointwise. Indeed, for a fixed point x, the condition $f_k(x) \to 0$ means that for any $\varepsilon > 0$ there is n such that for all

$$k \geq n \Rightarrow f_k(x) < \varepsilon.$$

Taking sup for all $k \geq n$, we obtain $\widetilde{f_n}(x) \leq \varepsilon$. Since the sequence $\left\{\widetilde{f_n}\right\}$ is decreasing, this implies that $\widetilde{f_n}(x) \to 0$, which was claimed.

Therefore, the sequence $\left\{g - \widetilde{f_n}\right\}$ is increasing and $g - \widetilde{f_n} \to g$ pointwise. By Theorem 2.15, we conclude that

$$\int_M \left(g - \widetilde{f_n}\right) d\mu \to \int_M g \, d\mu$$

whence it follows that

$$\int_M \widetilde{f_n} \, d\mu \to 0.$$

Since $0 \leq f_n \leq \widetilde{f_n}$, it follows that also $\int_M f_n \, d\mu \to 0$, which finishes the proof. ■

Example. Both monotone convergence theorem and dominated convergence theorem says that, under certain conditions, we have

$$\lim_{k \to \infty} \int_M f_k \, d\mu = \int_M \left(\lim_{k \to \infty} f_k\right) d\mu,$$

that is, the integral and the pointwise limit are interchangeable. Let us show one application of this to differentiating the integrals depending on parameters. Let $f(t, x)$ be a function from $I \times M \to \mathbb{R}$ where I is an open interval in $\mathbb{R}$ and M is as above a set with a σ-finite measure μ; here $t \in I$ and $x \in M$. Consider the function

$$F(t) = \int_M f(t, x) \, d\mu(x)$$

where $d\mu(x)$ means that the variable of integration is x while t is regarded as a constant. However, after the integration, we regard the result as a function of t. We claim that under certain conditions, the operation of differentiation of F in t and integration are also interchangeable.

Claim. *Assume that $f(t,x)$ is continuously differentiable in $t \in I$ for any $x \in M$, and that the derivative $f' = \frac{\partial f}{\partial t}$ satisfies the estimate*

$$|f'(t,x)| \leq g(x) \text{ for all } t \in I,\ x \in M,$$

where g is an integrable function on M. Then

$$F'(t) = \int_M f'(t,x)\, d\mu. \tag{2.25}$$

Proof. Indeed, for any fixed $t \in I$, we have

$$\begin{aligned} F'(t) &= \lim_{h\to 0} \frac{F(t+h) - F(t)}{h} \\ &= \lim_{h\to 0} \int_M \frac{f(t+h,x) - f(t,x)}{h} d\mu \\ &= \lim_{h\to 0} \int_M f'(c,x)\, d\mu, \end{aligned}$$

where $c = c(h,x)$ is some number in the interval $(t, t+h)$, which exists by the Lagrange mean value theorem. Since $c(h,x) \to t$ as $h \to 0$ for any $x \in M$, it follows that

$$f'(c,x) \to f'(t,x) \text{ as } h \to 0.$$

Since $|f'(c,x)| \leq g(x)$ and g is integrable, we obtain by the dominated convergence theorem that

$$\lim_{h\to 0} \int_M f'(c,x)\, d\mu = f'(t,x)$$

whence (2.25) follows. ∎

Let us apply this Claim to the following particular function $f(t,x) = e^{-tx}$ where $t \in (0,+\infty)$ and $x \in [0,+\infty)$. Then

$$F(t) = \int_0^\infty e^{-tx} dx = \frac{1}{t}.$$

We claim that, for any $t > 0$,

$$F'(t) = \int_0^\infty \frac{\partial}{\partial t}\left(e^{-tx}\right) dx = -\int_0^\infty e^{-tx} x dx$$

Indeed, to apply the Claim, we need to check that $\left|\frac{\partial}{\partial t}\left(e^{-tx}\right)\right|$ is uniformly bounded by an integrable function $g(x)$. We have

$$\left|\frac{\partial}{\partial t}\left(e^{-tx}\right)\right| = xe^{-tx}.$$

For all $t > 0$, this function is bounded only by $g(x) = x$, which is not integrable. However, it suffices to consider t within some interval $(\varepsilon, +\infty)$ for $\varepsilon > 0$ which gives that

$$\left|\frac{\partial}{\partial t}\left(e^{-tx}\right)\right| \leq xe^{-\varepsilon x} =: g(x),$$

and this function g is integrable. Hence, we obtain

$$-\int_0^\infty e^{-tx} x dx = \left(\frac{1}{t}\right)' = -\frac{1}{t^2}$$

that is,

$$\int_0^\infty e^{-tx} x dx = \frac{1}{t^2}.$$

Differentiating further in t we obtain also the identity

$$\int_0^\infty e^{-tx} x^{n-1} dx = \frac{(n-1)!}{t^n}, \tag{2.26}$$

for any $n \in \mathbb{N}$. Of course, this can also be obtained using the integration by parts.

Recall that the gamma function $\Gamma(\alpha)$ for any $\alpha > 0$ is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx.$$

Setting in (2.26) $t = 1$, we obtain that, for any $\alpha \in \mathbb{N}$,

$$\Gamma(\alpha) = (\alpha - 1)!.$$

For non-integer α, $\Gamma(\alpha)$ can be considered as the generalization of the factorial.

2.7 Lebesgue function spaces L^p

Recall that if V is a linear space over $\mathbb{R}$ then a function $N : V \to [0, +\infty)$ is called a *semi-norm* if it satisfies the following two properties:

1. *(the triangle inequality)* $N(x+y) \le N(x) + N(y)$ for all $x, y \in V$;
2. *(the scaling property)* $N(\alpha x) = |\alpha| N(x)$ for all $x \in V$ and $\alpha \in \mathbb{R}$.

It follows from the second property that $N(0) = 0$ (where "0" denote both the zero of V and the zero of $\mathbb{R}$). If in addition $N(x) > 0$ for all $x \neq 0$ then N is called a *norm*. If a norm is chosen in the space V then it is normally denoted by $\|x\|$ rather than $N(x)$.

For example, if $V = \mathbb{R}^n$ then, for any $p \ge 1$, define the p-norm

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}.$$

Then it is known that the p-norm is indeed a norm. This definition extends to $p = \infty$ as follows:

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|,$$

and the ∞-norm is also a norm.

Consider in $\mathbb{R}^n$ the function $N(x) = |x_1|$, which is obviously a semi-norm. However, if $n \ge 2$ then $N(x)$ is not a norm because $N(x) = 0$ for $x = (0, 1, ...)$.

The purpose of this section is to introduce similar norms in the spaces of Lebesgue integrable functions. Consider first the simplest example when M is a finite set of n elements, say,

$$M = \{1, ..., n\},$$

and measure μ is defined on 2^M as the counting measure, that is, $\mu(A)$ is just the cardinality of $A \subset M$. Any function f on M is characterized by n real numbers $f(1), ..., f(n)$, which identifies the space F of all functions on M with $\mathbb{R}^n$. The p-norm on F is then given by

$$\|f\|_p = \left(\sum_{i=1}^{n} |f(i)|^p \right)^{1/p} = \left(\int_M |f|^p \, d\mu \right)^{1/p}, \ p < \infty,$$

and

$$\|f\|_\infty = \sup_M |f| .$$

This will motivate similar constructions for general sets M with measures.

2.7.1 The p-norm

Let μ be complete σ-finite measure on a σ-algebra $\mathcal{M}$ on a set M. Fix some $p \in [1, +\infty)$. For any measurable function $f : M \to \mathbb{R}$, define its p-norm (with respect to measure μ) by

$$\|f\|_p := \left(\int_M |f|^p \, d\mu \right)^{1/p} .$$

Note that $|f|^p$ is a non-negative measurable function so that the integral always exists, finite or infinite. For example,

$$\|f\|_1 = \int_M |f| \, d\mu \ \text{ and } \ \|f\|_2 = \left(\int_M f^2 \, d\mu \right)^{1/2} .$$

In this generality, the p-norm is not necessarily a norm, because the norm must always be finite. Later on, we are going to restrict the domain of the p-norm to those functions for which $\|f\|_p < \infty$ but before that we consider the general properties of the p-norm.

The scaling property of p-norm is obvious: if $\alpha \in \mathbb{R}$ then

$$\|\alpha f\|_p = \left(\int_M |\alpha|^p |f|^p \, d\mu \right)^{1/p} = |\alpha| \left(\int_M |f|^p \, d\mu \right)^{1/p} = |\alpha| \, \|f\|_p .$$

Then rest of this subsection is devoted to the proof of the triangle inequality.

Lemma 2.18 (The Hölder inequality) *Let $p, q > 1$ be such that*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{2.27}$$

Then, for all measurable functions f, g on M,

$$\int_M |fg| \, d\mu \leq \|f\|_p \, \|g\|_p \tag{2.28}$$

(the undefined product $0 \cdot \infty$ is understood as 0).

Remark. Numbers p, q satisfying (2.27) are called the *Hölder conjugate*, and the couple (p, q) is called a Hölder couple.

Proof. Renaming $|f|$ to f and $|g|$ to g, we can assume that f, g are non-negative. If $\|f\|_p = 0$ then by Theorem 2.11 we have $f = 0$ a.e.and (2.28) is trivially satisfied. Hence, assume in the sequel that $\|f\|_p$ and $\|g\|_q > 0$. If $\|f\|_p = \infty$ then (2.28) is again trivial. Hence, we can assume that both $\|f\|_p$ and $\|g\|_q$ are in $(0, +\infty)$.

Next, observe that inequality (2.28) is scaling invariant: if f is replaced by αf when $\alpha \in \mathbb{R}$, then the validity of (2.28) does not change (indeed, when multiplying f by α, the both sides of (2.28) are multiplied by $|\alpha|$). Hence, taking $\alpha = \frac{1}{\|f\|_p}$ and renaming αf to f, we can assume that $\|f\|_p = 1$. In the same way, assume that $\|g\|_q = 1$.

Next, we use the Young inequality:

$$ab \leq \frac{a^p}{p} + \frac{b^p}{q},$$

which is true for all non-negative reals a, b and all Hölder couples p, q. Applying to f, g and integrating, we obtain

$$\begin{aligned}
\int_M fg\, d\mu &\leq \int_M \frac{|f|^p}{p}\, d\mu + \int_M \frac{|g|^q}{q}\, d\mu \\
&= \frac{1}{p} + \frac{1}{q} \\
&= 1 \\
&= \|f\|_p \|g\|_q \,.
\end{aligned}$$

∎

Theorem 2.19 (The Minkowski inequality) *For all measurable functions f, g and for all $p \geq 1$,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \,. \tag{2.29}$$

Proof. If $p = 1$ then (2.29) is trivially satisfied because

$$\|f + g\|_1 = \int_M |f + g|\, d\mu \leq \int_M |f|\, d\mu + \int_M |g|\, d\mu = \|f\|_1 + \|g\|_1 \,.$$

Assume in the sequel that $p > 1$. Since

$$\|f + g\|_p = \||f + g|\|_p \leq \||f| + |g|\|_p \,,$$

it suffices to prove (2.29) for $|f|$ and $|g|$ instead of f and g, respectively. Hence, renaming $|f|$ to f and $|g|$ to g, we can assume that f and g are non-negative.

If $\|f\|_p$ or $\|g\|_p$ is infinity then (2.29) is trivially satisfied. Hence, we can assume in the sequel that both $\|f\|_p$ and $\|g\|_p$ are finite. Using the inequality

$$(a + b)^p \leq 2^p (a^p + b^p) \,,$$

we obtain that

$$\int_M (f + g)^p\, d\mu \leq 2^p \left(\int_M f^p\, d\mu + \int_M g^p\, d\mu \right) < \infty.$$

Hence, $\|f+g\|_p < \infty$. Also, we can assume that $\|f+g\|_p > 0$ because otherwise (2.29) is trivially satisfied.

Finally, we prove (2.29) as follows. Let $q > 1$ be the Hölder conjugate for p, that is, $q = \frac{p}{p-1}$. We have

$$\|f+g\|_p^p = \int_M (f+g)^p \, d\mu = \int_M f(f+g)^{p-1} \, d\mu + \int_M g(f+g)^{p-1} \, d\mu. \tag{2.30}$$

Using the Hölder inequality, we obtain

$$\int_M f(f+g)^{p-1} \, d\mu \le \|f\|_p \left\|(f+g)^{p-1}\right\|_q = \|f\|_p \left(\int_M (f+g)^{(p-1)q} \, d\mu\right)^{1/q} = \|f\|_p \|f+g\|_p^{p/q}.$$

In the same way, we have

$$\int_M g(f+g)^{p-1} \, d\mu \le \|g\|_p \|f+g\|_p^{p/q}.$$

Adding up the two inequalities and using (2.30), we obtain

$$\|f+g\|_p^p \le \left(\|f\|_p + \|g\|_q\right) \|f+g\|_p^{p/q}.$$

Since $0 < \|f+g\|_p < \infty$, dividing by $\|f+g\|_p^{p/q}$ and noticing that $p - \frac{p}{q} = 1$, we obtain (2.29). ■

2.7.2 Spaces L^p

It follows from Theorem 2.19 that the p-norm is a semi-norm. In general, the p-norm is not a norm. Indeed, by Theorem 2.11

$$\int_M |f|^p \, d\mu = 0 \Leftrightarrow f = 0 \text{ a.e.}$$

Hence, there may be non-zero functions with zero p-norm. However, this can be corrected if we identify all functions that differ at a set of measure 0 so that any function f such that $f = 0$ a.e. will be identified with 0. Recall that the relation $f = g$ a.e. is an equivalence relation. Hence, as any equivalence relation, it induces equivalence classes. For any measurable function f, denote (temporarily) by $[f]$ the class of all functions on M that are equal to f a.e., that is,

$$[f] = \{g : f = g \text{ a.e.}\}.$$

Define the linear operations over classes as follows:

$$\begin{aligned} [f] + [g] &= [f+g] \\ \alpha [f] &= [\alpha f] \end{aligned}$$

where f, g are measurable functions and $\alpha \in \mathbb{R}$. Also, define the p-norm on the classes by

$$\|[f]\|_p = \|f\|_p.$$

Of course, this definition does not depend on the choice of a particular representative f of the class $[f]$. In other words, if $f' = f$ a.e. then $\|f'\|_p = \|f\|_p$ and, hence, $\|[f]\|_p$ is well-defined. The same remark applies to the linear operations on the classes.

Definition. *(The Lebesgue function spaces)* For any real $p \geq 1$, denote by $L^p = L^p(M, \mu)$ the set of all equivalence classes $[f]$ of measurable functions such that $\|f\|_p < \infty$.

Claim. *The set L^p is a linear space with respect to the linear operations on the classes, and the p-norm is a norm in L^p.*

Proof. If $[f], [g] \in L^p$ then $\|f\|_p$ and $\|g\|_p$ are finite, whence also $\|[f+g]\|_p = \|f+g\|_p < \infty$ so that $[f] + [g] \in L^p$. It is trivial that also $\alpha[f] \in L^p$ for any $\alpha \in \mathbb{R}$. The fact that the p-norm is a semi-norm was already proved. We are left to observe that $\|[f]\|_p = 0$ implies that $f = 0$ a.e. and, hence, $[f] = [0]$. ■

Convention. It is customary to call the elements of L^p functions (although they are not) and to write $f \in L^p$ instead of $[f] \in L^p$, in order to simplify the notation and the terminology.

Recall that a normed space $(V, \|\cdot\|)$ is called complete (or Banach) if any Cauchy sequence in V converges. That is, if $\{x_n\}$ is a sequence of vectors from V and $\|x_n - x_m\| \to 0$ as $n, m \to \infty$ then there is $x \in V$ such that $x_n \to x$ as $n \to \infty$ (that is, $\|x_n - x\| \to 0$).

The second major theorem of this course (after Carathéodory's extension theorem) is the following.

Theorem 2.20 *For any $p \geq 1$, the space L^p is complete.*

Let us first prove the following criterion of convergence of series.

Lemma 2.21 *Let $\{f_n\}$ be a sequence of measurable functions on M. If*

$$\sum_{n=1}^{\infty} \|f_n\|_1 < \infty$$

then the series $\sum_{n=1}^{\infty} f_n(x)$ converges a.e. *on M and*

$$\int_M \sum_{n=1}^{\infty} f_n(x)\, d\mu = \sum_{n=1}^{\infty} \int_M f_n(x)\, d\mu. \tag{2.31}$$

Proof. We will prove that the series

$$\sum_{n=1}^{\infty} |f_n(x)|$$

converges a.e. which then implies that the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely a.e.. Consider the partial sums

$$F_k(x) = \sum_{n=1}^{k} |f_n(x)|.$$

Clearly, $\{F_k\}$ is an increasing sequence of non-negative measurable functions, and, for all k,

$$\int_M F_k\, d\mu \leq \sum_{n=1}^{\infty} \int_M |f_n|\, d\mu = \sum_{n=1}^{\infty} \|f_n\|_1 =: C.$$

By Theorem 2.16, we conclude that the sequence $\{F_k(x)\}$ converges a.e.. Let $F(x) = \lim_{k\to\infty} F_k(x)$. By Theorem 2.15, we have

$$\int_M F\,d\mu = \lim_{k\to\infty} \int_M F_k d\mu \leq C.$$

Hence, the function F is integrable. Using that, we prove (2.31). Indeed, consider the partial sum

$$G_k(x) = \sum_{n=1}^{k} f_n(x)$$

and observe that

$$|G_k(x)| \leq \sum_{n=1}^{k} |f_n(x)| = F_k(x) \leq F(x).$$

Since

$$G(x) := \sum_{n=1}^{\infty} f_n(x) = \lim_{k\to\infty} G_k(x),$$

we conclude by the dominated convergence theorem that

$$\int_M G\,d\mu = \lim_{k\to\infty} \int_M G_k\,d\mu,$$

which is exactly (2.31). ∎

Proof of Theorem 2.20. Consider first the case $p = 1$. Let $\{f_n\}$ be a Cauchy sequence in L^1, that is,

$$\|f_n - f_m\|_1 \to 0 \text{ as } n, m \to \infty \tag{2.32}$$

(more precisely, $\{[f_n]\}$ is a Cauchy sequence, but this indeed amounts to (2.32)). We need to prove that there is a function $f \in L^1$ such that $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

It is a general property of Cauchy sequences that a Cauchy sequence converges if and only if it has a convergent subsequence. Hence, it suffices to find a convergent subsequence $\{f_{n_k}\}$. It follows from (2.32) that there is a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ such that

$$\left\|f_{n_{k+1}} - f_{n_k}\right\|_1 \leq 2^{-k}.$$

Indeed, define n_k to be a number such that

$$\|f_n - f_m\|_1 \leq 2^{-k} \text{ for all } n, m \geq n_k.$$

Let us show that the subsequence $\{f_{n_k}\}$ converges. For simplicity of notation, rename f_{n_k} into f_k, so that we can assume

$$\|f_{k+1} - f_k\|_1 < 2^{-k}.$$

In particular,

$$\sum_{k=1}^{\infty} \|f_{k+1} - f_k\|_1 < \infty,$$

which implies by Lemma 2.21 that the series

$$\sum_{k=1}^{\infty} (f_{k+1} - f_k)$$

converges a.e.. Since the partial sums of this series are $f_n - f_1$, it follows that the sequence $\{f_n\}$ converges a.e..

