Measure and integral

E. Kowalski (with some minor additions of J. Teichmann for spring term 2012)

ETH Zürich
kowalski@math.ethz.ch,jteichma@math.ethz.ch

Contents

Preamble

This course is an English translation and adaptation with minor changes of the French lecture notes I had written for a course of integration, Fourier analysis and probability in Bordeaux, which I taught in 2001/02 and the following two years.

The only difference worth mentioning between these notes and other similar treatments of integration theory is the incorporation "within the flow" of notions of probability theory (instead of having a specific chapter on probability). These probabilistic asides – usually identified with a grey bar on the left margin – can be disregarded by readers who are interested only in measure theory and integration for classical analysis. The index will (when actually present) highlight those sections in a specific way so that they are easy to read independently.

Acknowledgements. Thanks to the students who pointed out typos and inacurracies in the text as I was preparing it, in particular S. Tornier.

Introduction

The integral of a function was defined first for fairly regular functions defined on a closed interval; its two fundamental properties are that the operation of indefinite integration (with variable end point) is the operation inverse of differentiation, while the definite integral over a fixed interval has a geometric interpretation in terms of area (if the function is non-negative, it is the area under the graph of the function and above the axis of the variable of integration).

However, the integral, as it is first learnt, and as it was defined until the early 20th century (the Riemann integral), has a number of flaws which become quite obvious and problematic when it comes to dealing with situations beyond those of a continuous function $f : [a, b] \to \mathbf{C}$ defined on a closed interval in $\mathbf{R}$.

To describe this, we recall briefly the essence of the definition. It is based on approaching the desired integral

$$\int_a^b f(t)dt$$

with Riemann sums of the type

$$S(f) = \sum_{i=1}^{N} (y_i - y_{i-1}) \sup_{y_{i-1} \leqslant x \leqslant y_i} f(x)$$

where we have a subdivision

$$a = y_0 < y_1 < \cdots < y_N = b \tag{0.1}$$

of $[a, b]$. Indeed, if f is Riemann-integrable, such sums will converge to the integral when a sequence of subdivisions where the steps $\max |y_i - y_{i-1}|$ converge to 0.

Here are some of the difficulties that arise in working with this integral.

- Compatibility with other limiting processes: let (f_n) be a sequence of (continuous) functions on $[a, b]$ “converging” to a function f, in some sense. It is not always true that

$$\lim_{n \to +\infty} \int_a^b f_n(t)dt = \int_a^b f(t)dt, \tag{0.2}$$

  and in fact, with Riemann's definition, f may fail to be integrable. The problem, as often in analysis, is to justify a certain exchange of two limits, and this turns out to be quite difficult.

  The only case where the formula can be shown to hold very easily is when $f_n \to f$ uniformly on $[a, b]$: then f is continuous, and (0.2) holds. However, it is often very hard to check uniform convergence, and it often fails to hold. This becomes even more of a problem when working with integrals over unbounded sets like $]-\infty, +\infty[$, which are themselves limits of integrals over $]a, b[$ as $a \to -\infty$ and $b + \infty$: in effect, we then have three limiting processes to juggle.

The following simple example illustrates that one will always need some care: let $[a,b]=[0,1]$ and

$$f_n(x)=\begin{cases} 2n^2x & \text{if } 0\leqslant x\leqslant \frac{1}{2n}\\ 2n-2n^2x & \text{if } \frac{1}{2n}\leqslant x\leqslant \frac{1}{n}\\ 0 & \text{if } \frac{1}{n}\leqslant x\leqslant 1\end{cases} \tag{0.3}$$

so that $f_n(x)\to 0$ for all $x\in[0,1]$ (since $f_n(0)=0$ and the sequence becomes constant for all $n>x^{-1}$ if $x\in]0,1]$), while

$$\int_0^1 f_n(x)dx = 1/2$$

for all n.

- Multiple integrals: just like that integral

$$\int_a^b f(t)dt$$

has a geometric interpretation when $f(t)\geqslant 0$ on $[a,b]$ as the area of the plane domain situated between the x-axis and the graph of f, one is tempted to write down integrals with more than one variable to compute the volume of a solid in $\mathbf{R}^3$, or the surface area bounded by such a solid (e.g., the surface of a sphere). Riemann's definition encounters very serious difficulties here because the lack of a natural ordering of the plane makes it difficult to find suitable analogues of the subdivisions used for one-variable integration. And even if some definition is found, the justification of the expected formula of exchange of the variables of integration

$$\int_a^b\int_c^d f(x,y)dxdy=\int_c^d\int_a^b f(x,y)dydx \tag{0.4}$$

is usually very difficult.

- Probability: suppose one wants to give a precise mathematical sense to such intuitive statements as "a number taken at random between 0 and 1 has probability 1/2 of being $>1/2$"; it quickly seems natural to say that the probability that $x\in[0,1]$ satisfy a certain property $\mathcal{P}$ is

$$(b-a)=\int_a^b dx$$

if the set $I(\mathcal{P})$ of those $x\in[0,1]$ where $\mathcal{P}$ holds is equal to the interval $[a,b]$. Very quickly, however, one is led to construct properties which are quite natural, yet $I(\mathcal{P})$ is not an interval, or even a finite disjoint union of intervals! For example, let $\mathcal{P}$ be the property: "there is no 1 in the digital expansion $x=0,d_1d_2\cdots$ of x in basis 3, $d_i\in\{0,1,2\}$." What is the probability of this? This is related to the sixth Hilbert problem from 1900.