Set $f(x) = \lim_{n\to\infty} f_n(x)$ and prove that $\|f_n - f\|_1 \to 0$ as $n \to \infty$. Use again the condition (2.32): for any $\varepsilon > 0$ there exists N such that for all $n, m \geq N$

$$\|f_n - f_m\|_1 < \varepsilon.$$

Since $|f_n - f_m| \to |f_n - f|$ a.e. as $m \to \infty$, we conclude by Fatou's lemma that, for all $n \geq N$,

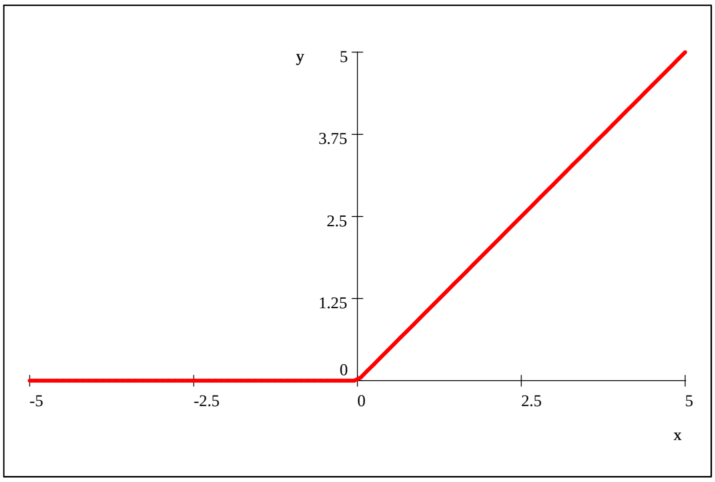
$$\|f_n - f\|_1 \leq \varepsilon,$$

which finishes the proof.

Consider now the case $p > 1$. Given a Cauchy sequence $\{f_n\}$ in L^p, we need to prove that it converges. If $f \in L^p$ then $\int_M |f|^p \, d\mu < \infty$ and, hence, $\int_M f_+^p \, d\mu$ and $\int_M f_-^p \, d\mu$ are also finite. Hence, both f_+ and f_- belong to L^p. Let us show that the sequences $\left\{(f_n)_+\right\}$ and $\left\{(f_n)_-\right\}$ are also Cauchy in L^p. Indeed, set

$$\varphi(t) = t_+ = \begin{cases} t, & t > 0, \\ 0, & t \leq 0, \end{cases}$$

so that $f_+ = \varphi(f)$.



Function $\varphi(t) = t_+$

It is obvious that

$$|\varphi(a) - \varphi(b)| \leq |a - b|.$$

Therefore,

$$\left|(f_n)_+ - (f_m)_+\right| = |\varphi(f_n) - \varphi(f_m)| \leq |f_n - f_m|,$$

whence

$$\left\|(f_n)_+ - (f_m)_+\right\|_p \leq \|f_n - f_m\|_p \to 0$$

as $n, m \to \infty$. This proves that the sequence $\left\{(f_n)_+\right\}$ is Cauchy, and the same argument applies to $\left\{(f_n)_-\right\}$.

It suffices to prove that each sequence $\left\{(f_n)_+\right\}$ and $\left\{(f_n)_-\right\}$ converges in L^p. Indeed, if we know already that $(f_n)_+ \to g$ and $(f_n)_- \to h$ in L^p then

$$f_n = (f_n)_+ - (f_n)_- \to g - h.$$

Therefore, renaming each of the sequence $\left\{(f_n)_+\right\}$ and $\left\{(f_n)_-\right\}$ back to $\{f_n\}$, we can assume in the sequel that $f_n \geq 0$.

The fact that $f_n \in L^p$ implies that $f_n^p \in L^1$. Let us show that the sequence $\{f_n^p\}$ is Cauchy in L^1. For that we use the following elementary inequalities.

Claim. *For all $a, b \geq 0$*

$$|a-b|^p \leq |a^p - b^p| \leq p\,|a-b|\left(a^{p-1} + b^{p-1}\right). \tag{2.33}$$

If $a = b$ then (2.33) is trivial. Assume now that $a > b$ (the case $a < b$ is symmetric). Applying the mean value theorem to the function $\varphi(t) = t^p$, we obtain that, for some $\xi \in (b, a)$,

$$a^p - b^p = \varphi'(\xi)(a-b) = p\xi^{p-1}(a-b).$$

Observing that $\xi^{p-1} \leq a^{p-1}$, we obtain the right hand side inequality of (2.33).

The left hand side inequality in (2.33) amounts to

$$(a-b)^p + b^p \leq a^p,$$

which will follow from the following inequality

$$x^p + y^p \leq (x+y)^p, \tag{2.34}$$

that holds for all $x, y \geq 0$. Indeed, without loss of generality, we can assume that $x \geq y$ and $x > 0$. Then, using the Bernoulli inequality, we obtain

$$(x+y)^p = x^p\left(1 + \frac{y}{x}\right)^p \geq x^p\left(1 + p\frac{y}{x}\right) \geq x^p\left(1 + \left(\frac{y}{x}\right)^p\right) = x^p + y^p.$$

Coming back to a Cauchy sequence $\{f_n\}$ in L^p such that $f_n \geq 0$, let us show that $\{f_n^p\}$ is Cauchy in L^1. We have by (2.33)

$$|f_n^p - f_m^p| \leq p\,|f_n - f_m|\left(f_n^{p-1} + f_m^{p-1}\right).$$

Recall the Hölder inequality

$$\int_M |fg|\, d\mu \leq \left(\int_M |f|^p\, d\mu\right)^{1/p}\left(\int_M |g|^q\, d\mu\right)^{1/q},$$

where q is the Hölder conjugate to p, that is, $\frac{1}{p} + \frac{1}{q} = 1$. Using this inequality, we obtain

$$\int_M |f_n^p - f_m^p|\, d\mu \leq p\left(\int_M |f_n - f_m|^p\, d\mu\right)^{1/p}\left(\int_M \left(f_n^{p-1} + f_m^{p-1}\right)^q\, d\mu\right)^{1/q}. \tag{2.35}$$

To estimate the last integral, use the inequality

$$(x+y)^q \le 2^q \left(x^q + y^q\right),$$

which yields

$$\left(f_n^{p-1} + f_m^{p-1}\right)^q \le 2^q \left(f_n^{(p-1)q} + f_m^{(p-1)q}\right) = 2^q \left(f_n^p + f_m^p\right), \tag{2.36}$$

where we have used the identity $(p-1)\,q = p$. Note that the numerical sequence $\left\{\|f_n\|_p\right\}$ is bounded. Indeed, by the Minkowski inequality,

$$\left|\|f_n\|_p - \|f_m\|_p\right| \le \|f_n - f_m\|_p \to 0$$

as $n, m \to \infty$. Hence, the sequence $\left\{\|f_n\|_p\right\}_{n=1}^{\infty}$ is a numerical Cauchy sequence and, hence, is bounded. Integrating (2.36) we obtain that

$$\int_M \left(f_n^{p-1} + f_m^{p-1}\right)^q d\mu \le 2^q \left(\int_M f_n^p \, d\mu + \int_M f_m^p \, d\mu\right) \le \text{const},$$

where the constant const is independent of n, m. Hence, by (2.35)

$$\int_M \left|f_n^p - f_m^p\right| d\mu \le \text{const} \left(\int_M |f_n - f_m|^p \, d\mu\right)^{1/p} = \text{const}\, \|f_n - f_m\|_p \to 0 \text{ as } n, m \to \infty,$$

whence it follows that the sequence $\{f_n^p\}$ is Cauchy in L^1. By the first part of the proof, we conclude that this sequence converges to a function $g \in L^1$. By Exercise 52, we have $g \ge 0$. Set $f = g^{1/p}$ and show that $f_n \to f$ in L^p. Indeed, by (2.33) we have

$$\int_M |f_n - f|^p \, d\mu \le \int_M \left|f_n^p - f^p\right| d\mu = \int_M \left|f_n^p - g\right| d\mu \to 0$$

which was to be proved. ■

Finally, there is also the space L^∞ (that is, L^p with $p = \infty$), which is defined in Exercise 56 and which is also complete.

2.8 Product measures and Fubini's theorem

2.8.1 Product measure

Let M_1, M_2 be two non-empty sets and S_1, S_2 be families of subsets of M_1 and M_2, respectively. Consider on the set $M = M_1 \times M_2$ the family of subsets

$$S = S_1 \times S_2 := \{A \times B : A \in S_1, B \in S_2\}.$$

Recall that, by Exercise 6, if S_1 and S_2 are semi-rings then $S_1 \times S_2$ is also a semi-ring.

Let μ_1 and μ_2 be two two functionals on S_1 and S_2, respectively. As before, define their product $\mu = \mu_1 \times \mu_2$ as a functional on S by

$$\mu(A \times B) = \mu_1(A)\, \mu_2(B) \text{ for all } A \in S_1,\ B \in S_2.$$

By Exercise 20, if μ_1 and μ_2 are finitely additive measures on semi-rings S_1 and S_2 then μ is finitely additive on S. We are going to prove (independently of Exercise 20), that if μ_1 and μ_2 are σ-additive then so is μ. A particular case of this result when $M_1 = M_2 = \mathbb{R}$ and $\mu_1 = \mu_2$ is the Lebesgue measure was proved in Lemma 1.11 using specific properties of $\mathbb{R}$. Here we consider the general case.

Theorem 2.22 *Assume that μ_1 and μ_2 are σ-finite measures on semi-rings S_1 and S_2 on the sets M_1 and M_2 respectively. Then $\mu = \mu_1 \times \mu_2$ is a σ-finite measure on $S = S_1 \times S_2$.*

By Carathéodory's extension theorem, μ can be then uniquely extended to the σ-algebra of measurable sets on M. The extended measure is also denoted by $\mu_1 \times \mu_2$ and is called the *product measure* of μ_1 and μ_2. Observe that the product measure is σ-finite (and complete). Hence, one can define by induction the product $\mu_1 \times ... \times \mu_n$ of a finite sequence of σ-finite measures $\mu_1, ..., \mu_n$. Note that this product satisfies the associative law (Exercise 57).

Proof. Let C, C_n be sets from S such that $C = \bigsqcup_n C_n$ where n varies in a finite or countable set. We need to prove that

$$\mu(C) = \sum_n \mu(C_n). \tag{2.37}$$

Let $C = A \times B$ and $C_n = A_n \times B_n$ where $A, A_n \in S_1$ and $B, B_n \in S_2$. Clearly, $A = \bigcup_n A_n$ and $B = \bigcup_n B_n$.

Consider function f_n on M_1 defined by

$$f_n(x) = \begin{cases} \mu_2(B_n), & x \in A_n, \\ 0, & x \notin A_n \end{cases}$$

that is $f_n = \mu_2(B_n) 1_{A_n}$.

Claim. *For any $x \in A$,*

$$\sum_n f_n(x) = \mu_2(B). \tag{2.38}$$

We have

$$\sum_n f_n(x) = \sum_{n:x\in A_n} f_n(x) = \sum_{n:x\in A_n} \mu_2(B_n).$$

If $x \in A_n$ and $x \in A_m$ and $n \neq m$ then the sets B_n, B_m must be disjoint. Indeed, if $y \in B_n \cap B_m$ then $(x, y) \in C_n \cap C_m$, which contradicts the hypothesis that C_n, C_m are disjoint. Hence, all sets B_n with the condition $x \in A_n$ are disjoint. Furthermore, their union is B because the set $\{x\} \times B$ is covered by the union of the sets C_n. Hence,

$$\sum_{n:x\in A_n} \mu_2(B_n) = \mu_2\left(\bigsqcup_{n:x\in A_n} B_n\right) = \mu_2(B),$$

which finishes the proof of the Claim.

Integrating the identity (2.38) over A, we obtain

$$\int_A \left(\sum_n f_n\right) d\mu_1 = \mu_2(B)\ \mu_1(A) = \mu(C).$$

On the other hand, since $f_n \geq 0$, the partials sums of the series $\sum_n f_n$ form an increasing sequence. Applying the monotone convergence theorem to the partial sums, we obtain that the integral is interchangeable with the infinite sum, whence

$$\int_A \left(\sum_n f_n\right) d\mu_1 = \sum_n \int_A f_n\, d\mu_1 = \sum_n \mu_2(B_n)\, \mu_1(A_n) = \sum_n \mu(C_n).$$

Comparing the above two lines, we obtain (2.37).

Finally, let us show that measure μ is σ-finite. Indeed, since μ_1 and μ_2 are σ-finite, there are sequences $\{A_n\} \subset S_1$ and $\{B_n\} \subset S_2$ such that $\mu_1(A_n) < \infty$, $\mu_2(B_n) < \infty$ and $\bigcup_n A_n = M_1$, $\bigcup_n B_n = M_2$. Set $C_{nm} = A_n \times B_m$. Then $\mu(C_{nm}) = \mu_1(A_n)\,\mu_2(B_m) < \infty$ and $\bigcup_{n,m} C_{nm} = M$. Since the sequence $\{C_{nm}\}$ is countable, we conclude that measure μ is σ-finite. ■

2.8.2 Cavalieri principle

Before we state the main result, let us discuss integration of functions taking value ∞. Let μ be an complete σ-finite measure on a set M and let f be a measurable function on M with values in $[0, \infty]$. If the set

$$S := \{x \in M : f(x) = \infty\}$$

has measure 0 then there is no difficulty in defining $\int_M f\,d\mu$ since the function f can be modified on set S to have finite values. If $\mu(S) > 0$ then we define the integral of f by

$$\int_M f\,d\mu = \infty. \tag{2.39}$$

For example, let f be a simple function, that is, $f = \sum_k a_k 1_{A_k}$ where $a_k \in [0, \infty]$ are distinct values of f. Assume that one of a_k is equal to ∞. Then $S = A_k$ with this k, and if $\mu(A_k) > 0$ then

$$\sum_k a_k \mu(A_k) = \infty,$$

which matches the definition (2.39).

Many properties of integration remain true if the function under the integral is allowed to take value ∞. We need the following extension of the monotone convergence theorem.

Extended monotone convergence theorem *If $\{f_k\}_{k=1}^{\infty}$ is an increasing sequence of measurable functions on M taking values in $[0, \infty]$ and $f_k \to f$ a.e. as $k \to \infty$ then*

$$\int_M f\,d\mu = \lim_{k\to\infty} \int_M f_k\,d\mu. \tag{2.40}$$

Proof. Indeed, if f is finite a.e. then this was proved in Theorem 2.15. Assume that $f = \infty$ on a set of positive measure. If one of f_k is also equal to ∞ on a set of positive measure then $\int_M f_k\,d\mu = \infty$ for large enough k and the both sides of (2.40) are equal to ∞. Consider the remaining case when all f_k are finite a.e.. Set

$$C = \lim_{k\to\infty} \int_M f_k\,d\mu.$$

We need to prove that $C = \infty$. Assume from the contrary that $C < \infty$. Then, for all k,

$$\int_M f_k\,d\mu \le C,$$

and by the second monotone convergence theorem (Theorem 2.16) we conclude that the limit $f(x) = \lim_{k\to\infty} f_k(x)$ is finite a.e., which contradicts the assumption that $\mu\{f = \infty\} > 0$. ■

Now, let μ_1 and μ_2 be σ-finite measures defined on σ-algebras $\mathcal{M}_1$ and $\mathcal{M}_2$ on sets M_1, M_2. Let $\mu = \mu_1 \times \mu_2$ be the product measure on $M = M_1 \times M_2$, which is defined on a σ-algebra $\mathcal{M}$. Consider a set $A \in \mathcal{M}$ and, for any $x \in M$ consider the x-section of A, that is, the set

$$A_x = \{y \in M_2 : (x, y) \in A\}.$$

Define the following function on M_1:

$$\varphi_A(x) = \begin{cases} \mu_2(A_x), & \text{if } A_x \in \mathcal{M}_2, \\ 0 & \text{otherwise} \end{cases} \tag{2.41}$$

Theorem 2.23 (The Cavalieri principle) *If $A \in \mathcal{M}$ then, for almost all $x \in M_1$, the set A_x is a measurable subset of M_2 (that is, $A_x \in \mathcal{M}_2$), the function $\varphi_A(x)$ is measurable on M_1, and*

$$\mu(A) = \int_{M_1} \varphi_A \, d\mu_1. \tag{2.42}$$

Note that the function $\varphi_A(x)$ takes values in $[0, \infty]$ so that the above discussion of the integration of such functions applies.

Example. The formula (2.42) can be used to evaluate the areas in $\mathbb{R}^2$ (and volumes in $\mathbb{R}^n$). For example, consider a measurable function function $\varphi : (a, b) \to [0, +\infty)$ and let A be the subgraph of φ, that is,

$$A = \left\{(x, y) \in \mathbb{R}^2 : a < x < b, 0 \le y \le \varphi(x)\right\}.$$

Considering $\mathbb{R}^2$ as the product $\mathbb{R} \times \mathbb{R}$, and the two-dimensional Lebesgue measure λ_2 as the product $\lambda_1 \times \lambda_1$, we obtain

$$\varphi_A(x) = \lambda_1(A_x) = \lambda_1[0, \varphi(x)] = \varphi(x),$$

provided $x \in (a, b)$ and $\varphi_A(x) = 0$ otherwise. Hence,

$$\lambda_2(A) = \int_{(a,b)} \varphi d\lambda_1.$$

If φ is Riemann integrable then we obtain

$$\lambda_2(A) = \int_a^b \varphi(x) \, dx,$$

which was also proved in Exercise 23.

Proof of Theorem 2.23. Denote by $\mathcal{A}$ the family of all sets $A \in \mathcal{M}$ for which all the claims of Theorem 2.23 are satisfied. We need to prove that $\mathcal{A} = \mathcal{M}$.

Consider first the case when the measures μ_1, μ_2 are finite. The finiteness of measure μ_2 implies, in particular, that the function $\varphi_A(x)$ is finite. Let show that $\mathcal{A} \supset S := \mathcal{M}_1 \times \mathcal{M}_2$ (note that S is a semi-ring). Indeed, any set $A \in S$ has the form $A = A_1 \times A_2$ where $A_i \in \mathcal{M}_i$. Then we have

$$A_x = \begin{cases} A_2, & x \in A_1, \\ \emptyset, & x \notin A_1, \end{cases} \in \mathcal{M}_2,$$

$$\varphi_A(x) = \begin{cases} \mu_2(A_2), & x \in A_1, \\ 0, & x \notin A_1, \end{cases}$$

so that φ_A is a measurable function on M_1, and

$$\int_{M_1} \varphi_A \, d\mu_1 = \mu_2(A_2)\,\mu_1(A_1) = \mu(A).$$

All the required properties are satisfied and, hence, $A \in \mathcal{A}$.