- Infinite integrals: as already mentioned, in Riemann's method, an integral like

$$\int_a^{+\infty} f(t)dt$$

is defined a posteriori as limit of integrals from a to b, with $b\to+\infty$. This implies that all desired properties of these integrals be checked separately – and involves therefore an additional limiting process. This becomes even more problematic when one tries to define and study functions defined by such integrals, like the Fourier transform of a function $f : \mathbf{R} \to \mathbf{C}$ by the formula

$$\hat{f}(t) = \int_{-\infty}^{+infty} f(x)e^{-ixt}dx$$

for $t \in \mathbf{R}$.

Henri Lebesgue, who introduced the new integration that we will study in this course as the beginning of the 20th century used to present the fundamental idea of his method as follows. Suppose you wish to compute the sum S (seen as a discrete analogue of the integral) of finitely many real numbers, say

$$a_1, \ a_2, \ \ldots, \ a_n$$

(which might be, as in the nice illustration that Lebesgue used, the successive coins found in your pocket, as you try to ascertain your wealth). The "Riemann" way is to take each coin in turn and add their values one by one:

$$S = a_1 + \cdots + a_n.$$

However, there is a different way to do this, which in fact is almost obvious (in the example of counting money): one can take all the coins, sort them into stacks corresponding to the possible values they represent, and compute the sum by adding the different values multiplied by the number of each of the coins that occur: if $b_1, \ldots, b_m$ are the different values among the a_i's, let

$$N_j = \text{number of } i \leqslant n \text{ with } a_i = b_j,$$

denote the corresponding multiplicity, and then we have

$$S = b_1 \times N_1 + \cdots + b_m \times N_m.$$

One essential difference is that in the second procedure, the order in which the coins appear is irrelevant; this is what will allow the definition of integration with respect to a measure to be uniformly valid and suitable for integrating over almost arbitrary sets, and in particular over any space $\mathbf{R}^n$ with $n \geqslant 1$.

To implement the idea behind Lebesgue's story, the first step, however, is to be able to give a meaning to the analogue of the multiplicities N_j when dealing with functions taking possibly infinitely many values, and defined on continuous sets. The intuitive meaning is that it should be a measure of the size of the set of those x where a function $f(x)$ takes values equal, or close, to some fixed value. Such a set, for an arbitrary function, may be quite complicated, and the construction requires a certain amount of work and formalism. But there are rich rewards that come with this effort.

Notation

Measure theory will require (as a convenient way to avoid many technical complications where otherwise a subdivision in cases would be needed) the definition of some arithmetic operations in the set

$$[0, +\infty] = [0, +\infty[\cup \{+\infty\}$$

where the extra element is a symbol subject to the following rules:

(1) Adding and multiplying elements $\neq +\infty$ is done as usual;
(2) We have $x + \infty = \infty + x = \infty$ for all $x \geqslant 0$;
(3) We have $x \cdot \infty = \infty \cdot x = \infty$ *if* $x > 0$;
(4) We have $0 \cdot \infty = \infty \cdot 0 = 0$.

Note, in particular, this last rule, which is somewhat surprising: it will be justified quite quickly.

It is immediately checked that these two operations on $[0, +\infty]$ are commutative, associative, with multiplication distributive with respect to addition.

The set $[0, +\infty]$ is also ordered in the obvious way that is suggested by the intuition: we have $a \leqslant +\infty$ for all $a \in [0, +\infty]$ and $a < +\infty$ if and only if $a \in [0, +\infty[\subset \mathbf{R}$. Note that

$$a \leqslant b, c \leqslant d \text{ imply } a + c \leqslant b + d \text{ and } ac \leqslant bd$$

if a, b, c and $d \in [0, +\infty]$. Note also, however, that, $a + b = a + c$ or $ab = ac$ do not imply $b = c$ now (think of $a = +\infty$).

Remark. Warning: using the subtraction and division operations is not permitted when working with elements which may be $= +\infty$.

For readers already familiar with topology, we note that there is an obvious topology on $[0, +\infty]$ extending the one on non-negative numbers, for which a sequence (a_n) of real numbers converges to $+\infty$ if and only $a_n \to +\infty$ in the usual sense. This topology is such that the open neighborhoods of $+\infty$ are exactly the sets which are complements of bounded subsets of $\mathbf{R}$.

We also recall the following fact: given $a_n \in [0, +\infty]$, the series $\sum a_n$ is always convergent in $[0, +\infty]$; its sum is real (i.e., not $= +\infty$) if and only if all terms are real and the partial sums form a bounded sequence: for some $M < +\infty$, we have

$$\sum_{n=1}^{N} a_n \leqslant M$$

for all N.

We also recall the definition of the limsup and liminf of a sequence (a_n) of real numbers, namely

$$\limsup_n a_n = \lim_{k \to +\infty} \sup_{n \geqslant k} a_n, \quad \liminf_n a_n = \lim_{k \to +\infty} \inf_{n \geqslant k} a_n \ ;$$

each of these limits exist in $[-\infty, +\infty]$ as limits of monotonic sequences (non-increasing for limsup and non-decreasing for the liminf). The sequence (a_n) converges if and only if $\limsup_n a_n = \liminf_n a_n$, and the limit is this common value.