Let us show that $\mathcal{A}$ is closed under monotone difference (that is, $\mathcal{A}^- = \mathcal{A}$). Indeed, if $A, B \in \mathcal{A}$ and $A \supset B$ then

$$(A - B)_x = A_x - B_x,$$

so that $(A - B)_x \in \mathcal{M}_2$ for almost all $x \in M_1$,

$$\varphi_{A-B}(x) = \mu_2(A_x) - \mu_2(B_x) = \varphi_A(x) - \varphi_B(x)$$

so that the function φ_{A-B} is measurable, and

$$\int_{M_1} \varphi_{A-B} \, d\mu_1 = \int_{M_1} \varphi_A \, d\mu_1 - \int_{M_1} \varphi_B \, d\mu_1 = \mu(A) - \mu(B) = \mu(A - B).$$

In the same way, $\mathcal{A}$ is closed under taking finite disjoint unions.

Let us show that $\mathcal{A}$ is closed under monotone limits. Assume first that $\{A_n\} \subset \mathcal{A}$ is an increasing sequence and let $A = \lim_{n\to\infty} A_n = \bigcup_n A_n$. Then

$$A_x = \bigcup_n (A_n)_x$$

so that $A_x \in \mathcal{M}_2$ for almost all $x \in M_1$,

$$\varphi_A(x) = \mu_2(A_x) = \lim_{n\to\infty} \mu_2((A_n)_x) = \lim_{n\to\infty} \varphi_{A_n}(x).$$

In particular, φ_A is measurable as the pointwise limit of measurable functions. Since the sequence $\{\varphi_{A_n}\}$ is monotone increasing, we obtain by the monotone convergence theorem

$$\int_{M_1} \varphi_A \, d\mu_1 = \lim_{n\to\infty} \int_{M_1} \varphi_{A_n} \, d\mu_1 = \lim_{n\to\infty} \mu(A_n) = \mu(A). \tag{2.43}$$

Hence, $A \in \mathcal{A}$. Note that this argument does not require the finiteness of measures μ_1 and μ_2 because if these measures are only σ-finite and, hence, the functions φ_{A_k} are not necessarily finite, the monotone convergence theorem can be replaced by the extended monotone convergence theorem. This remark will be used below.

Let now $\{A_n\} \subset \mathcal{A}$ be decreasing and let $A = \lim_{n\to\infty} A_n = \bigcap_n A_n$. Then all the above argument remains true, the only difference comes in justification of (2.43). We cannot use the monotone convergence theorem, but one can pass to the limit under the sign of the integral by the dominated convergence theorem because all functions φ_{A_n} are bounded by the function φ_{A_1}, and the latter is integrable because it is bounded by $\mu_2(M_2)$ and the measure $\mu_1(M_1)$ is finite.

Let Σ be the minimal σ-algebra containing S. From the above properties of $\mathcal{A}$, it follows that $\mathcal{A} \supset \Sigma$. Indeed, let R be the minimal algebra containing S, that is, R consists of finite disjoint union of elements of S. Since $\mathcal{A}$ contains S and $\mathcal{A}$ is closed under taking finite disjoint unions, we conclude that $\mathcal{A}$ contains R. By Theorem 1.14, $\Sigma = R^{\lim}$, that is, Σ is the extension of R by monotone limits. Since $\mathcal{A}$ is closed under monotone limits, we conclude that $A \supset \Sigma$.

To proceed further, we need a refinement of Theorem 1.10. The latter says that if $A \in \mathcal{M}$ then there is $B \in \Sigma$ such that $\mu(A \vartriangle B) = 0$. In fact, set B can be chosen to contain A, as is claimed in the following stronger statement.

Claim (Exercise 54) *For any set $A \in \mathcal{M}$, there is a set $B \in \Sigma$ such that $B \supset A$ and $\mu(B - A) = 0$.*

Using this Claim, let us show that $\mathcal{A}$ contains sets of measure 0. Indeed, for any set A such that $\mu(A) = 0$, there is $B \in \Sigma$ such that $A \subset B$ and $\mu(B) = 0$ (the latter being equivalent to $\mu(B - A) = 0$). It follows that

$$\int_{M_1} \varphi_B \, d\mu_1 = \mu(B) = 0,$$

which implies that $\mu_2(B_x) = \varphi_B(x) = 0$ for almost all $x \in M_1$. Since $A_x \subset B_x$, it follows that $\mu_2(A_x) = 0$ for almost all $x \in M_1$. In particular, the set A_x is measurable for almost all $x \in M_1$. Furthermore, we have $\varphi_A(x) = 0$ a.e. and, hence,

$$\int_{M_1} \varphi_A \, d\mu_1 = 0 = \mu(A),$$

which implies that $A \in \mathcal{A}$.

Finally, let us show that $\mathcal{A} = \mathcal{M}$. By the above Claim, for any set $A \in \mathcal{M}$, there is $B \in \Sigma$ such that $B \supset A$ and $\mu(B - A) = 0$. Then $B \in \mathcal{A}$ by the previous argument, and the set $C = B - A$ is also in $\mathcal{A}$ as a set of measure 0. Since $A = B - C$ and $\mathcal{A}$ is closed under monotone difference, we conclude that $A \in \mathcal{A}$.

Consider now the general case of σ-finite measures μ_1, μ_2. Consider an increasing sequences $\{M_{ik}\}_{k=1}^{\infty}$ of measurable sets in M_i $(i = 1, 2)$ such that $\mu_i(M_{ik}) < \infty$ and $\bigcup_k M_{ik} = M_i$. Then measure μ_i is finite in M_{ik} so that the first part of the proof applies to sets

$$A_k = A \cap (M_{1k} \times M_{2k}),$$

so that $A_k \in \mathcal{A}$. Note that $\{A_k\}$ is an increasing sequence and $A = \bigcup_{k=1} A_k$. Since the family $\mathcal{A}$ is closed under increasing monotone limits, we obtain that $A \in \mathcal{A}$, which finishes the proof. ■

Example. Let us evaluate the area of the disc in $\mathbb{R}^2$

$$D = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < r^2\right\},$$

where $r > 0$. We have

$$D_x = \left\{y \in \mathbb{R} : |y| < \sqrt{r^2 - x^2}\right\}$$

and

$$\varphi_D(x) = \lambda_1(D_x) = 2\sqrt{r^2 - x^2},$$

provided $|x| < r$. If $|x| \geq r$ then $D_x = \emptyset$ and $\varphi_D(x) = 0$. Hence, by Theorem 2.23,

$$\begin{aligned}
\lambda_2(D) &= \int_{-r}^{r} 2\sqrt{r^2 - x^2}dx \\
&= 2r^2 \int_{-1}^{1} \sqrt{1 - t^2}dt \\
&= 2r^2 \frac{\pi}{2} = \pi r^2.
\end{aligned}$$

Now let us evaluate the volume of the ball in $\mathbb{R}^3$

$$B = \left\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < r^2\right\}.$$

We have

$$B_x = \left\{(y, z) \in \mathbb{R}^2 : y^2 + z^2 < r^2 - x^2\right\},$$

which is a disk of radius $\sqrt{r^2 - x^2}$ in $\mathbb{R}^2$, whence

$$\varphi_B(x) = \lambda_2(B_x) = \pi\left(r^2 - x^2\right)$$

provided $|x| < r$, and $\varphi_B(x) = 0$ otherwise. Hence,

$$\lambda_3(B) = \int_{-r}^{r} \pi\left(r^2 - x^2\right) dx = \frac{4}{3}\pi r^3.$$

One can show by induction that if $B_n(r)$ is the n-dimensional ball of radius r then

$$\lambda_n(B_n(r)) = c_n r^n$$

with some constant c_n (see Exercise 59).

Example. Let B be a measurable subset of $\mathbb{R}^{n-1}$ of a finite measure, p be a point in $\mathbb{R}^n$ with $p_n > 0$, and C be a cone over B with the pole at p defined by

$$C = \{(1 - t)p + ty : y \in B,\ 0 < t < 1\}.$$

Let us show that

$$\lambda_n(C) = \frac{\lambda_{n-1}(B)\,p_n}{n}.$$

Considering $\mathbb{R}^n$ as the product $\mathbb{R} \times \mathbb{R}^{n-1}$, we obtain, for any $0 < z < p_n$,

$$\begin{aligned}
C_z &= \{x \in C : x_n = z\} = \{(1 - t)p + ty : y \in B,\ \ (1 - t)p_n = z\} \\
&= (1 - t)p + tB,
\end{aligned}$$

where $t = 1 - \frac{z}{p_n}$; besides, if $z \notin (0, p_n)$ then $C_z = \emptyset$. Hence, for any $0 < z < p_n$, we have

$$\varphi_z(C) = \lambda_{n-1}(C_z) = t^{n-1}\lambda_{n-1}(B),$$

and

$$\lambda_n(C) = \lambda_{n-1}(B)\int_0^{p_n}\left(1 - \frac{z}{p_n}\right)^{n-1} dz = \lambda_{n-1}(B)\,p_n\int_0^1 s^{n-1}ds = \frac{\lambda_{n-1}(B)\,p_n}{n}.$$

2.8.3 Fubini's theorem

Theorem 2.24 (Fubini's theorem) *Let μ_1 and μ_2 be complete σ-finite measures on the sets M_1, M_2, respectively. Let $\mu = \mu_1 \times \mu_2$ be the product measure on $M = M_1 \times M_2$ and let $f(x, y)$ be a measurable function on M with values in $[0, \infty]$, where $x \in M_1$ and $y \in M_2$. Then*

$$\int_M f \, d\mu = \int_{M_1} \left(\int_{M_2} f(x, y) \, d\mu_2(y) \right) d\mu_1(x). \tag{2.44}$$

More precisely, the function $y \mapsto f(x, y)$ is measurable on M_2 for almost all $x \in M_1$, the function

$$x \mapsto \int_{M_2} f(x, y) \, d\mu_2$$

is measurable on M_1, and its integral over M_1 is equal to $\int_M f \, d\mu$.

Furthermore, if f is a (signed) integrable function on M, then the function $y \mapsto f(x, y)$ is integrable on M_2 for almost all $x \in M_1$, the function

$$x \mapsto \int_{M_2} f(x, y) \, d\mu_2$$

is integrable on M_1, and the identity (2.44) *holds.*

The first part of Theorem 2.24 dealing with non-negative functions is also called *Tonneli's theorem.* The expression on the right hand side of (2.44) is referred to as a *repeated* (or *iterated*) integral. Switching x and y in (2.44) yields the following identity, which is true under the same conditions:

$$\int_M f \, d\mu = \int_{M_2} \left(\int_{M_1} f(x, y) \, d\mu_1(x) \right) d\mu_2(y).$$

By induction, this theorem extends to a finite product $\mu = \mu_1 \times \mu_2 \times ... \times \mu_n$ of measures.

Proof. Consider the product

$$\widetilde{M} = M \times \mathbb{R} = M_1 \times M_2 \times \mathbb{R}$$

and measure $\widetilde{\mu}$ on $\widetilde{M}$, which is defined by

$$\widetilde{\mu} = \mu \times \lambda = \mu_1 \times \mu_2 \times \lambda,$$

where $\lambda = \lambda_1$ is the Lebesgue measure on $\mathbb{R}$ (note that the product of measure satisfies the associative law – see Exercise 57). Consider the set $A \subset \widetilde{M}$, which is the subgraph of function f, that is,

$$A = \left\{ (x, y, z) \in \widetilde{M} : 0 \leq z \leq f(x, y) \right\}$$

(here $x \in M_1$, $y \in M_2$, $z \in \mathbb{R}$). By Theorem 2.23, we have

$$\mu(A) = \int_M \lambda\left(A_{(x,y)}\right) d\mu.$$

Since $A_{(x,y)} = [0, f(x,y)]$ and $\lambda\left(A_{(x,y)}\right) = f(x,y)$, it follows that

$$\mu(A) = \int_M f\, d\mu. \tag{2.45}$$

On the other hand, we have

$$\widetilde{M} = M_1 \times (M_2 \times \mathbb{R})$$

and

$$\widetilde{\mu} = \mu_1 \times (\mu_2 \times \lambda).$$

Theorem 2.23 now says that

$$\mu(A) = \int_{M_1} (\mu_2 \times \lambda)(A_x)\, d\mu_1(x)$$

where

$$A_x = \{(y,z) : 0 \le f(x,y) \le z\}.$$

Applying Theorem 2.23 to the measure $\mu_2 \times \lambda$, we obtain

$$(\mu_2 \times \lambda)(A_x) = \int_{M_2} \lambda\left((A_x)_y\right) d\mu_2 = \int_{M_2} f(x,y)\, d\mu_2(y).$$

Hence,

$$\mu(A) = \int_{M_1} \left(\int_{M_2} f(x,y)\, d\mu_2(y) \right) d\mu_1(x).$$

Comparing with (2.45), we obtain (2.44). The claims about the measurability of the intermediate functions follow from Theorem 2.23.

Now consider the case when f is integrable. By definition, function f is integrable if f_+ and f_- are integrable. By the first part, we have

$$\int_M f_+ d\mu = \int_{M_1} \left(\int_{M_2} f_+(x,y)\, d\mu_2(y) \right) d\mu_1.$$

Since the left hand side here is finite, it follows that the function

$$F(x) = \int_{M_2} f_+(x,y)\, d\mu_2(y)$$

is integrable and, hence, is finite a.e.. Consequently, the function $f_+(x,y)$ is integrable in y for almost all x. In the same way, we obtain

$$\int_M f_- d\mu = \int_{M_1} \left(\int_{M_2} f_-(x,y)\, d\mu_2(y) \right) d\mu_1,$$

so that the function

$$G(x) = \int_{M_2} f_-(x,y)\, d\mu_2(y)$$

is integrable and, hence, is finite a.e.. Therefore, the function $f_-(x,y)$ is integrable in y for almost all x. We conclude that $f = f_+ - f_-$ is integrable in y for almost all x, the function

$$\int_{M_2} f(x,y)\, d\mu_2(y) = F(x) - G(x)$$

is integrable on M_1, and

$$\int_{M_1}\left(\int_{M_2} f(x,y)\, d\mu_2(y)\right) d\mu_1(x) = \int_{M_1} F\, d\mu_1 - \int_{M_1} G\, d\mu_1 = \int_M f\, d\mu.$$

■

Example. Let Q be the unit square in $\mathbb{R}^2$, that is,

$$Q = \left\{(x,y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 1\right\}.$$

Let us evaluate

$$\int_Q (x+y)^2\, d\lambda_2.$$

Since $Q = (0,1) \times (0,1)$ and $\lambda_2 = \lambda_1 \times \lambda_1$, we obtain by (2.44)

$$\begin{aligned}
\int_Q (x+y)^2\, d\lambda_2 &= \int_0^1 \left(\int_0^1 (x+y)^2\, dx\right) dy \\
&= \int_0^1 \left[\frac{(x+y)^3}{3}\right]_0^1 dy \\
&= \int_0^1 \frac{(1+y)^3 - y^3}{3} dy \\
&= \int_0^1 \left(y^2 + y + \frac{1}{3}\right) dy \\
&= \frac{1}{3} + \frac{1}{2} + \frac{1}{3} = \frac{7}{6}.
\end{aligned}$$

3 Integration in Euclidean spaces and in probability spaces

3.1 Change of variables in Lebesgue integral

Recall the following rule of change of variable in a Riemann integral:

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f(\varphi(y))\,\varphi'(y)\,dy$$

if φ is a continuously differentiable function on $[\alpha, \beta]$ and $a = \varphi(\alpha)$, $b = \varphi(\beta)$.

Let us rewrite this formula for Lebesgue integration. Assume that $a < b$ and $\alpha < \beta$. If φ is an increasing diffeomorphism from $[\alpha, \beta]$ onto $[a, b]$ then $\varphi(\alpha) = a$, $\varphi(\beta) = b$, and we can rewrite this formula for the Lebesgue integration:

$$\int_{[a,b]} f(x)\,d\lambda_1(x) = \int_{[\alpha,\beta]} f(\varphi(y))\,|\varphi'(y)|\,d\lambda_1(y)\,, \tag{3.1}$$

where we have used that $\varphi' \geq 0$.

If φ is a decreasing diffeomorphism from $[\alpha, \beta]$ onto $[a, b]$ then $\varphi(\alpha) = b$, $\varphi(\beta) = a$, and we have instead

$$\begin{aligned}\int_{[a,b]} f(x)\,d\lambda_1(x) &= -\int_{[\alpha,\beta]} f(\varphi(y))\,\varphi'(y)\,d\lambda_1(y)\\ &= \int_{[\alpha,\beta]} f(\varphi(y))\,|\varphi'(y)|\,d\lambda_1(y)\end{aligned}$$

where we have used that $\varphi' \leq 0$.

Hence, the formula (3.1) holds in the both cases when φ is an increasing or decreasing diffeomorphism. The next theorem generalizes (3.1) for higher dimensions.

Theorem 3.1 (Transformationsatz) *Let $\mu = \lambda_n$ be the Lebesgue measure in $\mathbb{R}^n$. Let U, V be open subsets of $\mathbb{R}^n$ and $\Phi : U \to V$ be a diffeomorphism. Then, for any non-negative measurable function $f : V \to \mathbb{R}$, the function $f \circ \Phi : U \to \mathbb{R}$ is measurable and*

$$\int_V f\,d\mu = \int_U (f \circ \Phi)\,|\det \Phi'|\;d\mu. \tag{3.2}$$

The same holds for any integrable function $f : V \to \mathbb{R}$.

Recall that $\Phi : U \to V$ is a diffeomorphism if the inverse mapping $\Phi^{-1} : V \to U$ exists and both Φ and Φ' are continuously differentiable. Recall also that Φ' is the total derivative of Φ, which in this setting coincides with the Jacobi matrix, that is,

$$\Phi' = J_\Phi = \left(\frac{\partial \Phi_i}{\partial x_j}\right)_{i,j=1}^n.$$

Let us rewrite (3.2) in the form matching (3.1):

$$\int_V f(x)\;d\mu(x) = \int_U f(\Phi(y))\,|\det \Phi'(y)|\;d\mu(y)\,. \tag{3.3}$$

The formula (3.3) can be explained and memorized as follows. In order to evaluate $\int_V f(x)\, d\mu(x)$, one finds a suitable substitution $x = \Phi(y)$, which maps one-to-one $y \in U$ to $x \in V$, and substitutes x via $\Phi(y)$, using the rule

$$d\mu(x) = |\det \Phi'(y)|\, d\mu(y).$$

Using the notation $\frac{dx}{dy}$ instead of $\Phi'(y)$, this rule can also be written in the form

$$\frac{d\mu(x)}{d\mu(y)} = \left|\det \frac{dx}{dy}\right|, \tag{3.4}$$

which is easy to remember. Although presently the identity (3.4) has no rigorous meaning, it can be turned into a theorem using a proper definition of the quotient $d\mu(x)/d\mu(y)$.

Example. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping that is, $\Phi(y) = Ay$ where A is a non-singular $n \times n$ matrix. Then $\Phi'(y) = A$ and (3.2) becomes

$$\int_{\mathbb{R}^n} f(x)\, d\mu = |\det A| \int_{\mathbb{R}^n} f(Ay)\, d\mu(y).$$

In particular, applying this to $f = 1_S$ where S is a measurable subset of $\mathbb{R}^n$, we obtain

$$\mu(S) = |\det A|\, \mu\left(A^{-1}S\right)$$

or, renaming $A^{-1}S$ to S,

$$\mu(AS) = |\det A|\, \mu(S). \tag{3.5}$$

In particular, if $|\det A| = 1$ then $\mu(AS) = \mu(S)$ that is, the Lebesgue measure is preserved by such linear mappings. Since all orthogonal matrices have determinants ± 1, it follows that the Lebesgue measure is preserved by orthogonal transformations of $\mathbb{R}^n$, in particular, by rotations.

Example. Let S be the unit cube, that is,

$$S = \{x_1 e_1 + ... + x_n e_n : 0 < x_i < 1, i = 1, ..., n\},$$

where $e_1, ..., e_n$ the canonical basis in $\mathbb{R}^n$. Setting $a_i = Ae_i$, we obtain

$$AS = \{x_1 a_1 + ... + x_n a_n : 0 < x_i < 1, i = 1, ..., n\}.$$

This set is called a *parallelepiped* spanned by the vectors $a_1, ..., a_n$ (its particular case for $n = 2$ is a parallelogram). Denote it by $\Pi(a_1, ..., a_n)$. Note that the columns of the matrix A are exactly the column-vectors a_i. Therefore, we obtain that from (3.5)

$$\mu(\Pi(a_1, ..., a_n)) = |\det A| = |\det(a_1, a_2, ..., a_n)|.$$

Example. Consider a *tetrahedron* spanned by $a_1, ..., a_n$, that is,

$$T(a_1, ..., a_n) = \{x_1 a_1 + ... + x_n a_n : x_1 + x_2 + ... + x_n < 1,\ x_i > 0, i = 1, ..., n\}.$$

It is possible to prove that

$$\mu\left(T\left(a_1,...,a_n\right)\right)=\frac{1}{n!}\left|\det\left(a_1,a_2,...,a_n\right)\right|$$

(see Exercise 58).

Example. (The polar coordinates) Let (r,θ) be the polar coordinates in $\mathbb{R}^2$. The change from the Cartesian coordinates (x,y) to the polar coordinates (r,θ) can be regarded as a mapping

$$\Phi : U \to \mathbb{R}^2$$

where $U=\{(r,\theta) : r>0,\ -\pi<\theta<\pi\}$ and

$$\Phi\left(r,\theta\right)=\left(r\cos\theta,r\sin\theta\right).$$

The image of U is the open set

$$V=\mathbb{R}^2\setminus\left\{(x,0):x\leq 0\right\},$$

and Φ is a diffeomorphism between U and V. We have

$$\Phi'=\begin{pmatrix}\partial_r\Phi_1 & \partial_\theta\Phi_1\\ \partial_r\Phi_2 & \partial_\theta\Phi_2\end{pmatrix}=\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix}$$

whence

$$\det\Phi'=r.$$

Hence, we obtain the formula for computing the integrals in the polar coordinates:

$$\int_V f\left(x,y\right)\,d\lambda_2\left(x,y\right)=\int_U f\left(r\cos\theta,r\sin\theta\right)r\lambda_2\left(r,\theta\right). \tag{3.6}$$

If function f is given on $\mathbb{R}^2$ then

$$\int_{\mathbb{R}^2} f d\lambda_2=\int_V f d\lambda_2$$

because the difference $\mathbb{R}^2\setminus V$ has measure 0. Also, using Fubini's theorem in U, we can express the right hand side of (3.6) via repeated integrals and, hence, obtain

$$\int_{\mathbb{R}^2} f\left(x,y\right)\,d\lambda_2\left(x,y\right)=\int_0^\infty\left(\int_{-\pi}^{\pi} f\left(r\cos\theta,r\sin\theta\right)d\theta\right)rdr. \tag{3.7}$$

For example, if $f=1_D$ where $D=\{(x,y):x^2+y^2<R^2\}$ is the disk of radius R then

$$\lambda_2\left(D\right)=\int_{\mathbb{R}^2}1_D\left(x,y\right)d\lambda_2=\int_0^R\left(\int_{-\pi}^{\pi}d\theta\right)rdr=2\pi\int_0^R rdr=\pi R^2.$$

Example. Let us evaluate the integral

$$I=\int_{-\infty}^{\infty}e^{-x^2}dx.$$

Using Fubini's theorem, we obtain

$$\begin{aligned}
I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right) \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right) e^{-x^2} dx \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-\left(x^2+y^2\right)} dy\right) dx \\
&= \int_{\mathbb{R}^2} e^{-\left(x^2+y^2\right)} d\lambda_2 (x, y).
\end{aligned}$$

Then by (3.7)

$$\int_{\mathbb{R}^2} e^{-\left(x^2+y^2\right)} d\lambda_2 = \int_0^{\infty}\left(\int_{-\pi}^{\pi} e^{-r^2} d\theta\right) r dr = 2\pi \int_0^{\infty} e^{-r^2} r dr = \pi \int_0^{\infty} e^{-r^2} dr^2 = \pi.$$

Hence, $I^2 = \pi$ and $I = \sqrt{\pi}$, that is,

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

The proof of Theorem 3.1 will be preceded by a number of auxiliary claims.

We say that a mapping Φ is *good* if (3.2) holds for all non-negative measurable functions. Theorem 3.1 will be proved if we show that all diffeomorphisms are good.

Claim 0. *Φ is good if and only if for any measurable set $B \subset V$, the set $A = \Phi^{-1}(B)$ is also measurable and*

$$\mu(B) = \int_A |\det \Phi'| \, d\mu. \tag{3.8}$$

Proof. Applying (3.2) for indicator functions, that is, for functions $f = 1_B$ where B is a measurable subset of V, we obtain that the function $1_A = 1_B \circ \Phi$ is measurable and

$$\mu(B) = \int_U 1_B(\Phi(y)) |\det \Phi'(y)| \, d\mu(y) = \int_U 1_A(y) |\det \Phi'| \, d\mu = \int_A |\det \Phi'| \, d\mu. \tag{3.9}$$

Conversely, if (3.8) is true then (3.2) is true for indicators functions $f = 1_B$. By the linearity of the integral and the monotone convergence theorem for series, we obtain (3.2) for simple functions:

$$f = \sum_{k=1}^{\infty} b_k 1_{B_k}.$$

Finally, any non-negative measurable function f is an increasing pointwise limit of simple functions, whence (3.2) follows for f. ■

The idea of the proof of (3.8). We will first prove it for affine mappings, that is, the mappings of the form

$$\Psi(x) = Cx + D$$

where C is a constant $n \times n$ matrix and $D \in \mathbb{R}^n$. A general diffeomorphism Φ can be approximated in a neighborhood of any point $x_0 \in U$ by the tangent mapping

$$\Psi_{x_0}(x) = \Phi(x_0) + \Phi'(x_0)(x - x_0),$$

which is affine. This implies that if A is a set in a small neighborhood of x_0 then

$$\mu\left(\Phi\left(A\right)\right) \approx \mu\left(\Psi_{x_0}\left(A\right)\right) = \int_A \left|\det \Psi'_{x_0}\right| d\mu \approx \int_A \left|\det \Phi'\right| \, d\mu.$$

Now split an arbitrary measurable set $A \subset U$ into small pieces $\{A_k\}_{k=1}^{\infty}$ and apply the previous approximate identity to each A_k:

$$\mu\left(\Phi\left(A\right)\right) = \sum_k \mu\left(\Phi\left(A_k\right)\right) \approx \sum_k \int_{A_k} \left|\det \Phi'\right| \, d\mu = \int_A \left|\det \Phi'\right| \, d\mu.$$

The error of approximation can be made arbitrarily small using finer partitions of A. ■

Claim 1. *If $\Phi : U \to V$ and $\Psi : V \to W$ are good mappings then $\Psi \circ \Phi$ is also good.*

Proof. Let $f : W \to \mathbb{R}$ be a non-negative measurable function. We need to prove that

$$\int_W f \, d\mu = \int_U \left(f \circ \Psi \circ \Phi\right) \left|\det \left(\Psi \circ \Phi\right)'\right| d\mu.$$

Since Ψ is good, we have

$$\int_W f \, d\mu = \int_V \left(f \circ \Psi\right) \left|\det \Psi'\right| d\mu.$$

Set $g = f \circ \Psi \left|\det \Phi'\right|$ so that g is a function on V. Since Φ is good, we have

$$\int_V g \, d\mu = \int_U \left(g \circ \Phi\right) \left|\det \Phi'\right| d\mu.$$

Combining these two lines and using the chain rule in the form

$$\left(\Psi \circ \Phi\right)' = \left(\Psi' \circ \Phi\right) \Phi'$$

and it its consequence

$$\det \left(\Psi \circ \Phi\right)' = \det \left(\Psi' \circ \Phi\right) \det \Phi',$$

we obtain

$$\begin{aligned} \int_W f \, d\mu &= \int_U \left(f \circ \Psi \circ \Phi\right) \left|\det \Psi' \circ \Phi\right| \left|\det \Phi'\right| d\mu \\ &= \int_U \left(f \circ \Psi \circ \Phi\right) \left|\det \left(\Psi \circ \Phi\right)'\right| d\mu, \end{aligned}$$

which was to be proved ■

Claim 2. *All translations $\mathbb{R}^n$ are good.*

Proof. A translation of $\mathbb{R}^n$ is a mapping of the form $\Phi\left(x\right) = x + v$ where v is a constant vector from $\mathbb{R}^n$. Obviously, $\Phi' = \mathrm{id}$ and $\det \Phi' \equiv 1$. Hence, the fact that Φ is good amounts by (3.8) to

$$\mu\left(B\right) = \mu\left(B - \nu\right),$$

for any measurable subset $B \subset \mathbb{R}^n$. This is obviously true when B is a box. Then following the procedure of construction of the Lebesgue measure, we obtain that this is true for any measurable set B. ■

Note that the identity (3.2) for translations amounts to

$$\int_{\mathbb{R}^n} f(x)\, d\mu(x) = \int_{\mathbb{R}^n} f(y+v)\, d\mu(y), \tag{3.10}$$

for any non-negative measurable function f on $\mathbb{R}^n$.

Claim 3. *All homotheties of* $\mathbb{R}^n$ *are good.*

Proof. A homothety of $\mathbb{R}^n$ is a mapping $\Phi(x) = cx$ where c is a non-zero real. Since $\Phi' = c\,\mathrm{id}$ and, hence, $\det \Phi' = c^n$, the fact that Φ is good is equivalent to the identity

$$\mu(B) = |c|^n \mu\left(c^{-1}B\right), \tag{3.11}$$

for any measurable set $B \subset \mathbb{R}^n$. This is true for boxes and then extends to all measurable sets by the measure extension procedure. ■

The identity (3.2) for homotheties amounts to

$$\int_{\mathbb{R}^n} f(x)\, d\mu(x) = |c|^n \int_{\mathbb{R}^n} f(cy)\, d\mu(y), \tag{3.12}$$

for any non-negative measurable function f on $\mathbb{R}^n$.

Claim 4. *All linear mappings are good.*

Proof. Let Φ be a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. Extending a function f, initially defined on V, to the whole $\mathbb{R}^n$ by setting $f = 0$ in $\mathbb{R}^n \setminus V$, we can assume that $V = U = \mathbb{R}^n$. Let us use the fact from linear algebra that any non-singular linear mapping can be reduced by column-reduction of the matrices to the composition of finitely many *elementary* linear mappings, where an elementary linear mapping is one of the following:

1. $\Phi(x_1, \ldots, x_n) = (cx_1, x_2, \ldots, x_n)$ for some $c \neq 0$;
2. $\Phi(x_1, \ldots, x_n) = (x_1 + cx_2, x_2, \ldots, x_n)$;
3. $\Phi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n)$ (switching the variables x_i and x_j).

Hence, in the view of Claim 1, it suffices to prove Claim 4 if Φ is an elementary mapping. If Φ is of the first type then, using Fubini's theorem and setting $\lambda = \lambda_1$, we can write

$$\begin{aligned}
\int_{\mathbb{R}^n} f\, d\mu &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, ..., x_n)\, d\lambda(x_1) \right) d\lambda(x_2) \ldots d\lambda(x_n) \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(ct, ..., x_n)\, |c|\, d\lambda(t) \right) d\lambda(x_2) \ldots d\lambda(x_n) \\
&= \int_{\mathbb{R}^n} f(cx_1, x_2, \ldots, x_n)\, |c|\, d\mu \\
&= \int_{\mathbb{R}^n} f(\Phi(x))\, |\det \Phi'|\, d\mu,
\end{aligned}$$

where we have used the change of integral under a homothety in $\mathbb{R}^1$ and $\det \Phi' = c$.

If Φ is of the second type then

$$\begin{aligned}
\int_{\mathbb{R}^n} f\,d\mu &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, ..., x_n)\, d\lambda(x_1) \right) d\lambda(x_2) \dots d\lambda(x_n) \\
&= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1 + cx_2, ..., x_n)\, d\lambda(x_1) \right) d\lambda(x_2) \dots d\lambda(x_n) \\
&= \int_{\mathbb{R}^n} f(x_1 + cx_2, x_2, \dots, x_n)\, d\mu \\
&= \int_{\mathbb{R}^n} f(\Phi(x)) \left|\det \Phi'\right|\, d\mu,
\end{aligned}$$

where we have used the translation invariance of the integral in $\mathbb{R}^1$ and $\det \Phi' = 1$.

Let Φ be of the third type. For simplicity of notation, let

$$\Phi(x_1, x_2, \dots, x_n) = (x_2, x_1, \dots, x_n).$$

Then, using twice Fubini's theorem with different orders of the repeated integrals, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^n} f\,d\mu &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, x_2, ..., x_n)\, d\lambda(x_2) \right) d\lambda(x_1) \dots d\lambda(x_n) \\
&= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y_2, y_1, y_3.., y_n)\, d\lambda(y_1) \right) d\lambda(y_2) \dots d\lambda(y_n) \\
&= \int_{\mathbb{R}^n} f(y_2, y_1, \dots, y_n)\, d\mu(y) \\
&= \int_{\mathbb{R}^n} f(\Phi(y)) \left|\det \Phi'\right| d\mu.
\end{aligned}$$

In the second line we have changed notation $y_1 = x_2$, $y_2 = x_1$, $y_k = x_k$ for $k > 2$ (let us emphasize that this is not a change of variable but just a change of notation in each of the repeated integrals), and in the last line we have used $\det \Phi' = -1$. ■

It follows from Claims 1, 2 and 4 that all affine mappings are good.

It what follows we will use the ∞-norm in $\mathbb{R}^n$:

$$\|x\| = \|x\|_\infty := \max_{1 \le i \le n} |x_i|.$$

Let Φ be a diffeomorphism between open sets U and V in $\mathbb{R}^n$. Let K be a compact subset of U.

Claim 5. *For any $\eta > 0$ there is $\delta > 0$ such that, for any interval $[x, y] \subset K$ with the condition $\|y - x\| < \delta$, the following is true:*

$$\|\Phi(y) - \Phi(x) - \Phi'(x)(y - x)\| \le \eta \|y - x\|.$$

Proof. Consider the function

$$f(t) = \Phi((1 - t)x + ty), \quad 0 \le t \le 1,$$

so that $f(0)=\Phi(x)$ and $f(1)=\Phi(y)$. Since the interval $[x,y]$ is contained in U, the function f is defined for all $t\in[0,1]$. By the fundamental theorem of calculus, we have

$$\Phi(y)-\Phi(x)=\int_0^1 f'(t)\,dt=\int_0^1 \Phi'((1-t)x+ty)(y-x)\,dt.$$

Since the function $\Phi'(x)$ is uniformly continuous on K, δ can be chosen so that

$$x,z\in K,\ \|z-x\|<\delta\Rightarrow\|\Phi(z)-\Phi(x)\|<\eta.$$

Applying this with $z=(1-t)x+ty$ and noticing that $\|z-x\|\le\|y-x\|<\delta$, we obtain

$$\|\Phi'((1-t)x+ty)-\Phi'(x)\|<\eta.$$

Denoting

$$F(t)=\Phi'((1-t)x+ty)-\Phi'(x),$$

we obtain

$$\begin{aligned}\Phi(y)-\Phi(x) &= \int_0^1(\Phi'(x)+F(t))(y-x)\,dt\\ &= \Phi'(x)(y-x)+\int_0^1 F(t)(y-x)\,dt\end{aligned}$$

and

$$\|\Phi(y)-\Phi(x)-\Phi'(x)(y-x)\|=\|\int_0^1 F(t)(y-x)\|dt\le\eta\|y-x\|.$$

■

For any $x\in U$, denote by $\Psi_x(y)$ the tangent mapping of Φ at x, that is,

$$\Psi_x(y)=\Phi(x)+\Phi'(x)(y-x). \tag{3.13}$$

Clearly, Ψ_x is a non-singular affine mapping in $\mathbb{R}^n$. In particular, Ψ_x is good. Note also that $\Psi_x(x)=\Phi(x)$ and $\Psi_x'\equiv\Phi'(x)$.

The Claim 5 can be stated as follows: for any $\eta>0$ there is $\delta>0$ such that if $[x,y]\subset K$ and $\|y-x\|<\delta$ then

$$\|\Phi(y)-\Psi_x(y)\|\le\eta\|y-x\|. \tag{3.14}$$

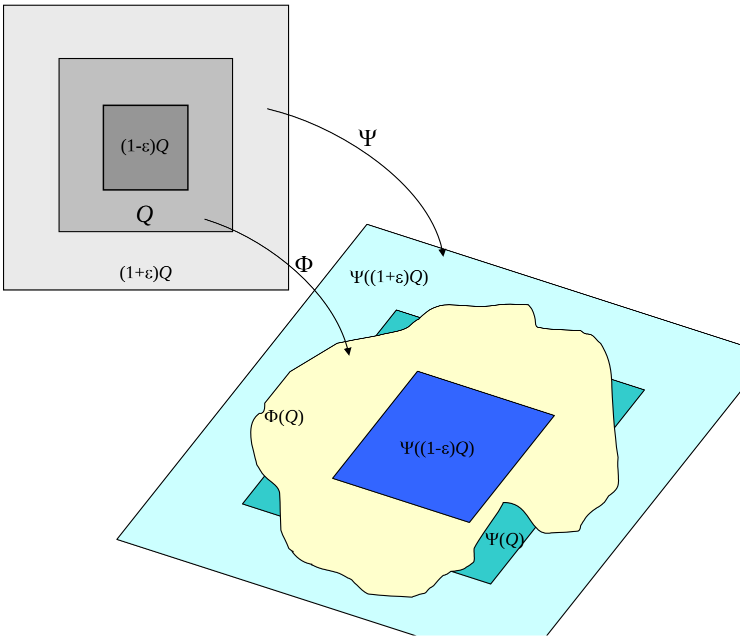
Consider a metric balls $Q_r(x)$ with respect to the ∞-norm, that is,

$$Q_r(x)=\{y\in\mathbb{R}^n:\|y-x\|<r\}.$$

Clearly, $Q_r(x)$ is the cube with the center x and with the edge length $2r$. If $Q=Q_r(x)$ and $c>0$ then set $cQ=Q_{cr}(x)$.