Note also that if $a_n \leqslant b_n$, we get

(0.5) $$\limsup_n a_n \leqslant \limsup_n b_n \text{ and } \liminf_n a_n \leqslant \liminf_n b_n.$$

We finally recall some notation concerning equivalence relations and quotient sets (this will be needed to defined L^p spaces). For any set X, an equivalence relation $\sim$ on X is a relation between elements of X, denoted $x \sim y$, such that

$$x \sim x, \quad x \sim y \text{ if and only if } y \sim x, \quad x \sim y \text{ and } y \sim z \text{ imply } x \sim z$$

note that equality $x = y$ satisfies these facts. For instance, given a set Y and a map $f : X \to Y$, we can define an equivalence relation by defining $x \sim_f y$ if and only if $f(x) = f(y)$.

Given an equivalence relation $\sim$ and $x \in X$, the *equivalence class* $\pi(x)$ of x is the set

$$\pi(x) = \{y \in X \mid y \sim x\} \subset X.$$

The set of all equivalence classes, denoted $X/\sim$, is called the quotient space of X modulo $\sim$; we then obtain a map $\pi : X \to X/\sim$, which is such that $\pi(x) = \pi(y)$ *if and only if* $x \sim y$. In other words, using π it is possible to transform the test for the relation $\sim$ into a test for equality. Moreover, note that – by construction – this map π is surjective.

To construct a map $f : Y \to X/\sim$, where Y is an arbitrary set, it *is enough* to construct a map $f_1 : Y \to X$ and define $f = \pi \circ f_1$. Constructing a map from $X/\sim$, $g : X/\sim \to Y$, on the other hand, is completely equivalent with constructing a map $g_1 : X \to Y$ such that $x \sim y$ implies $g_1(x) = g_1(y)$. Indeed, if that is the case, we see that the value $g_1(x)$ depends only on the equivalence class $\pi(x)$, and one may define unambiguously $g(\pi(x)) = g(x)$ for any $\pi(x) \in X/\sim$. The map g is said to be *induced* by g_1.

As a special case, if V is a vector space, and if $W \subset V$ is a vector subspace, one denotes by V/W the quotient set of V modulo the relation defined so that $x \sim y$ if and only if $x - y \in W$. Then, the maps induced by

$$+ : (x, y) \mapsto x + y \text{ et } \cdot : (\lambda, x) \mapsto \lambda x$$

are well-defined (using the fact that W is itself a vector space); they define on the quotient V/W a vector-space structure, and it is easy to check that the map $\pi : V \to V/W$ is then a surjective linear map.

Some constructions and paradoxical-looking facts in measure theory and integration theory depend on the Axiom of Choice. Here is its formulation:

Axiom. Let X be an arbitrary set, and let $(X_i)_{i \in I}$ be an arbitrary family of *non-empty* subsets of X, with arbitrary index set I. Then there exists a (non-unique, in general) map

$$f : I \to X$$

with the property that $f(i) \in X_i$ for all i.

In other words, the map f "chooses" one element out of each of the sets X_i. This seems quite an obvious fact, but we will see some strange-looking consequences...

We will use the following notation:

(1) For a set X, $|X| \in [0, +\infty]$ denotes its cardinal, with $|X| = \infty$ if X is infinite. There is no distinction between the various infinite cardinals.
(2) If E and F are vector-spaces (over $\mathbf{R}$ or $\mathbf{C}$), we denote $L(E, F)$ the vector space of linear maps from E to F, and write $L(E)$ instead of $L(E, E)$.

CHAPTER 1

Measure theory

The first step in measure theory is somewhat unintuitive. The issue is the following: in following on the idea described at the end of the introduction, it would seem natural to try to define a "measure of size" $\lambda(X) \geqslant 0$ of subsets $X \subset \mathbf{R}$ from which integration can be built. This would generalize the natural size $b - a$ of an interval $[a, b]$, and in this respect the following assumptions seem perfectly natural:

- For any interval $[a, b]$, we have $\lambda([a, b]) = b - a$, and $\lambda(\emptyset) = 0$; moreover $\lambda(X) \leqslant \lambda(Y)$ if $X \subset Y$;
- For any subset $X \subset \mathbf{R}$ and $t \in \mathbf{R}$, we have
$$\lambda(X_t) = \lambda(X), \text{ where } X_t = t + X = \{x \in \mathbf{R} \mid x - t \in X\}$$
(invariance by translation).
- For any sequence $(X_n)_{n\geqslant 1}$ of subsets of $\mathbf{R}$, such that $X_n \cap X_m = \emptyset$ if $n \neq m$, we have
$$\lambda\Bigl(\bigcup_{n\geqslant 1} X_n\Bigr) = \sum_{n\geqslant 1} \lambda(X_n), \tag{1.1}$$
where the sum on the right-hand side makes sense in $[0, +\infty]$.

However, it is a fact that *because of the Axiom of Choice*, these conditions are not compatible, and there is no map λ defined on all subsets of $\mathbf{R}$ with these properties.