Claim 6. *For any $\varepsilon\in(0,1)$, there is $\delta>0$ such that, for any cube $Q=Q_r(x)\subset K$ with radius $0<r<\delta$,*

$$\Psi_x((1-\varepsilon)Q)\subset\Phi(Q)\subset\Psi_x((1+\varepsilon)Q). \tag{3.15}$$

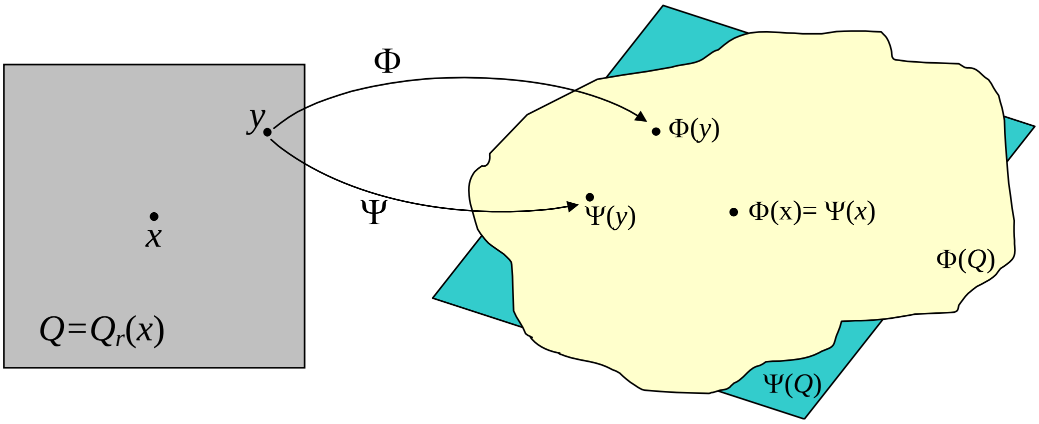


The set $\Phi(Q)$ is "squeezed" between $\Psi((1-\varepsilon)Q)$ and $\Psi((1+\varepsilon)Q)$

Proof. Fix some $\eta > 0$ (which will be chosen below as a function of ε) and let $\delta > 0$ be the same as in Claim 5. Let Q be a cube as in the statement. For any $y \in Q$, we have $\|y - x\| < r < \delta$ and $[x, y] \subset Q \subset K$, which implies (3.14), that is,

$$\|\Phi(y) - \Psi(y)\| < \eta r, \tag{3.16}$$

where $\Psi = \Psi_x$.



Mappings Φ and Ψ

Now we will apply Ψ^{-1} to the difference $\Phi(y) - \Psi(y)$. Note that by (3.13), for any $a \in \mathbb{R}^n$,

$$\Psi^{-1}(a) = \Phi'(x)^{-1}(a - \Phi(x)) + x$$

whence it follows that, for all $a, b \in \mathbb{R}^n$,

$$\Psi^{-1}(a) - \Psi^{-1}(b) = \Phi'(x)^{-1}(a - b). \tag{3.17}$$

By the compactness of K, we have

$$C := \sup_K \| (\Phi')^{-1} \| < \infty.$$

It follows from (3.17) that

$$\left\| \Psi^{-1}(a) - \Psi^{-1}(b) \right\| \leq C \|a - b\|$$

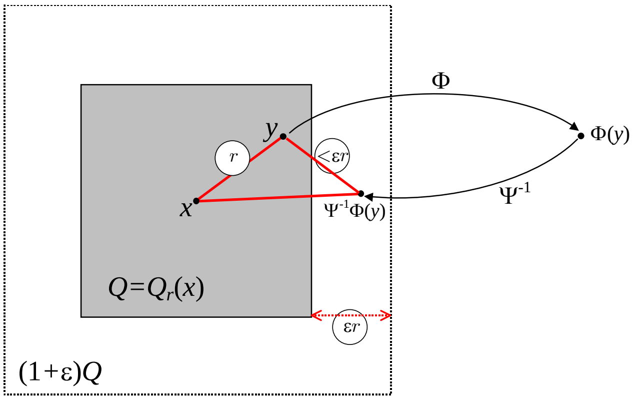
(it is important that the constant C does not depend on a particular choice of $x \in K$ but is determined by the set K).

In particular, setting here $a = \Phi(y)$, $b = \Psi(y)$ and using (3.16), we obtain

$$\left\| \Psi^{-1}\Phi(y) - y \right\| \leq C \|\Phi(y) - \Psi(y)\| < C\eta r.$$

Now we choose η to satisfy $C\eta = \varepsilon$ and obtain

$$\|\Psi^{-1}\Phi(y) - y\| < \varepsilon r. \tag{3.18}$$



Comparing the points $y \in Q$ and $\Psi^{-1}\Phi(y)$

It follows that

$$\|\Psi^{-1}\Phi(y) - x\| \leq \|\Psi^{-1}\Phi(y) - y\| + \|y - x\| < \varepsilon r + r = (1 + \varepsilon) r$$

whence

$$\Psi^{-1}\Phi(y) \in Q_{(1+\varepsilon)r}(x) = (1 + \varepsilon) Q$$

and

$$\Phi(Q) \subset \Psi((1 + \varepsilon) Q),$$

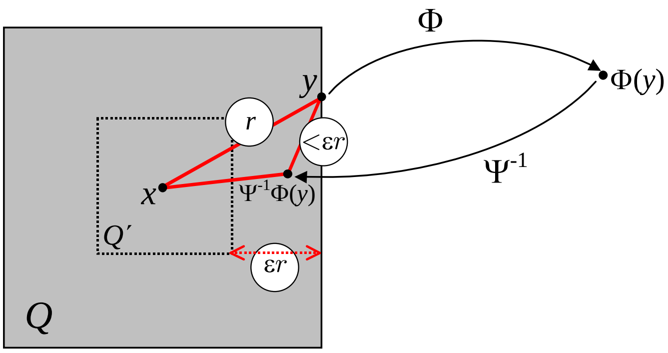
which is the right inclusion in (3.15).

To prove the left inclusion in (3.15), observe that, for any $y \in \partial Q := \overline{Q} \setminus Q$, we have $\|y - x\| = r$ and, hence,

$$\|\Psi^{-1}\Phi(y) - x\| \geq \|y - x\| - \|\Psi^{-1}\Phi(y) - y\| > r - \varepsilon r = (1 - \varepsilon) r,$$

that is,

$$\Psi^{-1}\Phi(y) \notin (1 - \varepsilon) Q =: Q'.$$



Comparing the points $y \in \partial Q$ and $\Psi^{-1}\Phi(y)$

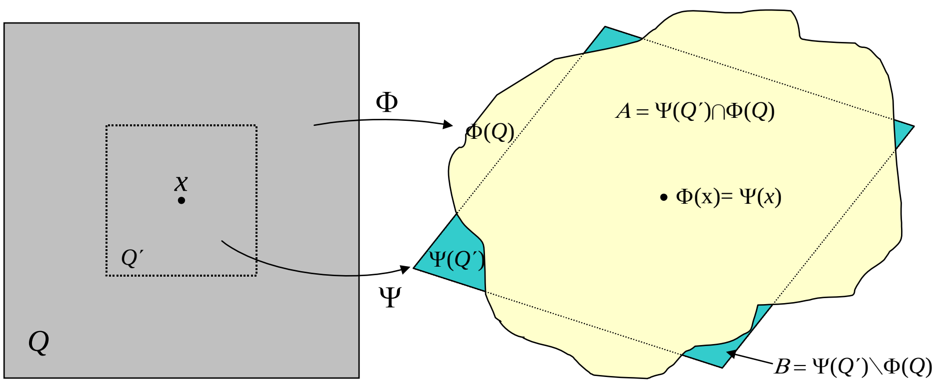
It follows that $\Phi(y) \notin \Psi(Q')$ for any $y \in \partial Q$, that is,

$$\Psi(Q') \cap \Phi(\partial Q) = \emptyset. \tag{3.19}$$

Consider two sets

$$\begin{aligned} A &= \Psi(Q') \cap \Phi(Q), \\ B &= \Psi(Q') \setminus \Phi(Q), \end{aligned}$$

which are obviously disjoint and their union is $\Psi(Q')$.



Sets A and B (the latter is shown as non-empty although it is in fact empty)

Observe that A is open as the intersection of two open sets[1]. Using (3.19) we can write

$$B = \Psi(Q') \setminus \Phi(Q) \setminus \Phi(\partial Q) = \Psi(Q') \setminus \Phi\left(\overline{Q}\right).$$

Then B is open as the difference of an open and closed set. Hence, the open set $\Psi(Q')$ is the disjoint union of two open sets A, B. Since $\Psi(Q')$ is connected as a continuous image of a connected set Q', we conclude that one of the sets A, B is $\emptyset$ and the other is $\Psi(Q')$.

[1] Recall that diffeomorphisms map open sets to open and closed sets to closed.

Note that A is non-empty because A contains the point $\Phi(x) = \Psi(x)$. Therefore, we conclude that $A = \Psi(Q')$, which implies

$$\Phi(Q) \supset \Psi(Q').$$

■

Claim 7. *If set $S \subset U$ has measure 0 then also the set $\Phi(S)$ has measure 0.*

Proof. Exhausting U be a sequence of compact sets, we can assume that S is contained in some compact set K such that $K \subset U$. Choose some $\rho > 0$ and denote by K_ρ the closed ρ-neighborhood of K, that is, $K_\rho = \{x \in \mathbb{R}^n : \|x - K\| \le \rho\}$. Clearly, ρ be can taken so small $K_\rho \subset U$.

The hypothesis $\mu(S) = 0$ implies that, for any $\varepsilon > 0$, there is a sequence $\{B_k\}$ of boxes such that $S \subset \bigcup_k B_k$ and

$$\sum_k \mu(B_k) < \varepsilon. \tag{3.20}$$

For any $r > 0$, each box B in $\mathbb{R}^n$ can be covered by a sequence of cubes of radii $\le r$ and such that the sum of their measures is bounded by $2\mu(B)$[2]. Find such cubes for any box B_k from (3.20), and denote by $\{Q_j\}$ the sequence of all the cubes across all B_k. Then the sequence $\{Q_j\}$ covers S, and

$$\sum_j \mu(Q_j) < 2\varepsilon.$$

Clearly, we can remove from the sequence $\{Q_j\}$ those cubes that do not intersect S, so that we can assume that each Q_j intersects S and, hence, K. Choosing r to be smaller than $\rho/2$, we obtain that all Q_j are contained in K_ρ.

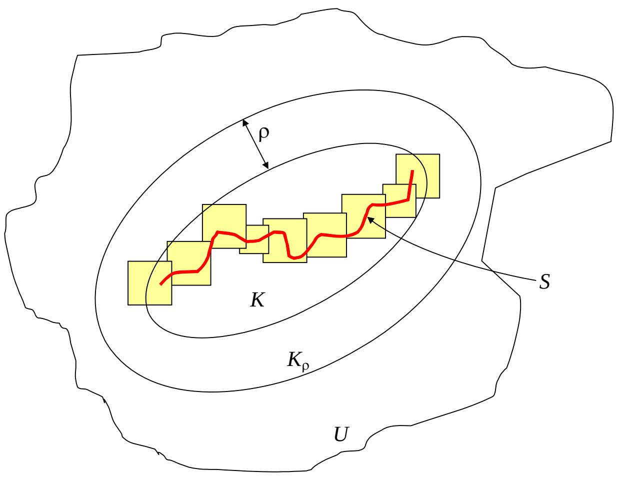


Figure 1: Covering set S of measure 0 by cubes of radii $< \rho/2$

By Claim 6, there is $\delta > 0$ such that, for any cube $Q \subset K_\rho$ of radius $< \delta$ and center x,

$$\Phi(Q) \subset \Psi_x(2Q).$$

[2]Indeed, consider the cubes whose vertices have the coordinates that are multiples of a, where a is sufficiently small. Then choose cubes those that intersect B. Clearly, the chosen cubes cover B and their total measure can be made arbitrarily close to $\mu(B)$ provided a is small enough.

Hence, assuming from the very beginning that $r < \delta$ and denoting by x_j the center of Q_j, we obtain

$$\Phi(Q_j) \subset \Psi_{x_j}(2Q_j) =: A_j.$$

By Claim 4, we have

$$\mu(A_j) = \left|\det \Psi'_{x_k}\right| \mu(2Q_j) = \left|\det \Phi'(x_k)\right| 2^n \mu(Q_j).$$

By the compactness of K_ρ,

$$C := \sup_{K_\rho} \left|\det \Phi'\right| < \infty.$$

It follows that

$$\mu(A_j) \le C2^n \mu(Q_j)$$

and

$$\sum_j \mu(A_j) \le C2^n \sum_k \mu(Q_j) \le C2^{n+1}\varepsilon.$$

Since $\{A_j\}$ covers S, we obtain by the sub-additivity of the outer measure that $\mu^*(S) \le C2^{n+1}\varepsilon$. Since $\varepsilon > 0$ is arbitrary, we obtain $\mu^*(S) = 0$, which was to be proved. ■

Claim 8. *For any diffeomorphism $\Phi : U \to V$ and for any measurable set $A \subset U$, the set $\Phi(A)$ is also measurable.*

Proof. Any measurable set $A \subset \mathbb{R}^n$ can be represented in the form $B - C$ where B is a Borel set and C is a set of measure 0 (see Exercise 54). Then $\Phi(A) = \Phi(B) - \Phi(C)$. By Lemma 2.1, the set $\Phi(B)$ is Borel, and by Claim 7, the set $\Phi(C)$ has measure 0. Hence, $\Phi(A)$ is measurable. ■

Proof of Theorem 3.1. By Claim 0, it suffices to prove that, for any measurable set $B \subset V$, the set $A = \Phi^{-1}(B)$ is measurable and

$$\mu(B) = \int_A \left|\det \Phi'\right| d\mu.$$

Applying Claim 8 to Φ and Φ^{-1}, we obtain that A is measurable if and only if B is measurable. Hence, it suffices to prove that, for any measurable set $A \subset U$,

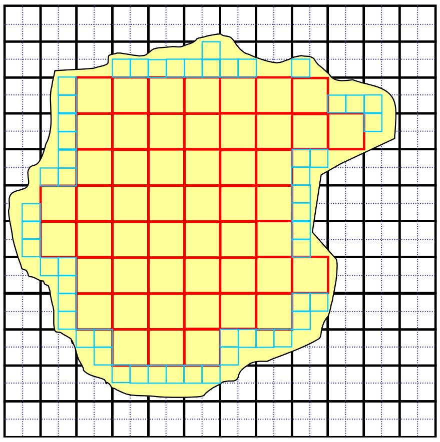
$$\mu(\Phi(A)) = \int_A \left|\det \Phi'\right| d\mu, \tag{3.21}$$

By the monotone convergence theorem, the identity (3.21) is stable under taking monotone limits for increasing sequences of sets A. Therefore, exhausting U by compact sets, we can reduce the problem to the case when the set A is contained in such a compact set, say K. Since $\left|\det \Phi'\right|$ is bounded on K and, hence, is integrable, it follows from the dominated convergence theorem that (3.21) is stable under taking monotone decreasing limits as well. Clearly, (3.21) survives also taking monotone difference of sets.

Now we use the facts that any measurable set is a monotone difference of a Borel set and a set of measure 0 (Exercise 54), and the Borel sets are obtained from the open sets by monotone differences and monotone limits (Theorem 1.14). Hence, it suffices to prove (3.21) for two cases: if set A has measure 0 and if A is an open set. For the sets of measure 0 (3.21) immediately follows from Claim 7 since the both sides of (3.21) are 0.

So, assume in the sequel that A is an open set. We claim that, for any open set, and in particular for A, and for any $\rho > 0$, there is a disjoint sequence of cubes $\{Q_k\}$ or radii $\leq \rho$ such that $Q_k \subset A$ and

$$\mu\left(A \setminus \bigcup_k Q_k\right) = 0.$$



Using dyadic grids with steps $2^{-k}\rho$, $k = 0, 1, ...$, to split an open set A into disjoint union of cubes modulo a set of measure zero.

Furthermore, if ρ is small enough then, by Claim 6, we have

$$\Psi_{x_k}\left((1-\varepsilon) Q_k\right) \subset \Phi\left(Q_k\right) \subset \Psi_{x_k}\left((1+\varepsilon) Q_k\right),$$

where x_k is the center of Q_k and $\varepsilon > 0$ is any given number. By the uniform continuity of $\ln\left|\det \Phi'(x)\right|$ on K, if ρ is small enough then, for $x, y \in K$,

$$\|x - y\| < \rho \Rightarrow \ 1 - \varepsilon < \frac{\left|\det \Phi'(y)\right|}{\left|\det \Phi'(x)\right|} < 1 + \varepsilon.$$

Since

$$\begin{aligned}
\mu\left(\Psi_{x_k}\left((1+\varepsilon) Q_k\right)\right) &= \left|\det \Psi'_{x_k}\right| (1+\varepsilon)^n \mu\left(Q_k\right) \\
&= \left|\det \Phi'\left(x_k\right)\right| (1+\varepsilon)^n \mu\left(Q_k\right) \\
&\leq (1+\varepsilon)^{n+1} \int_{Q_k} \det\left|\Phi'(x)\right| d\mu,
\end{aligned}$$

it follows that

$$\mu\left(\Phi(A)\right) = \sum_k \mu\left(\Phi\left(Q_k\right)\right) \leq \sum_k (1+\varepsilon)^{n+1} \int_{Q_k} \det\left|\Phi'(x)\right| d\mu = (1+\varepsilon)^{n+1} \int_A \left|\det \Phi'\right| d\mu.$$

Similarly, one proves that

$$\mu\left(\Phi(A)\right) \geq (1-\varepsilon)^{(n+1)} \int_A \left|\det \Phi'\right| d\mu.$$

Since ε can be made arbitrarily small, we obtain (3.21). $\blacksquare$

3.2 Random variables and their distributions

Recall that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a triple of a set Ω, a σ-algebra $\mathcal{F}$ (also frequently called a σ-field) and a probability measure $\mathbb{P}$ on $\mathcal{F}$, that is, a measure with the total mass 1 (that is, $\mathbb{P}(\Omega) = 1$). Recall also that a function $X : \Omega \to \mathbb{R}$ is called measurable with respect to $\mathcal{F}$ if for any $x \in \mathbb{R}$ the set

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$$

is measurable, that is, belong to $\mathcal{F}$. In other words, $\{X \leq x\}$ is an event.

Definition. Any measurable function X on Ω is called a *random variable* (Zufallsgröße).

For any event X and any $x \in \mathbb{R}$, the probability $\mathbb{P}(X \leq x)$ is defined, which is referred to as the probability that the random variable X is bounded by x.

For example, if A is any event, then the indicator function $\mathbf{1}_A$ on Ω is a random variable.

Fix a random variable X on Ω. Recall that by Lemma 2.1 for any Borel set $A \subset \mathbb{R}$, the set

$$\{X \in A\} = X^{-1}(A)$$

is measurable. Hence, the set $\{X \in A\}$ is an event and we can consider the probability $\mathbb{P}(X \in A)$ that X is in A. Set for any Borel set $A \subset \mathbb{R}$,

$$P_X(A) = \mathbb{P}(X \in A).$$

Then we obtain a real-valued functional $P_X(A)$ on the σ-algebra $\mathcal{B}(\mathbb{R})$ of all Borel sets in $\mathbb{R}$.