Proof. Here is one of the well-known constructions that shows this (due to the italian mathematician Giuseppe Vitali). We consider the quotient set $X = \mathbf{R}/\mathbf{Q}$ (i.e., the quotient modulo the equivalence relation given by $x \sim y$ if $x - y \in \mathbf{Q}$); by the Axiom of Choice, there exists a map
$$f : X \to \mathbf{R}$$
which associates a single element in it to each equivalence class; by shifing using integers, we can in fact assume that $f(X) \subset [0, 1]$ (replace $f(x)$ by its fractional part if needed). Denote then $N = f(X) \subset \mathbf{R}$. Considering $\lambda(N)$ we will see that this can not exist! Indeed, note that, by definition of equivalence relations, we have
$$\mathbf{R} = \bigcup_{t\in\mathbf{Q}} (t + N),$$
over the *countable* set $\mathbf{Q}$. Invariance under translation implies that $\lambda(t + N) = \lambda(N)$ for all N, and hence we see that if $\lambda(N)$ were equal to 0, it would follow by the countable additivity that $\lambda(\mathbf{R}) = 0$, which is absurd. But if we assume that $\lambda(N) = \lambda(t+N) = c > 0$ for all $t \in \mathbf{Q}$, we obtain another contradiction by considering the (still disjoint) union
$$M = \bigcup_{t\in[0,1]\cap\mathbf{Q}} (t + N),$$

because $M \subset [0,2]$ (recall $N \subset [0,1]$), and thus

$$2 = \lambda([0,2]) \geqslant \lambda(M) = \sum_{t\in[0,1]\cap\mathbf{Q}} \lambda(t+N) = \sum_{t\in[0,1]\cap\mathbf{Q}} c = +\infty!$$

$\square$

Since abandoning the requirement (1.1) turns out to be too drastic to permit a good theory (it breaks down any limiting argument), the way around this difficulty has been to restrict the sets for which one tries to define such quantities as the measure $\lambda(X)$. By considering only suitable "well-behaved sets", it turns out that one can avoid the problem above. However, it was found that there is no unique notion of "well-behaved sets" suitable for all sets on which we might want to integrate functions,[1] and therefore one proceeds by describing *axiomatically* the common properties that characterize the collection of well-behaved sets.

1.1. Algebras, σ-algebras, etc

Here is the first formal definition of a collection of "well-behaved sets", together with the description of those maps which are adapted to this type of structures.

Definition 1.1.1. Let X be any set.

(1) A σ-algebra on X is a family $\mathcal{M}$ of subsets of X such that the following conditions hold:

(i) We have $\emptyset \in \mathcal{M}$, $X \in \mathcal{M}$.

(ii) If $Y \in \mathcal{M}$, then the complement set $X - Y = \{x \in X \mid x \notin Y\}$ is also in $\mathcal{M}$.

(iii) If (Y_n) is any countable family of subsets $Y_n \in \mathcal{M}$, then

$$\bigcup_{n\geqslant 1} Y_n \in \mathcal{M} \text{ and } \bigcap_{n\geqslant 1} Y_n \in \mathcal{M}. \tag{1.2}$$

A set $Y \in \mathcal{M}$ is said to be *measurable* for $\mathcal{M}$. The pair $(X, \mathcal{M})$ is called a *measurable space*.

(2) Let $(X, \mathcal{M})$ and $(X', \mathcal{M}')$ be measurable spaces. A map $f : X \to X'$ is *measurable* with respect to $\mathcal{M}$ and $\mathcal{M}'$ if, for all $Y \in \mathcal{M}'$, the inverse image

$$f^{-1}(Y) = \{x \in X \mid f(x) \in Y\}$$

is in $\mathcal{M}$.

Remark 1.1.2. In addition to the above, note already that if Y and Z are measurable sets, then $Y - Z = Y \cap (X - Z)$ is also measurable.

The following lemma is essentially obvious but also very important:

Lemma 1.1.3. (1) *Let $(X, \mathcal{M})$ be a measurable space. The identify map $(X, \mathcal{M}) \to (X, \mathcal{M})$ sending x to itself is measurable.*

(2) *Any constant map is measurable.*

(3) *Let $X \xrightarrow{f} X'$ and $X' \xrightarrow{g} X''$ be measurable maps. The composite $g \circ f : X \to X''$ is also measurable.*

[1] For subsets of $\mathbf{R}$, it is possible to proceed by using the Carathéodory extension theorem which gives a fairly direct approach.

PROOF. For (1) and (2), there is almost nothing to say; for (3), it is enough to remark that

$$(g \circ f)^{-1}(Y'') = f^{-1}(g^{-1}(Y'')).$$

□

EXAMPLE 1.1.4. (1) For any X, one can take $\mathcal{M} = \mathcal{M}_{min} = \{X, \emptyset\}$; this is the smallest possible σ-algebra on X.

(2) Similarly, the largest possible σ-algebra on X is the set $\mathcal{M}_{max}$ of all subsets of X. Of course, any map $(X, \mathcal{M}_{max}) \to (X', \mathcal{M}')$ is measurable. Although we shall see that $\mathcal{M}_{max}$ is not suitable for defining integration when X is a "big" set, it is the most usual σ-algebra used when X is either finite or countable.