Lemma 3.2 *For any random variable X, P_X is a probability measure on $\mathcal{B}(\mathbb{R})$. Conversely, given any probability measure μ on $\mathcal{B}(\mathbb{R})$, there exists a probability space and a random variable X on it such that $P_X = \mu$.*

Proof. We can write $P_X(A) = \mathbb{P}(X^{-1}(A))$. Since X^{-1} preserves all set-theoretic operations, by this formula the probability measure $\mathbb{P}$ on $\mathcal{F}$ induces a probability measure on $\mathcal{B}$. For example, check additivity:

$$P_X(A \sqcup B) = \mathbb{P}(X^{-1}(A \sqcup B)) = \mathbb{P}(X^{-1}(A) \sqcup X^{-1}(B)) = \mathbb{P}(X^{-1}(A)) + \mathbb{P}(X^{-1}(B)).$$

The σ-additivity is proved in the same way. Note also that

$$P_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1.$$

For the converse statement, consider the probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}, \mu)$$

and the random variable on it

$$X(x) = x.$$

Then

$$P_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(x : X(x) \in A) = \mu(A).$$

$\blacksquare$

Hence, each random variable X induces a probability measure P_X on real Borel sets. The measure P_X is called *the distribution* of X. In probabilistic terminology, $P_X(A)$ is the probability that the value of the random variable X occurs to be in A.

Any probability measure on $\mathcal{B}$ is also called *a distribution.* As we have seen, any distribution is the distribution of some random variable.

Consider examples distributions on $\mathbb{R}$. There is a large class of distributions possessing a *density*, which can be described as follows. Let f be a non-negative Lebesgue integrable function on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} f d\lambda = 1, \tag{3.22}$$

where $\lambda = \lambda_1$ is the one-dimensional Lebesgue measure. Define the measure μ on $\mathcal{B}(\mathbb{R})$ by

$$\mu(A) = \int_A f d\lambda.$$

By Theorem 2.12, μ is a measure. Since $\mu(\mathbb{R}) = 1$, μ is a probability measure. Function f is called the *density* of μ or the *density function* of μ.

Here are some frequently used examples of distributions which are defined by their densities.

1. The uniform distribution $\mathcal{U}(I)$ on a bounded interval $I \subset \mathbb{R}$ is given by the density function $f = \frac{1}{\ell(I)} 1_I$.

2. The normal distribution $\mathcal{N}(a, b)$:

$$f(x) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b}\right)$$

where $a \in \mathbb{R}$ and $b > 0$. For example, $\mathcal{N}(0, 1)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

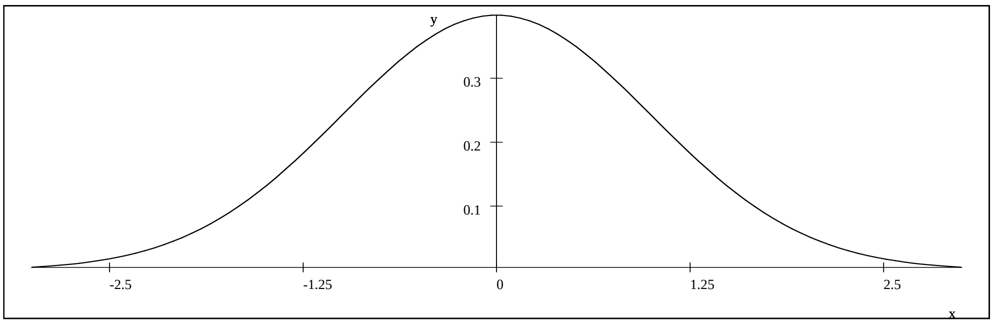
To verify total mass identity (3.22), use the change $y = \frac{x-a}{\sqrt{2b}}$ in the following integral:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b}\right) dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-y^2\right) dy = 1, \tag{3.23}$$

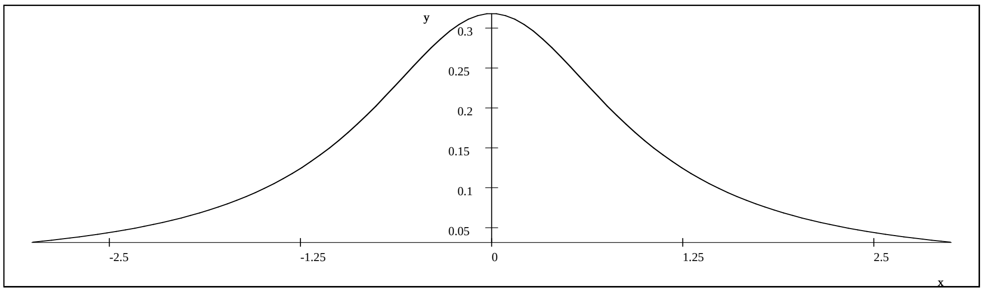
where the last identity follows from

$$\int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}$$

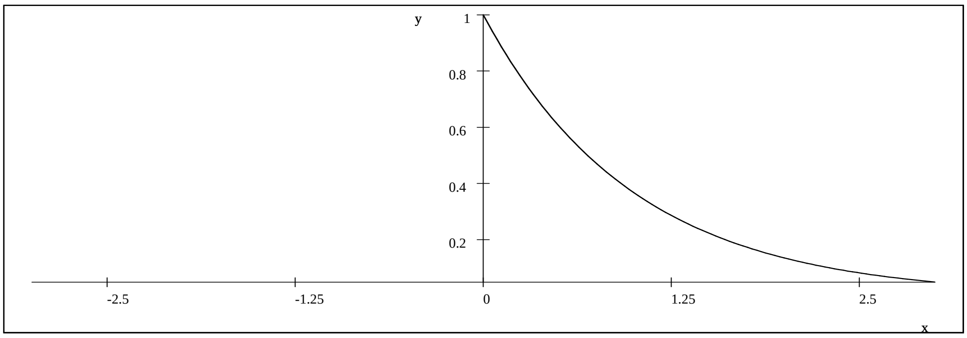
(see an Example in Section 3.1). Here is the plot of f for $\mathcal{N}(0, 1)$:



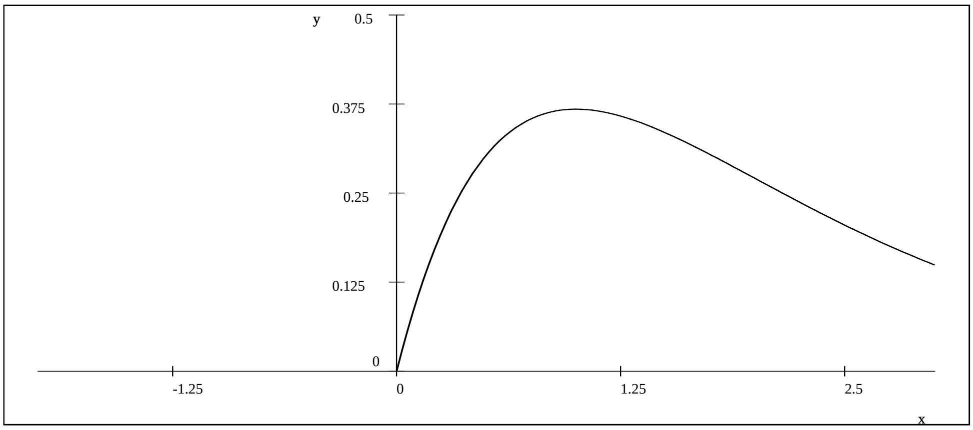
3. The Cauchy distribution:
$$f(x) = \frac{a}{\pi(x^2 + a^2)},$$
where $a > 0$. Here is the plot of f in the case $a = 1$:



4. The exponential distribution:
$$f(x) = \begin{cases} ae^{-ax}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$
where $a > 0$. The case $a = 1$ is plotted here:



5. The gamma distribution:
$$f(x) = \begin{cases} c_{a,b}x^{a-1}\exp(-x/b), & x > 0, \\ 0, & x \leq 0, \end{cases},$$
where $a > 1, b > 0$ and $c_{a,b}$ is chosen from the normalization condition (3.22), which gives
$$c_{a,b} = \frac{1}{\Gamma(a)b^a}.$$
For the case $a = 2$ and $b = 1$, we have $c_{a,b} = 1$ and $f(x) = xe^{-x}$, which is plotted below.



Another class of distributions consists of *discrete distributions* (or *atomic distributions*). Choose a finite or countable sequence $\{x_k\} \subset \mathbb{R}$ of distinct reals, a sequence $\{p_k\}$ of non-negative numbers such that

$$\sum_k p_k = 1, \tag{3.24}$$

and define for any set $A \subset \mathbb{R}$ (in particular, for any Borel set A)

$$\mu(A) = \sum_{x_k \in A} p_k.$$

By Exercise 2, μ is a measure, and by (3.24) μ is a probability measure. Clearly, measure μ is concentrated at points x_k which are called *atoms* of this measure. If X is a random variable with distribution μ then X takes the value x_k with probability p_k.

Here are two examples of such distributions, assuming $x_k = k$.

1. The binomial distribution $B(n, p)$ where $n \in \mathbb{N}$ and $0 < p < 1$:

   $$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, ..., n.$$

   The identity (3.24) holds by the binomial formula. In particular, if $n = 1$ then one gets Bernoulli's distribution

   $$p_0 = p, \quad p_1 = 1 - p.$$

2. The Poisson distribution $Po(\lambda)$:

   $$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, 2, \dots .$$

   where $\lambda > 0$. The identity (3.24) follows from the expansion of e^{λ} into the power series.

3.3 Functionals of random variables

Definition. If X is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then its *expectation* is defined by

$$\mathbb{E}\,(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega),$$

provided the integral in the right hand side exists (that is, either X is non-negative or X is integrable, that is $\mathbb{E}\,(|X|) < \infty$).

In other words, the notation $\mathbb{E}\,(X)$ is another (shorter) notation for the integral $\int_\Omega X d\mathbb{P}$.

By simple properties of Lebesgue integration and a probability measure, we have the following properties of $\mathbb{E}$:

1. $\mathbb{E}(X+Y) = \mathbb{E}\,(X) + \mathbb{E}\,(Y)$ provided all terms make sense;
2. $\mathbb{E}\,(\alpha X) = \alpha \mathbb{E}\,(X)$ where $\alpha \in \mathbb{R}$;
3. $\mathbb{E}1 = 1$.
4. $\inf X \le \mathbb{E}X \le \sup X$.
5. $|\mathbb{E}X| \le \mathbb{E}\,|X|$.

Definition. If X is an integrable random variable then its *variance* is defined by

$$\operatorname{var} X = \mathbb{E}\left((X-c)^2\right), \tag{3.25}$$

where $c = \mathbb{E}\,(X)$.

The variance measures the quadratic mean deviation of X from its mean value $c = \mathbb{E}\,(X)$. Another useful expression for variance is the following:

$$\operatorname{var} X = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}\,(X)\right)^2. \tag{3.26}$$

Indeed, from (3.25), we obtain

$$\operatorname{var} X = \mathbb{E}\left(X^2\right) - 2c\mathbb{E}\,(X) + c^2 = \mathbb{E}\left(X^2\right) - 2c^2 + c^2 = \mathbb{E}\left(X^2\right) - c^2.$$

Since by (3.25) $\operatorname{var} X \ge 0$, it follows from (3.26) that

$$\left(\mathbb{E}\,(X)\right)^2 \le \mathbb{E}(X^2). \tag{3.27}$$

Alternatively, (3.27) follows from a more general Cauchy-Schwarz inequality: for any two random variables X and Y,

$$\left(\mathbb{E}\,(|XY|)\right)^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2). \tag{3.28}$$

The definition of expectation is convenient to prove the properties of $\mathbb{E}$ but no good for computation. For the latter, there is the following theorem.

Theorem 3.3 *Let X be a random variable and f be a Borel function on $(-\infty, +\infty)$. Then $f(X)$ is also a random variable and*

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x) dP_X(x). \tag{3.29}$$

If measure P_X has the density $g(x)$ then

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x) g(x)\, d\lambda(x). \tag{3.30}$$

Proof. Function $f(X)$ is measurable by Theorem 2.2. It suffices to prove (3.29) for non-negative f. Consider first an indicator function

$$f(x) = \mathbf{1}_A(x),$$

for some Borel set $A \subset \mathbb{R}$. For this function, the integral in (3.29) is equal to

$$\int_A dP_X(A) = P_X(A).$$

On the other hand,

$$f(X) = \mathbf{1}_A(X) = \mathbf{1}_{\{X \in A\}}$$

and

$$\mathbb{E}(f(X)) = \int_\Omega \mathbf{1}_{\{X \in A\}} d\mathbb{P} = \mathbb{P}(X \in A) = P_X(A).$$

Hence, (3.29) holds for the indicator functions. In the same way one proves (3.30) for indicator functions.

By the linearity of the integral, (3.29) (and (3.30)) extends to functions which are finite linear combinations of indicator functions. Then by the monotone convergence theorem the identity (3.29) (and (3.30)) extends to infinite linear combinations of the indicator functions, that is, to simple functions, and then to arbitrary Borel functions f. ■

In particular, it follows from (3.29) and (3.30) that

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, dP_X(x) = \int_{\mathbb{R}} x g(x)\, d\lambda(x). \tag{3.31}$$

Also, denoting $c = \mathbb{E}(X)$, we obtain

$$\operatorname{var} X = \mathbb{E}(X - c)^2 = \int_{\mathbb{R}} (x - c)^2 dP_X = \int_{\mathbb{R}} (x - c)^2 g(x)\, d\lambda(x).$$

Example. If $X \sim \mathcal{N}(a, b)$ then one finds

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b}\right) dx = \int_{-\infty}^{+\infty} \frac{y+a}{\sqrt{2\pi b}} \exp\left(-\frac{y^2}{2b}\right) dy = a,$$

where we have used the fact that

$$\int_{-\infty}^{+\infty} \frac{y}{\sqrt{2\pi b}} \exp\left(-\frac{y^2}{2b}\right) dy = 0$$

because the function under the integral is odd, and

$$\int_{-\infty}^{+\infty} \frac{a}{\sqrt{2\pi b}} \exp\left(-\frac{y^2}{2b}\right) dy = a,$$

which is true by (3.23). Similarly, we have

$$\operatorname{var} X = \int_{-\infty}^{+\infty} \frac{(x-a)^2}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b}\right) dx = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi b}} \exp\left(-\frac{y^2}{2b}\right) dy = b, \quad (3.32)$$

where the last identity holds by Exercise 50. Hence, for the normal distribution $\mathcal{N}(a,b)$, the expectation is a and the variance is b.

Example. If P_X is a discrete distribution with atoms $\{x_k\}$ and values $\{p_k\}$ then (3.31) becomes

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} x_k p_k.$$

For example, if $X \sim Po(\lambda)$ then one finds

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$$

Similarly, we have

$$\begin{aligned} \mathbb{E}\left(X^2\right) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda^2 + \lambda, \end{aligned}$$

whence

$$\operatorname{var} X = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 = \lambda.$$

Hence, for the Poisson distribution with the parameter λ, both the expectation and the variance are equal to λ.

3.4 Random vectors and joint distributions

Definition. A mapping $X : \Omega \to \mathbb{R}^n$ is called a *random vector* (or a vector-valued random variable) if it is measurable with respect to the σ-algebra $\mathcal{F}$.

Let $X_1, ..., X_n$ be the components of X. Recall that the measurability of X means that the following set

$$\{\omega \in \Omega : X_1(\omega) \le x_1, ..., X_n(\omega) \le x_n\} \quad (3.33)$$

is measurable for all reals $x_1, ..., x_n$.

Lemma 3.4 *A mapping $X : \Omega \to \mathbb{R}^n$ is a random vector if and only if all its components $X_1, ..., X_n$ are random variables.*

Proof. If all components are measurable then all sets $\{X_1 \leq x_1\}, ..., \{X_n \leq x_n\}$ are measurable and, hence, the set (3.33) is also measurable as their intersection. Conversely, let the set (3.33) is measurable for all x_k. Since

$$\{X_1 \leq x\} = \bigcup_{m=1}^{\infty} \{X_1 \leq x, X_2 \leq m, X_3 \leq m, ..., X_n \leq m\}$$

we see that $\{X_1 \leq x\}$ is measurable and, hence, X_1 is a random variable. In the same way one handles X_k with $k > 1$. ■

Corollary. *If $X : \Omega \to \mathbb{R}^n$ is a random vector and $f : \mathbb{R}^n \to \mathbb{R}^m$ is a Borel mapping then $f \circ X : \Omega \to \mathbb{R}^m$ is a random vector.*

Proof. Let $f_1, ..., f_m$ be the components of f. By the argument similar to Lemma 3.4, each f_k is a Borel function on $\mathbb{R}^n$. Since the components $X_1, ..., X_n$ are measurable functions, by Theorem 2.2 the function $f_k(X_1, ..., X_n)$ is measurable, that is, a random variable. Hence, again by Lemma 3.4 the mapping $f(X_1, ..., X_n)$ is a random vector. ■

As follows from Lemma 2.1, for any random vector $X : \Omega \to \mathbb{R}^n$ and for any Borel set $A \subset \mathbb{R}^n$, the set $\{X \in A\} = X^{-1}(A)$ is measurable. Similarly to the one-dimensional case, introduce the distribution measure P_X on $\mathcal{B}(\mathbb{R}^n)$ by

$$P_X(A) = \mathbb{P}(X \in A).$$

Lemma 3.5 *If X is a random vector then P_X is a probability measure on $\mathcal{B}(\mathbb{R}^n)$. Conversely, if μ is any probability measure on $\mathcal{B}(\mathbb{R}^n)$ then there exists a random vector X such that $P_X = \mu$.*

The proof is the same as in the dimension 1 (see Lemma 3.2) and is omitted.

If $X_1, ..., X_n$ is a sequence of random variable then consider the random vector $X = (X_1, ..., X_n)$ and its distribution P_X.

Definition. The measure P_X is referred to as the *joint distribution* of the random variables $X_1, X_2, ..., X_n$. It is also denoted by $P_{X_1 X_2 ... X_n}$.

As in the one-dimensional case, any non-negative measurable function $g(x)$ on $\mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} g(x)dx = 1,$$

is associated with a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, which is defined by

$$\mu(A) = \int_A g(x)\lambda_n(x).$$

The function g is called *the density* of μ. If the distribution P_X of a random vector X has the density g then we also say that X has the density g (or the density function g).

Theorem 3.6 *If $X = (X_1, ..., X_n)$ is a random vector and $f : \mathbb{R}^n \to \mathbb{R}$ is a Borel function then*

$$\mathbb{E}f(X) = \int_{\mathbb{R}^n} f(x)dP_X(x).$$

If X has the density $g(x)$ then

$$\mathbb{E}f(X) = \int_{\mathbb{R}^n} f(x)g(x)\,d\lambda_n(x). \tag{3.34}$$

The proof is exactly as that of Theorem 3.3 and is omitted.

One can also write (3.34) in a more explicit form:

$$\mathbb{E}f(X_1, X_2, ..., X_n) = \int_{\mathbb{R}^n} f(x_1, ..., x_n) g(x_1, ..., x_n)\, d\lambda_n(x_1, ..., x_n). \tag{3.35}$$

As an example of application of Theorem 3.6, let us show how to recover the density of a component X_k given the density of X.

Corollary. *If a random vector X has the density function g then its component X_1 has the density function*

$$g_1(x) = \int_{\mathbb{R}^{n-1}} g(x, x_2, ..., x_n)\, d\lambda_{n-1}(x_2, ..., x_n). \tag{3.36}$$

Similar formulas take places for other components.

Proof. Let us apply (3.35) with function $f(x) = \mathbf{1}_{\{x_1 \in A\}}$ where A is a Borel set on $\mathbb{R}$; that is,

$$f(x) = \begin{cases} 1, & x_1 \in A \\ 0, & x_1 \notin A \end{cases} = \mathbf{1}_{A \times \mathbb{R} \times ... \times R}.$$

Then

$$\mathbb{E}f(X_1, ..., X_n) = \mathbb{P}(X_1 \in A) = P_{X_1}(A),$$

and (3.35) together with Fubini's theorem yield

$$\begin{aligned} P_{X_1}(A) &= \int_{\mathbb{R}^n} 1_{A \times \mathbb{R} \times ... \times \mathbb{R}}(x_1, ..., x_n)\, g(x_1, ..., x_n)\, d\lambda_n(x_1, ..., x_n) \\ &= \int_{A \times \mathbb{R}^{n-1}} g(x_1, ..., x_n)\, d\lambda_n(x_1, ..., x_n) \\ &= \int_A \left(\int_{\mathbb{R}^{n-1}} g(x_1, ..., x_n)\, d\lambda_{n-1}(x_2, ..., x_n) \right) d\lambda_1(x_1). \end{aligned}$$

This implies that the function in the brackets (which is a function of x_1) is the density of X_1, which proves (3.36). ■

Example. Let X, Y be two random variables with the joint density function

$$g(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) \tag{3.37}$$

(the two-dimensional normal distribution). Then X has the density function

$$g_1(x) = \int_{-\infty}^{+\infty} g(x, y) dy = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

that is, $X \sim \mathcal{N}(0, 1)$. In the same way $Y \sim \mathcal{N}(0, 1)$.

Theorem 3.7 *Let $X : \Omega \to \mathbb{R}^n$ be a random variable with the density function g. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism. Then the random variable $Y = \Phi(X)$ has the following density function*

$$h = g \circ \Phi^{-1} \left| \det \left(\Phi^{-1}\right)' \right|.$$

Proof. We have for any open set $A \in \mathbb{R}^n$

$$\mathbb{P}(\Phi(X) \in A) = \mathbb{P}\left(X \in \Phi^{-1}(A)\right) = \int_{\Phi^{-1}(A)} g(x)\, d\lambda_n(x).$$

Substituting $x = \Phi^{-1}(y)$ we obtain by Theorem 3.1

$$\int_{\Phi^{-1}(A)} g(x)\, d\lambda_n(x) = \int_A g\left(\Phi^{-1}(y)\right) \left|\det\left(\Phi^{-1}(y)\right)'\right| d\lambda_n(y) = \int_A h d\lambda_n,$$

so that

$$\mathbb{P}(Y \in A) = \int_A h d\lambda_n. \tag{3.38}$$

Since on the both sides of this identity we have finite measures on $\mathcal{B}(\mathbb{R}_n)$ that coincide on open sets, it follows from the uniqueness part of the Carathéodory extension theorem that the measures coincides on all Borel sets. Hence, (3.38) holds for all Borel sets A, which was to be proved. ■

Example. For any $\alpha \in \mathbb{R} \setminus \{0\}$, the density function of $Y = \alpha X$ is

$$h(x) = g\left(\frac{x}{\alpha}\right) |\alpha|^{-n}.$$

In particular, if $X \sim \mathcal{N}(a, b)$ then $\alpha X \sim \mathcal{N}(\alpha a, \alpha^2 b)$.

As another example of application of Theorem 3.7, let us prove the following statement.

Corollary. *Assuming that X, Y are two random variables with the joint density function $g(x, y)$. Then the random variable $U = X + Y$ has the density function*

$$h_1(u) = \frac{1}{2} \int_{\mathbb{R}} g\left(\frac{u+v}{2}, \frac{u-v}{2}\right) d\lambda(v) = \int_{\mathbb{R}} g(t, u-t)\, dt, \tag{3.39}$$

and $V = X - Y$ has the density function

$$h_2(v) = \frac{1}{2} \int_{\mathbb{R}} g\left(\frac{u+v}{2}, \frac{u-v}{2}\right) d\lambda(u) = \int_{\mathbb{R}} g(t, t-v)\, dt. \tag{3.40}$$

Proof. Consider the mapping $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$\Phi \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x-y \end{pmatrix},$$

and the random vector

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} X+Y \\ X-Y \end{pmatrix} = \Phi(X, Y).$$

Clearly, Φ is a diffeomorphism, and

$$\Phi^{-1} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{u+v}{2} \\ \frac{u-v}{2} \end{pmatrix}, \quad \left(\Phi^{-1}\right)' = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}, \quad \left|\det\left(\Phi^{-1}\right)'\right| = \frac{1}{2}.$$

Therefore, the density function h of (U,V) is given by

$$h(u,v) = \frac{1}{2} g\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

By Corollary to Theorem 3.6, the density function h_1 of U is given by

$$h_1(u) = \int_{\mathbb{R}} h(u,v)\,d\lambda(v) = \frac{1}{2}\int_{\mathbb{R}} g\left(\frac{u+v}{2}, \frac{u-v}{2}\right) d\lambda(v).$$

The second identity in (3.39) is obtained using the substitution $t = \frac{u+v}{2}$. In the same way one proves (3.40). ∎

Example. Let X, Y be again random vectors with joint density (3.37). Then by (3.39) we obtain the density of $X+Y$:

$$\begin{aligned} h_1(u) &= \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{(u+v)^2 + (u-v)^2}{8}\right) dv \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{4} - \frac{v^2}{4}\right) dv \\ &= \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{4}u^2}. \end{aligned}$$

Hence, $X+Y \sim \mathcal{N}(0,2)$ and in the same way $X-Y \sim \mathcal{N}(0,2)$.

3.5 Independent random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space as before.

Definition. Two random vectors $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ are called *independent* if, for all Borel sets $A \in \mathcal{B}_n$ and $B \in \mathcal{B}_m$, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent, that is

$$\mathbb{P}(X \in A \text{ and } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Similarly, a sequence $\{X_i\}$ of random vectors $X_i : \Omega \to \mathbb{R}^{n_i}$ is called independent if, for any sequence $\{A_i\}$ of Borel sets $A_i \in \mathcal{B}_{n_i}$, the events $\{X_i \in A_i\}$ are independent. Here the index i runs in any index set (which may be finite, countable, or even uncountable).

If $X_1, ..., X_k$ is a finite sequence of random vectors, such that $X_i : \Omega \to \mathbb{R}^{n_i}$ then we can form a vector $X = (X_1, ..., X_k)$ whose components are those of all X_i; that is, X is an n-dimensional random vector where $n = n_1 + ... + n_k$. The distribution measure P_X of X is called the joint distribution of the sequence $X_1, ..., X_k$ and is also denoted by $P_{X_1 ... X_k}$.

A particular case of the notion of a joint distribution for the case when all X_i are random variables, was considered above.

Theorem 3.8 *Let X_i be a n_i-dimensional random vector, $i = 1, ..., k$. The sequence $X_1, X_2, ..., X_k$ is independent if and only if their joint distribution $P_{X_1 ... X_k}$ coincides with the product measure of $P_{X_1}, P_{X_2}, ..., P_{X_k}$, that is*

$$P_{X_1 ... X_k} = P_{X_1} \times ... \times P_{X_k}. \tag{3.41}$$

If $X_1, X_2, ..., X_k$ are independent and in addition X_i has the density function $f_i(x)$ then the sequence $X_1, ..., X_k$ has the joint density function

$$f(x) = f_1(x_1)f_2(x_2)...f_k(x_k),$$

where $x_i \in \mathbb{R}^{n_i}$ and $x = (x_1, ..., x_k) \in \mathbb{R}^n$.

Proof. If (3.41) holds then, for any sequence $\{A_i\}_{i=1}^k$ of Borel sets $A_i \subset \mathbb{R}^{n_i}$, consider their product

$$A = A_1 \times A_2 \times ... \times A_k \subset \mathbb{R}^n \tag{3.42}$$

and observe that

$$\begin{aligned}
\mathbb{P}(X_1 \in A_1, ..., X_k \in A_k) &= \mathbb{P}(X \in A) \\
&= P_X(A) \\
&= P_{X_1} \times ... \times P_{X_k}(A_1 \times ... \times A_k) \\
&= P_{X_1}(A_1)...P_{X_k}(A_k) \\
&= \mathbb{P}(X_1 \in A_1)...\mathbb{P}(X_k \in A_k).
\end{aligned}$$

Hence, $X_1, ..., X_k$ are independent.

Conversely, if $X_1, ..., X_k$ are independent then, for any set A of the product form (3.42), we obtain

$$\begin{aligned}
P_X(A) &= \mathbb{P}(X \in A) \\
&= \mathbb{P}(X_1 \in A_1,\ X_2 \in A_2,\ ...,\ X_k \in A_k) \\
&= \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2)...\mathbb{P}(X_k \in A_k) \\
&= P_{X_1} \times ... \times P_{X_k}(A).
\end{aligned}$$

Hence, the measure P_X and the product measure $P_{X_1} \times ... \times P_{X_k}$ coincide on the sets of the form (3.42). Since $\mathcal{B}_n$ is the minimal σ-algebra containing the sets (3.42), the uniqueness part of the Carathéodory extension theorem implies that these two measures coincide on $\mathcal{B}_n$, which was to be proved.

To prove the second claim, observe that, for any set A of the form (3.42), we have by Fubini's theorem

$$\begin{aligned}
\mathbb{P}_X(A) &= P_{X_1}(A_1)...P_{X_k}(A_k) \\
&= \int_{A_1} f_1(x_1)\,d\lambda_{n_1}(x_1)...\int_{A_k} f_k(x_k)\,d\lambda_{n_k}(x_k) \\
&= \int_{A_1\times...\times A_k} f_1(x_1)...f_k(x_k)\,d\lambda_n(x) \\
&= \int_A f(x)\,d\lambda_n(x),
\end{aligned}$$

where $x = (x_1, ..., x_k) \in \mathbb{R}^n$. Hence, the two measures $P_X(A)$ and

$$\mu(A) = \int_A f(x)\,d\lambda_n(x)$$

coincide of the product sets, which implies by the uniqueness part of the Carathéodory extension theorem that they coincide on all Borel sets $A \subset \mathbb{R}^n$. $\blacksquare$

Corollary. *If X and Y are independent integrable random variables then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) \tag{3.43}$$

and

$$\operatorname{var}(X+Y) = \operatorname{var} X + \operatorname{var} Y. \tag{3.44}$$

Proof. Using Theorems 3.6, 3.8, and Fubini's theorem, we have

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xy dP_{XY} = \int_{\mathbb{R}^2} xy d\left(P_X \times P_Y\right) = \left(\int_{\mathbb{R}} x dP_X\right)\left(\int_{\mathbb{R}} y dP_Y\right) = \mathbb{E}(X)\mathbb{E}(Y).$$

To prove the second claim, observe that $\operatorname{var}(X+c) = \operatorname{var} X$ for any constant c. Hence, we can assume that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. In this case, we have $\operatorname{var} X = \mathbb{E}(X^2)$. Using (3.43) we obtain

$$\operatorname{var}(X+Y) = \mathbb{E}(X+Y)^2 = \mathbb{E}\left(X^2\right) + 2\mathbb{E}(XY) + \mathbb{E}\left(Y^2\right) = \mathbb{E}X^2 + \mathbb{E}Y^2 = \operatorname{var} X + \operatorname{var} Y.$$

$\blacksquare$

The identities (3.43) and (3.44) extend to the case of an arbitrary finite sequence of independent integrable random variables $X_1, ..., X_n$ as follows:

$$\mathbb{E}\left(X_1...X_n\right) = \mathbb{E}\left(X_1\right)...\mathbb{E}\left(X_n\right),$$

$$\operatorname{var}\left(X_1 + ... + X_n\right) = \operatorname{var} X_1 + ... + \operatorname{var} X_n \;,$$

and the proofs are the same.

Example. Let X, Y be independent random variables and

$$X \sim \mathcal{N}(a, b), \quad Y \sim \mathcal{N}(a', b').$$

Let us show that

$$X + Y \sim \mathcal{N}(a + a', b + b').$$

Let us first simply evaluate the expectation and the variance of $X + Y$:

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = a + a',$$

and, using the independence of X, Y and (3.44),

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) = b + b'.$$

However, this does not yet give the distribution of $X + Y$. To simplify the further argument, rename $X - a$ to X, $Y - a'$ to Y, so that we can assume in the sequel that $a = a' = 0$. Also for simplicity of notation, set $c = b'$. By Theorem 3.8, the joint density function of X, Y is as follows:

$$\begin{aligned} g(x,y) &= \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{x^2}{2b}\right) \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{y^2}{2c}\right) \\ &= \frac{1}{2\pi\sqrt{bc}} \exp\left(-\frac{x^2}{2b} - \frac{y^2}{2c}\right). \end{aligned}$$

By Corollary to Theorem 3.7, we obtain that the random variable $U = X + Y$ has the density function

$$\begin{aligned} h(u) &= \int_{\mathbb{R}} g(t, u-t)\, d\lambda(t) \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{bc}} \exp\left(-\frac{t^2}{2b} - \frac{(u-t)^2}{2c}\right) dt \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{bc}} \exp\left(-\left(\frac{1}{b} + \frac{1}{c}\right)\frac{t^2}{2} + \frac{ut}{c} - \frac{u^2}{2c}\right) dt. \end{aligned}$$

Next compute

$$\int_{-\infty}^{+\infty} \exp\left(-\alpha t^2 + 2\beta t\right) dt$$

where $\alpha > 0$ and $\beta \in \mathbb{R}$. We have

$$\begin{aligned} -\alpha t^2 + 2\beta t &= -\alpha\left(t^2 - \frac{2\beta}{\alpha}t + \frac{\beta^2}{\alpha^2} - \frac{\beta^2}{\alpha^2}\right) \\ &= -\alpha\left(t - \frac{\beta}{\alpha}\right)^2 + \frac{\beta^2}{\alpha}. \end{aligned}$$

Hence, substituting $\sqrt{\alpha}\left(t - \frac{\beta}{\alpha}\right) = s$, we obtain

$$\int_{-\infty}^{+\infty} \exp\left(-\alpha t^2 + 2\beta t\right) dt = \frac{e^{\beta^2/\alpha}}{\sqrt{\alpha}} \int_{-\infty}^{\infty} \exp\left(-s^2\right) ds = \sqrt{\frac{\pi}{\alpha}} e^{\beta^2/\alpha}.$$

Setting here $\alpha = \frac{1}{2}\left(\frac{1}{b} + \frac{1}{c}\right)$ and $\beta = \frac{u}{2c}$, we obtain

$$h(u) = \frac{1}{2\pi\sqrt{bc}} e^{-u^2/(2c)} \sqrt{\frac{2bc\pi}{b+c}} e^{\frac{u^2}{4c^2}\frac{2bc}{b+c}} = \frac{1}{\sqrt{2\pi(b+c)}} \exp\left(-\frac{u^2}{2(b+c)}\right),$$

that is, $U \sim \mathcal{N}(0, b+c)$.

Example. Consider a sequence $\{X_i\}_{i=1}^n$ of independent Bernoulli variables taking 1 with probability p, and 0 with probability $1-p$, where $0 < p < 1$. Set $S = X_1 + ... + X_n$ and prove that, for any $k = 0, 1, ..., n$

$$\mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{3.45}$$

Indeed, the sum S is equal to k if and only if exactly k of the values $X_1, ..., X_n$ are equal to 1 and $n-k$ are equal to 0. The probability, that the given k variables from X_i, say $X_{i_1}, ..., X_{i_k}$ are equal to 1 and the rest are equal to 0, is equal to $p^k (1-p)^{n-k}$. Since the sequence $(i_1, ..., i_k)$ can be chosen in $\binom{n}{k}$ ways, we obtain (3.45). Hence, S has the binomial distribution $B(n,p)$. This can be used to obtain in an easy way the expectation and the variance of the binomial distribution – see Exercise 67d.

In the next statement, we collect some more useful properties of independent random variables.

Theorem 3.9 (a) *If $X_i : \Omega \to \mathbb{R}^{n_i}$ is a sequence of independent random vectors and $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^{m_i}$ is a sequence of Borel functions then the random vectors $\{f_i(X_i)\}$ are independent.*

(b) *If $\{X_1, X_2, ..., X_n, Y_1, Y_2, .., Y_m\}$ is a sequence of independent random variables then the random vectors*

$$X = (X_1, X_2, ..., X_n) \quad \text{and} \quad Y = (Y_1, Y_2, ..., Y_m)$$

are independent.

(c) *Under conditions of (b), for all Borel functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$, the random variables $f(X_1, ..., X_n)$ and $g(Y_1, ..., Y_m)$ are independent.*

Proof. (a) Let $\{A_i\}$ be a sequence of Borel sets, $A_i \in \mathcal{B}_{m_i}$. We need to show that the events $\{f_i(X_i) \in A_i\}$ are independent. Since

$$\{f_i(X_i) \in A_i\} = \left\{X_i \in f_i^{-1}(A_i)\right\} = \{X_i \in B_i\},$$

where $B_i = f^{-1}(A_i) \in \mathcal{B}_{n_i}$, these sets are independent by the definition of the independence of $\{X_i\}$.

(b) We have by Theorem 3.8

$$\begin{aligned} P_{XY} &= P_{X_1} \times ... \times P_{X_n} \times P_{Y_1} \times ... \times P_{Y_m} \\ &= P_X \times P_Y. \end{aligned}$$

Hence, X and Y are independent.

(c) The claim is an obvious combination of (a) and (b). ■

3.6 Sequences of random variables

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent random variables and consider their partial sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We will be interested in the behavior of S_n as $n \to \infty$. The necessity of such considerations arises in many mathematical models of random phenomena.

Example. Consider a sequence on n independent trials of coin tossing and set $X_i = 1$ if at the i-th tossing the coin lands showing heads, and $X_i = 0$ if the coin lands showing tails. Assuming that the heads appear with probability p, we see each X_i has the same Bernoulli distribution: $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$. Based on the experience, we can assume that the sequence $\{X_i\}$ is independent. Then S_n is exactly the number of heads in a series of n trials. If n is big enough then one can expect that $S_n \approx np$. The claims that justify this conjecture are called *laws of large numbers*. Two such statements of this kind will be presented below.

As we have seen in Section 1.10, for any positive integer n, there exist n independent events with prescribed probabilities. Taking their indicators, we obtain n independent Bernoulli random variables. However, in order to make the above consideration of infinite sequences of independent random variables meaningful, one has to make sure that such sequences do exist. The next theorem ensures that, providing enough supply of sequences of independent random variables.