(3) Let $\mathcal{M}'$ be a σ-algebra on a set X', and let $f : X \to X'$ be any map; then defining

$$\mathcal{M} = f^{-1}(\mathcal{M}') = \{f^{-1}(Y) \mid Y \in \mathcal{M}'\}$$

we obtain a σ-algebra on X, called the inverse image of $\mathcal{M}'$. Indeed, the formulas

$$f^{-1}(\emptyset) = \emptyset, \quad f^{-1}(X' - Y') = X - f^{-1}(Y'),$$

$$f^{-1}\Big(\bigcup_{i \in I} Y_i\Big) = \bigcup_i f^{-1}(Y_i), \quad f^{-1}\Big(\bigcap_{i \in I} Y_i\Big) = \bigcap_i f^{-1}(Y_i) \tag{1.3}$$

(valid for any index set I) show that $\mathcal{M}$ is a σ-algebra. (On the other hand, the direct image $f(\mathcal{M})$ is not a σ-algebra in general, since $f(X)$ might not be all of X', and this would prevent X' to lie in $f(\mathcal{M})$). This inverse image is the smallest σ-algebra such that

$$f : (X, f^{-1}(\mathcal{M})) \to (X', \mathcal{M}')$$

becomes measurable.

The following special case is often used without explicit mentioning: let $(X, \mathcal{M})$ be a measurable space and let $X' \subset X$ be any fixed subset of X (not necessarily in $\mathcal{M}$). Then we can define a σ-algebra on X' by putting

$$\mathcal{M}' = \{Y \cap X' \mid Y \in \mathcal{M}\} ;$$

it is simply the inverse image σ-algebra $i^{-1}(\mathcal{M})$ associated with the inclusion map $i : X' \hookrightarrow X$. Note that if $X' \in \mathcal{M}$, the following simpler description

$$\mathcal{M}' = \{Y \in \mathcal{M} \mid Y \subset X'\},$$

holds, but it is not valid if $X' \notin \mathcal{M}$.

(4) Let $(\mathcal{M}_i)_{i \in I}$ be σ-algebras on a fixed set X, with I an arbitrary index set. Then the intersection

$$\bigcap_i \mathcal{M}_i = \{Y \subset X \mid Y \in \mathcal{M}_i \text{ for all } i \in I\}$$

is still a σ-algebra on X. (Not so, in general, the union, as one can easily verify).

(5) Let $(X, \mathcal{M})$ be a measurable space, and $Y \subset X$ an arbitrary subset of X. Then $Y \in \mathcal{M}$ if and only if the *characteristic function*

$$\chi_Y : (X, \mathcal{M}) \to (\{0, 1\}, \mathcal{M}_{max})$$

defined by

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{otherwise} \end{cases}$$

is measurable. This is clear, since $\chi_Y^{-1}(\{0\}) = X - Y$ and $\chi_Y^{-1}(\{1\}) = Y$.

Remark 1.1.5. In probabilistic language, it is customary to denote a measurable space by (Ω, Σ); an element $\omega \in \Omega$ is called an "sample event" and $A \subset \Sigma$ is called an "event".

The intersection of two events corresponds to the logical "and", and the union to "or". Thus, for instance, if (A_n) is a countable family of events (or properties), one can say that the event "all the A_n hold" is still an event, since the intersection of the A_n is measurable. Similarly, the event "at least one of the A_n holds" is measurable (it is the union of the A_n).

It is difficult to describe completely explicitly the more interesting σ-algebras which are involved in integration theory. However, an indirect construction, which is often sufficiently handy, is given by the following construction:

Definition 1.1.6. (1) Let X be a set and $\mathcal{A}$ a family of subsets of X. The σ-algebra generated by $\mathcal{A}$, denoted $\sigma(\mathcal{A})$, is the smallest σ-algebra containing $\mathcal{A}$, i.e., it is given by

$$\sigma(\mathcal{A}) = \{Y \subset X \mid Y \in \mathcal{M} \text{ for any } \sigma\text{-algebra } \mathcal{M} \text{ with } \mathcal{A} \subset \mathcal{M}\}$$

(in other words, it is the intersection of all σ-algebras containing $\mathcal{A}$). This σ-algebra is called σ-algebra generated by $\mathcal{A}$.

(2) Let $(X, \mathcal{T})$ be a topological space. The Borel σ-algebra on X, denoted $\mathcal{B}_X$, is the σ-algebra generated by the collection $\mathcal{T}$ of open sets in X.

(3) Let $(X, \mathcal{M})$ and $(X', \mathcal{M}')$ be measurable spaces; the product σ-algebra on $X \times X'$ is the σ-algebra denoted $\mathcal{M} \otimes \mathcal{M}'$ which is generated by all the sets of the type $Y \times Y'$ where $Y \in \mathcal{M}$ and $Y' \in \mathcal{M}'$.