Theorem 3.10 *Let $\{\mu_i\}_{i\in I}$ be a family of probability measures on $\mathcal{B}(\mathbb{R})$ where I is an arbitrary index set. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family $\{X_i\}_{i\in I}$ of independent random variables on Ω such that $P_{X_i} = \mu_i$.*

The space Ω is constructed as the (infinite) product $\mathbb{R}^I$ and the probability measure $\mathbb{P}$ on Ω is constructed as the product $\bigotimes_{i\in I}\mu_i$ of the measures μ_i. The argument is similar to Theorem 2.22 but technically more involved. The proof is omitted.

Theorem 3.10 is a particular case of a more general *Kolmogorov's extension theorem* that allows constructing families of random variables with prescribed joint densities.

3.7 The weak law of large numbers

Definition. We say that a sequence $\{Y_n\}$ of random variables *converges in probability* to a random variable Y and write $Y_n \xrightarrow{\mathbb{P}} Y$ if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| > \varepsilon) = 0.$$

This is a particular case of the *convergence in measure* – see Exercise 41.

Theorem 3.11 (The weak law of large numbers) *Let $\{X_i\}_{i=1}^\infty$ be independent sequence of random variables having a common finite expectation $\mathbb{E}(X_i) = a$ and a common finite variance* $\operatorname{var} X_i = b$. *Then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} a. \tag{3.46}$$

Hence, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - a\right| \le \varepsilon\right) \to 1 \text{ as } n \to \infty$$

which means that, for large enough n, the average S_n/n concentrates around a with a probability close to 1.

Example. In the case of coin tossing, $a = \mathbb{E}(X_i) = p$ and $\operatorname{var} X_i$ is finite so that $\frac{S_n}{n} \xrightarrow{\mathbb{P}} p$. One sees that in some sense $S_n/n \approx p$, and the precise meaning of that is the following:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \le \varepsilon\right) \to 1 \text{ as } n \to \infty.$$

Example. Assume that all $X_i \sim \mathcal{N}(0,1)$. Using the properties of the sums of independent random variables with the normal distribution, we obtain

$$S_n \sim \mathcal{N}(0, n)$$

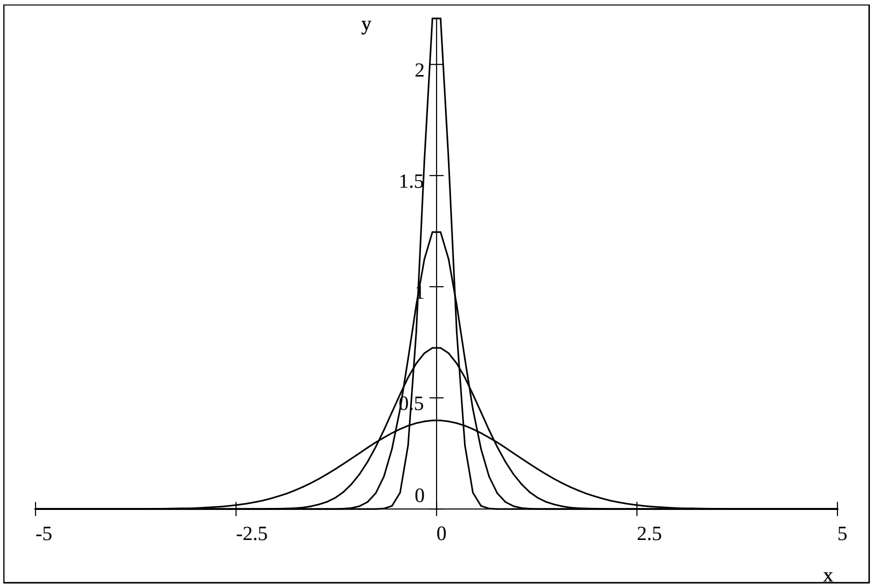
and

$$\frac{S_n}{n} \sim \mathcal{N}\left(0, \frac{1}{n}\right).$$

Hence, for large n, the variance of $\frac{S_n}{n}$ becomes very small and $\frac{S_n}{n}$ becomes more concentrated around its mean 0. On the plot below one can see the density functions of $\frac{S_n}{n}$, that is,

$$g_n(x) = \sqrt{\frac{n}{2\pi}} \exp\left(-n\frac{x^2}{2}\right),$$

which becomes steeper for larger n.



One can conclude (without even using Theorem 3.11) that

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \leq \varepsilon\right) \to 1 \text{ as } n \to \infty,$$

that is, $\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0$.

Proof of Theorem 3.11. We use the following simple inequality, which is called Chebyshev's inequality: if $Y \geq 0$ is a random variable then, for all $t > 0$,

$$\mathbb{P}(Y \geq t) \leq \frac{1}{t^2}\mathbb{E}\left(Y^2\right). \tag{3.47}$$

Indeed,

$$\mathbb{E}\left(Y^2\right) = \int_\Omega Y^2 d\mathbb{P} \geq \int_{\{Y \geq t\}} Y^2 d\mathbb{P} \geq t^2 \int_{\{Y \geq y\}} d\mathbb{P} = t^2\mathbb{P}(Y \geq t),$$

whence (3.47) follows.

Let us evaluate the expectation and the variance of S_n:

$$\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}X_i = an,$$

$$\operatorname{var} S_n = \sum_{k=i}^n \operatorname{var} X_i = bn,$$

where in the second line we have used the independence of $\{X_i\}$ and Corollary to Theorem 3.8. Hence, applying (3.47) with $Y = |S_n - an|$, we obtain

$$\mathbb{P}\left(|S_n - an| \geq \varepsilon n\right) \leq \frac{1}{(\varepsilon n)^2}\mathbb{E}\left(S_n - an\right)^2 = \frac{\operatorname{var} S_n}{(\varepsilon n)^2} = \frac{bn}{(\varepsilon n)^2} = \frac{b}{\varepsilon^2 n}.$$

Therefore,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right) \leq \frac{b}{\varepsilon^2 n}, \tag{3.48}$$

whence (3.46) follows. ■

3.8 The strong law of large numbers

In this section we use the convergence of random variables *almost surely* (a.s.) Namely, if $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables then we say that X_n converges to X a.s. and write $X_n \xrightarrow{\text{a.s.}} X$ if $X_n \to X$ almost everywhere, that is,

$$\mathbb{P}\left(X_n \to X\right) = 1.$$

In the expanded form, this means that

$$\mathbb{P}\left(\left\{\omega : \lim_{n\to\infty} X_n\left(\omega\right) = X\left(\omega\right)\right\}\right) = 1.$$

To understand the difference between the convergence a.s. and the convergence in probability, consider an example.

Example. Let us show an example where $X_n \xrightarrow{\mathbb{P}} X$ but X_n does not converges to X a.s. Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}\left([0,1]\right)$, and $\mathbb{P} = \lambda_1$. Let A_n be an interval in $[0,1]$ and set $X_n = \mathbf{1}_{A_n}$. Observe that if $\ell\left(A_n\right) \to 0$ then $X_n \xrightarrow{\mathbb{P}} 0$ because for all $\varepsilon \in (0,1)$

$$\mathbb{P}\left(X_n > \varepsilon\right) = \mathbb{P}\left(X_n = 1\right) = \ell\left(A_n\right).$$

However, for certain choices of A_n, we do not have $X_n \xrightarrow{\text{a.s.}} 0$. Indeed, one can construct $\{A_n\}$ so that $\ell\left(A_n\right) \to 0$ but nevertheless any point $\omega \in [0,1]$ is covered by these intervals infinitely many times. For example, one can take as $\{A_n\}$ the following sequence:

$$\begin{array}{l} [0,1], \\ \left[0,\frac{1}{2}\right], \left[\frac{1}{2},1\right], \\ \left[0,\frac{1}{3}\right], \left[\frac{1}{3},\frac{2}{3}\right], \left[\frac{2}{3},1\right], \\ \left[0,\frac{1}{4}\right], \left[\frac{1}{4},\frac{2}{4}\right], \left[\frac{2}{4},\frac{3}{4}\right], \left[\frac{3}{4},1\right] \\ \ldots \end{array}$$

For any $\omega \in [0,1]$, the sequence $\{X_n(\omega)\}$ contains the value 1 infinitely many times, which means that this sequence does not converge to 0. It follows that X_n does not converge to 0 a.s.; moreover, we have $\mathbb{P}\left(X_n \to 0\right) = 0$.

The following theorem states the relations between the two types of convergence.

Theorem 3.12 *Let $\{X_n\}$ be a sequence of random variables.*

(a) If $X_n \xrightarrow{\text{a.s.}} X$ then $X_n \xrightarrow{\mathbb{P}} X$ (hence, the convergence a.s. is stronger than the convergence in probability).

(b) If, for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) < \infty \tag{3.49}$$

then $X_n \xrightarrow{\text{a.s.}} X$.

(c) If $X_n \xrightarrow{\mathbb{P}} X$ then there exists a subsequence $X_{n_k} \xrightarrow{\text{a.s.}} X$.

The proof of Theorem 3.12 is contained in Exercise 42. In fact, we need only part (b) of this theorem. The condition (3.49) is called *the Borel-Cantelli condition.* It is obviously stronger that $X_n \xrightarrow{\mathbb{P}} X$ because the convergence of a series implies that the terms of the series go to 0. Part (b) says that the Borel-Cantelli condition is also stronger than $X_n \xrightarrow{\text{a.s.}} X$, which is not that obvious.

Let $\{X_i\}$ be a sequence of random variables. We say that this sequence is *identically distributed* if all X_i have the same distribution measure. If $\{X_i\}$ are identically distributed then their expectations are the same and the variances are the same.

As before, set $S_n = X_1 + ... + X_n$.

Theorem 3.13 (The strong law of large numbers) *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent identically distributed random variables with a finite expectation $\mathbb{E}(X_i) = a$ and a finite variance* $\operatorname{var} X_i = b$. *Then*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} a. \tag{3.50}$$

The word "strong" is the title of the theorem refers to the convergence a.s. in (3.50), as opposed to the weaker convergence in probability in Theorem 3.11 (cf. (3.46)).

The statement of Theorem 3.13 remains true if one drops the assumption of the finiteness of $\operatorname{var} X_n$. Moreover, the finiteness of the mean $\mathbb{E}(X_n)$ is not only sufficient but also necessary condition for the existence of the limit $\lim \frac{s_n}{n}$ a.s. (Kolmogorov's theorem). Another possibility to relax the hypotheses is to drop the assumption that X_n are identically distributed but still require that X_n have a common finite mean and a common finite variance. The proofs of these stronger results are much longer and will not be presented here.

Proof of Theorem 3.13. By Theorem 3.12(b), it suffices to verify the Borel-Cantelli condition, that is, for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) < \infty. \tag{3.51}$$

In the proof of Theorem 3.11, we have obtained the estimate

$$\mathbb{P}\left(\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right) \leq \frac{b}{\varepsilon^2 n}, \tag{3.52}$$

which however is not enough to prove (3.51), because

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

Nevertheless, taking in (3.52) n to be a perfect square k^2, where $k \in \mathbb{N}$, we obtain

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{k^2}}{k^2} - a\right| > \varepsilon\right) \leq \sum_{k=1}^{\infty} \frac{b}{\varepsilon^2 k^2} < \infty,$$

whence by Theorem 3.12(b)

$$\frac{S_{k^2}}{k^2} \xrightarrow{\text{a.s.}} a. \tag{3.53}$$

Now we will extend this convergence to the whole sequence S_n, that is, fill gaps between the perfect squares. This will be done under the assumption that all X_i are non-negative (the general case of a signed X_i will be treated afterwards). In this case the sequence S_n is increasing. For any positive integer n, find k so that

$$k^2 \leq n < (k+1)^2. \tag{3.54}$$

Using the monotonicity of S_n and (3.54), we obtain

$$\frac{S_{k^2}}{(k+1)^2} \leq \frac{S_n}{n} \leq \frac{S_{(k+1)^2}}{k^2}.$$

Since $k^2 \sim (k+1)^2$ as $k \to \infty$, it follows from (3.53) that

$$\frac{S_{k^2}}{(k+1)^2} \xrightarrow{\text{a.s.}} a \quad \text{and} \quad \frac{S_{(k+1)^2}}{k^2} \xrightarrow{\text{a.s.}} a\,,$$

whence we conclude that

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} a.$$

Note that in this argument we have not used the hypothesis that X_i are identically distributed. As in the proof of Theorem 3.11, we only have used that X_i are independent random variables and that the have a finite common expectation a and a finite common variance.

Now we get rid of the restriction $X_i \geq 0$. For a signed X_i, consider the positive X_i^+ part and the negative parts X_i^-. Then the random variables X_i^+ are independent because we can write $X_i^+ = f(X_i)$ where $f(x) = x_+$, and apply Theorem 3.9. Next, X_i^+ has finite expectation and variance by $X^+ \leq |X|$. Furthermore, by Theorem 3.3 we have

$$\mathbb{E}\left(X_i^+\right) = \mathbb{E}\left(f\left(X_i\right)\right) = \int_{\mathbb{R}} f\left(x\right) dP_{X_i}.$$

Since all measures P_{X_i} are the same, it follows that all the expectations $\mathbb{E}\left(X_i^+\right)$ are the same; set

$$a^+ := \mathbb{E}\left(X_i^+\right).$$

In the same way, the variances $\operatorname{var} X_i^+$ are the same. Setting

$$S_n^+ = X_1^+ + X_2^+ + ... + X_n^+,$$

we obtain by the first part of the proof that

$$\frac{S_n^+}{n} \xrightarrow{\text{a.s.}} a^+ \text{ as } n \to \infty.$$

In the same way, considering X_i^- and setting $a^- = \mathbb{E}\left(X_i^-\right)$ and

$$S_n^- = X_1^- + ... + X_n^- \ ,$$

we obtain

$$\frac{S_n^-}{n} \xrightarrow{\text{a.s.}} a^- \text{ as } n \to \infty.$$

Subtracting the two relations and using that

$$S_n^+ - S_n^- = S_n$$

and

$$a^+ - a^- = \mathbb{E}\left(X_i^+\right) - \mathbb{E}\left(X_i^-\right) = \mathbb{E}\left(X_i\right) = a,$$

we obtain

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} a \text{ as } n \to \infty,$$

which finishes the proof. ■

3.9 Extra material: the proof of the Weierstrass theorem using the weak law of large numbers

Here we show an example how the weak law of large numbers allows to prove the following purely analytic theorem.

Theorem 3.14 (The Weierstrass theorem) *Let $f(x)$ be a continuous function on a bounded interval $[a,b]$. Then, for any $\varepsilon > 0$, there exists a polynomial $P(x)$ such that*

$$\sup_{x\in[a,b]} |f(x) - P(x)| < \varepsilon.$$

Proof. It suffices to consider the case of the interval $[0,1]$. Consider a sequence $\{X_i\}_{i=1}^{\infty}$ of independent Bernoulli variables taking 1 with probability p, and 0 with probability $1-p$, and set as before $S_n = X_1 + ... + X_n$. Then, for $k = 0, 1, ..., n$,

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

which implies

$$\mathbb{E}f\left(\frac{S_n}{n}\right) = \sum_{k=0}^{n} f(\frac{k}{n})\mathbb{P}(S_n = k) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} p^k (1-p)^{n-k}.$$

The right hand side here can be considered as a polynomial in p. Denote

$$B_n(p) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} p^k (1-p)^{n-k}. \tag{3.55}$$

The polynomial $B_n(p)$ is called the *Bernstein's polynomial* of f. It turns out to be a good approximation for f. The idea is that S_n/n converges in some sense to p. Therefore, we may expect that $\mathbb{E}f(\frac{S_n}{n})$ converges to $f(p)$. In fact, we will show that

$$\lim_{n\to\infty} \sup_{p\in[0,1]} |f(p) - B_n(p)| = 0, \tag{3.56}$$

which will prove the claim of the theorem.

First observe that f is uniformly continuous so that for any $\delta > 0$ there exists $\varepsilon = \varepsilon(\delta) > 0$ such that if $|x - y| \le \varepsilon$ then $|f(x) - f(y)| \le \delta$. Using the binomial theorem, we obtain

$$1 = (p + (1-p))^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k},$$

whence

$$f(p) = \sum_{k=0}^{n} f(p)\binom{n}{k} p^k (1-p)^{n-k}.$$

Comparing with (3.55), we obtain

$$\begin{aligned}
|f(p) - B_n(p)| &\le \sum_{k=0}^{n} \left| f(p) - f(\frac{k}{n})\right| \binom{n}{k} p^k (1-p)^{n-k} \\
&= \left(\sum_{\left|\frac{k}{n}-p\right|\le\varepsilon} + \sum_{\left|\frac{k}{n}-p\right|>\varepsilon} \right) \left| f(p) - f(\frac{k}{n})\right| \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

By the choice of ε, in the first sum we have

$$\left|f(p) - f(\frac{k}{n})\right| \leq \delta,$$

so that the sum is bounded by δ.

In the second sum, we use the fact that f is bounded by some constant C so that

$$\left|f(p) - f(\frac{k}{n})\right| \leq 2C,$$

and the second sum is bounded by

$$2C \sum_{\left|\frac{k}{n}-p\right|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} = 2C \sum_{\left|\frac{k}{n}-p\right|>\varepsilon} \mathbb{P}(S_n = k) = 2C\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right).$$

Using the estimate (3.48) and $\mathbb{E}(X_i) = p$, we obtain

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{b}{\varepsilon^2 n},$$

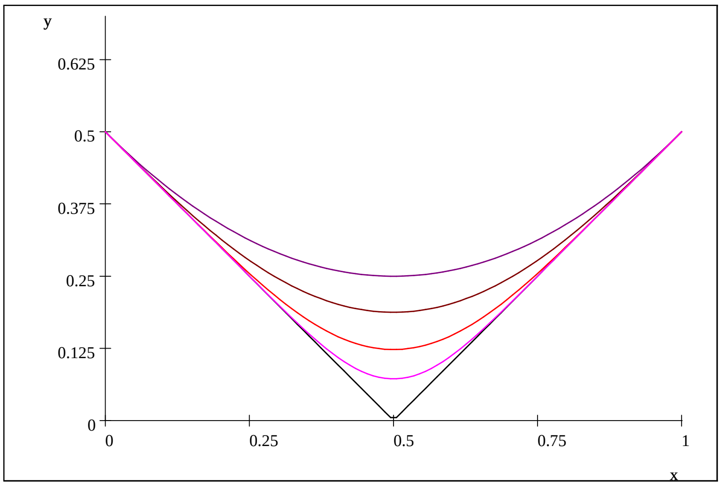
where $b = \operatorname{var} X_i = p(1-p) < 1$.

Hence, for all $p \in [0,1]$,

$$|f(p) - B_n(p)| \leq \delta + \frac{2C}{\varepsilon^2 n},$$

whence the claim follows. ■

To illustrate this theorem, the next plot contains a sequence of Bernstein's approximations with $n = 3, 4, 10, 30$ to the function $f(x) = |x - 1/2|$.



Remark. The upper bound for the sum

$$\sum_{\left|\frac{k}{n}-p\right|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k}$$

can be proved also analytically, which gives also another proof of the weak law of large numbers in the case when all X_i are Bernoulli variables. Such a proof was found by Jacob Bernoulli in 1713, which was historically the first proof of the law of large numbers.