Remark 1.1.7. (1) If $(X, \mathcal{T})$ is a topological space, we can immediately check that $\mathcal{B}$ is generated either by the closed sets or the open sets (since the closed sets are the complements of the open ones, and conversely). If $X = \mathbf{R}$ with its usual topology (which is the most important case), the Borel σ-algebra contains all intervals, whether closed, open, half-closed, half-infinite, etc. For instance:

$$[a, b] = \mathbf{R} - (]-\infty, a[\cup]b, +\infty[), \text{ and }]a, b] = [a, b] \cap]a, +\infty[. \tag{1.4}$$

Moreover, the Borel σ-algebra on $\mathbf{R}$ is in fact generated by the much smaller collection of closed intervals $[a, b]$, or by the intervals $]-\infty, a]$ where $a \in \mathbf{R}$. Indeed, using arguments as above, the σ-algbera $\mathcal{M}$ generated by those intervals contains all open intervals, and then one can use the following

Lemma 1.1.8. *Any open set U in $\mathbf{R}$ is a disjoint union of a family, either finite or countable, of open intervals.*

Proof. Indeed, these intervals are simply the connected components of the set U; there are at most countably many of them because, between any two of them, it is possible to put some rational number. □

By convention, when X is a topological space, we consider the Borel σ-algebra on X when speaking of measurability issues, unless otherwise specified. This applies in particular to functions $(X, \mathcal{M}) \to \mathbf{R}$: to say that such a function is measurable means with respect to the Borel σ-algebra.

(2) If $(X, \mathcal{M})$ and $(X', \mathcal{M}')$ are measurable spaces, one may check also that the product σ-algebra $\mathcal{M} \otimes \mathcal{M}'$ defined above is the smallest σ-algebra on $X \times X'$ such that the projection maps

$$p_1 : X \times X' \to X \text{ and } p_2 : X \times X' \to X'$$

are both measurable.

Indeed, for a given σ-algebra $\mathcal{N}$ on $X \times X'$, the projections are measurable if and only if $p_2^{-1}(Y') = X \times Y' \in \mathcal{N}$ and $p_1^{-1}(Y) = Y \times X' \in \mathcal{N}$ for any $Y \in \mathcal{M}$, $Y' \in \mathcal{M}'$. Since

$$Y \times Y' = (Y \times X') \cap (X \times Y'),$$

we see that these two types of sets generate the product σ-algebra, which is in turn generated by all $Y \times Y'$.

(3) The Borel σ-algebra on $\mathbf{C} = \mathbf{R}^2$ is the same as the product σ-algebra $\mathcal{B} \otimes \mathcal{B}$, where $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbf{R}$; this is due to the fact that any open set in $\mathbf{C}$ is a countable union of sets of the type $I_1 \times I_2$ where $I_i \subset \mathbf{R}$ is an open interval. Moreover, one can check that the restriction to the real line $\mathbf{R}$ of the Borel σ-algebra on $\mathbf{C}$ is simply the Borel σ-algebra on $\mathbf{R}$ (because the inclusion map $\mathbf{R} \hookrightarrow \mathbf{C}$ is continuous; see Corollary 1.1.10 below).

(4) If we state that $(X, \mathcal{M}) \to \mathbf{C}$ is a measurable function with no further indication, it is implied that the target $\mathbf{C}$ is given with the Borel σ-algebra. Similarly for target space $\mathbf{R}$.

In probability theory, a measurable map $\Omega \to \mathbf{C}$ is called a *random variable*.[2]

(5) In the next chapters, we will consider functions $f : X \to [0, +\infty]$. Measurability is then always considered with respect to the σ-algebra on $[0, +\infty]$ generated by $\mathcal{B}_{[0,+\infty[}$ and the singleton $\{+\infty\}$. In other words, f is measurable if and only if

$$f^{-1}(U), \qquad f^{-1}(+\infty) = \{x \mid f(x) = +\infty\}$$

are in $\mathcal{M}$, where U runs over all open sets of $[0, +\infty[$.

The following lemma is important to ensure that σ-algebras indirectly defined as generated by a collection of sets are accessible.

Lemma 1.1.9. (1) *Let $(X, \mathcal{M})$ and $(X', \mathcal{M}')$ be measurable spaces such that $\mathcal{M}' = \sigma(\mathcal{A}')$ is generated by a collection of subsets $\mathcal{A}'$. A map $f : X \to X'$ is measurable if and only if $f^{-1}(\mathcal{A}') \subset \mathcal{M}$.*

(2) *In particular, for any measurable space $(X, \mathcal{M})$, a function $f : X \to \mathbf{R}$ is measurable if and only if*

$$f^{-1}(]-\infty, a]) = \{x \in X \mid f(x) \leqslant a\}$$

is measurable for all $a \in \mathbf{R}$, and a function $f : X \to [0, +\infty]$ is measurable if and only if

$$f^{-1}(]-\infty, a]) = \{x \in X \mid f(x) \leqslant a\}$$

is measurable for all $a \in [0, +\infty[$.

(3) *Let $(X, \mathcal{M})$, $(X', \mathcal{M}')$ and $(X'', \mathcal{M}'')$ be measurable spaces. A map $f : X'' \to X \times X'$ is measurable for the product σ-algebra on $X \times X'$ if and only if $p_1 \circ f : X'' \to X$ and $p_2 \circ f : X'' \to X'$ are measurable.*

Proof. Part (2) is a special case of (1), if one takes further into account (in the case of functions taking values in $[0, +\infty]$) that

$$f^{-1}(+\infty) = \bigcap_{n \geqslant 1} \{x \in X \mid f(x) > n\} = \bigcap_{n \geqslant 1} (X - f^{-1}(]-\infty, n])).$$

This first point follows from the formula

$$f^{-1}(\sigma(\mathcal{A}')) = \sigma(f^{-1}(\mathcal{A}')), \tag{1.5}$$

[2]. Quite often, in the literature, random variables are assumed to be real-valued.

since we deduce then that, using the assumption $f^{-1}(\mathcal{A}') \subset \mathcal{M}$, that

$$f^{-1}(\mathcal{M}') = f^{-1}(\sigma(\mathcal{A}')) = \sigma(f^{-1}(\mathcal{A}')) \subset \sigma(\mathcal{M}) = \mathcal{M},$$

which is exactly what it means for f to be measurable.

The left-hand side of (1.5) is an inverse image σ-algebra that contains $f^{-1}(\mathcal{A}')$, and therefore it contains the σ-algebra generated by this family of sets, which is the right-hand side.

Conversely, notice that

$$\mathcal{M}'' = \{Y \mid f^{-1}(Y) \in \sigma(f^{-1}(\mathcal{A}'))\}$$

is a σ-algebra on X' (see (1.3)), which contains $\mathcal{A}'$, and therefore also $\sigma(\mathcal{A}')$. Consequently, we get

$$f^{-1}(\sigma(\mathcal{A}')) \subset f^{-1}(\mathcal{M}'') \subset \sigma(f^{-1}(\mathcal{A}')),$$

as desired.

As for (3), the composite maps $p_1 \circ f$ and $p_2 \circ f$ are of course measurable if f is (see Remark 1.1.7, (2)). Conversely, to check that f is measurable, it suffices by (1) to check that $f^{-1}(Y \times Y') \in \mathcal{M}''$ for any $Y \in \mathcal{M}$, $Y' \in \mathcal{M}'$. But since $f(x) = (p_1 \circ f(x), p_2 \circ f(x))$, we get

$$f^{-1}(Y \times Y') = (p_1 \circ f)^{-1}(Y) \cap (p_2 \circ f)^{-1}(Y'),$$

which gives the result. □

Corollary 1.1.10. (1) *Let $f : X \to X'$ be a continuous map between topological spaces. Then f is measurable with respect to the Borel σ-algebras on X and X'.*

(2) *Let $(X, \mathcal{M})$ be a measurable space, and let f, $g : X \to \mathbf{C}$ be measurable maps. Then $f \pm g$ and fg are measurable, and if $g(x) \neq 0$ for $x \in X$, the inverse $1/g$ is measurable. In particular, the set of complex-valued measurable functions on X is a vector space with operations given by addition of functions and multiplication by constants.*

(3) *A function $f : X \to \mathbf{C}$ is measurable for the Borel σ-algebra on $\mathbf{C}$ if and only if* $\mathrm{Re}(f)$ *are* $\mathrm{Im}(f)$ *are measurable as functions $X \to \mathbf{R}$.*

Proof. Part (1) is immediate from Lemma 1.1.9, (1), since continuity means that $f^{-1}(U)$ is open for any open set in X'. To prove (2), we write, for instance,

$$f + g = p \circ (f \times g),$$

where

$$p : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$$

is the addition map, and $f \times g : x \mapsto (f(x), g(x))$. According to Lemma 1.1.9, (3), this last map $f \times g$ is measurable, and according to (1), the map p is measurable (because it is continuous). By composition, $f + g$ is also measurable. Similar arguments apply to $f - g$, fg and $1/g$.

Finally, (3) is a special case of (2) according to Remark 1.1.7, (3). □

Example 1.1.11. In probabilistic language, if $X : \Omega \to \mathbf{C}$ is a random variable, an event of the type $X^{-1}(Y)$ is denoted simply $\{X \in Y\}$. For instance, one commonly writes

$$\{X > a\} = \{\omega \mid X(\omega) > a\}$$

for $a \in \mathbf{R}$. Moreover, if $f : \mathbf{C} \to \mathbf{C}$ is a measurable function, it will often be convenient to denote by $f(X)$ the composite random variable $Y = f \circ X$.

It may have been noticed that, up to now, Axiom (iii) in the definition of a σ-algebra has not been used in any essential way. It is however indispensible as soon as sequences of functions – which are so important in analysis – come into play. For instance, the following crucial lemma would not otherwise be accessible:

LEMMA 1.1.12. *Let $(X, \mathcal{M})$ be a measurable space.*

(1) *Let (f_n), $n \geqslant 1$, be a sequence of measurable real-valued functions, such that $f_n(x) \to f(x)$ for any $x \in X$, i.e., such that (f_n) converges pointwise to a limiting function f. Then this function f is measurable.*

(2) *More generally, the functions defined by*

$$(\limsup f_n)(x) = \limsup f_n(x), \quad \liminf f_n, \quad (\sup f_n)(x) = \sup_n f_n(x), \quad \inf f_n$$

are measurable.

Note that, even if the f_n are continuous (when this makes sense), there is no reason that the pointwise limit f is continuous.

PROOF. Because the limit $f(x)$ of $f_n(x)$ is also equal to its limsup, it is enough to prove the second part. Moreover, we have

$$\limsup f_n(x) = \lim_k \sup_{n \geqslant k} f_n(x) = \inf_k \sup_{n \geqslant k} f_n(x),$$

(since it is a monotonic limit of a non-increasing sequence), and hence it is enough to prove the result for $\inf f_n$ and $\sup f_n$ (for arbitrary sequences of functions); replacing f_n with $-f_n$, it is even enough to consider the case of $\sup f_n$.

Thus, let $g(x) = \sup_{n \geqslant 1} f_n(x)$, and assume this takes real values (i.e., the sequences $(f_n(x))$ are all bounded, for $x \in X$). By Lemma 1.1.9, (1) and Remark 1.1.7, (1), it suffices to prove that for any $a \in \mathbf{R}$, we have

$$E_a = \{x \mid g(x) \leqslant a\} \in \mathcal{M}.$$

But $g(x) = \sup f_n(x) \leqslant a$ if and only if $f_n(x) \leqslant a$ for all n, and hence we can write

$$E_a = \bigcap_n \{x \mid f_n(x) \leqslant a\},$$

which is a countable union of measurable sets (because each f_n is measurable); by (1.2), we get $E_a \in \mathcal{M}$ as desired.

Otherwise, if g takes the value ∞, notice that

$$\{g = \infty\} = \{x \in X \mid \sup f_n(x) = \infty\} = \cap_{M \geqslant 1} \cup_N \cap_{n \geqslant N} \{x \mid f_n(x) \geqslant M\}.$$

$\square$

REMARK 1.1.13. (1) Let f and g be measurables functions on $(X, \mathcal{M})$; it follows from this lemma, in particular, that $\sup(f, g)$ and $\inf(f, g)$ are also measurable.

It follows therefore that if f is real-valued, the absolute value

$$|f| = \sup(f, -f) \tag{1.6}$$

is measurable, and that the so-called *positive and negative parts* of f, defined by

$$f^+ = \sup(f, 0), \text{ and } f^- = -\inf(f, 0) \tag{1.7}$$

are also measurable. Note that $f^+ \geqslant 0$, $f^- \geqslant 0$ and

$$f = f^+ - f^- \text{ while } |f| = f^+ + f^- \tag{1.8}$$

These are often very convenient representations of f (resp. $|f|$) as difference (resp. sum) of two non-negative measurable functions.

(2) Let (f_n) be a sequence of complex-valued measurable functions on X. Then the set

$$Y = \{x \in X \mid f_n(x) \text{ converges to some limit } \}$$

is a measurable set in X. Indeed, by translating the Cauchy criterion in terms of set operations, one can write

$$Y = \bigcap_{k\geqslant 1} \bigcup_{N\geqslant 1} \bigcap_{n,m\geqslant N} \{x \mid |f_n(x) - f_m(x)| < 1/k\} \tag{1.9}$$

and the result follows because $|f_n - f_m|$ is measurable.

1.2. Measure on a σ-algebra

We have now defined (through axiomatic properties) what are well-behaved collections of sets that one may wish to "measure". Because of the generality, this measure (which may not be definable at all!) does not have an intrinsic, unique, meaning, and we also use abstract axiomatic properties to define a measure.

Definition 1.2.1. (1) Let $(X, \mathcal{M})$ be a measurable space. A *measure* μ on $(X, \mathcal{M})$ is a map

$$\mu : \mathcal{M} \to [0, +\infty]$$

such that $\mu(\emptyset) = 0$ and

$$\mu\Bigl(\bigcup_n Y_n\Bigr) = \sum_n \mu(Y_n) \tag{1.10}$$

for any countable family of pairwise disjoints measurable sets $Y_n \in \mathcal{M}$. The triple $(X, \mathcal{M}, \mu)$ is called a *measured space.*

(2) A measure μ is said to be *finite* if $\mu(X) < +\infty$, and is said to be *σ-finite* if one can write X as a countable union of subsets with finite measure: there exist $X_n \in \mathcal{M}$ such that

$$X = \bigcup_{n\geqslant 1} X_n, \text{ and } \mu(X_n) < +\infty \text{ for all } n.$$

(3) A *probability measure* is a measure μ such that $\mu(X) = 1$.

Remark 1.2.2. For any finite measure μ, if $\mu(X) > 0$, we can define $\mu'(Y) = \mu(Y)/\mu(X)$ for $Y \in \mathcal{M}$ and obtain a probability measure. So the theory of finite measures is almost equivalent with that of probability measures.

The most commonly used measures are σ-finite; this is important, for instance, in the theory of multiple integrals, as we will see.

In probabilistic language, one uses a probability measure on (Ω, Σ), commonly denoted P, and $P(E) \in [0, 1]$, for $E \in \Sigma$, is the probability of an event E. The triple (Ω, Σ, P) is called a *probability space.*

The definition implies quickly the following useful properties:

Proposition 1.2.3. *Let μ be a measure on $(X, \mathcal{M})$.*

(1) *For Y, $Z \in \mathcal{M}$, with $Y \subset Z$, we have $\mu(Y) \leqslant \mu(Z)$, and more precisely*

$$\mu(Z) = \mu(Y) + \mu(Z - Y).$$

(2) *For Y, $Z \in \mathcal{M}$, we have*

$$\mu(Y \cup Z) + \mu(Y \cap Z) = \mu(Y) + \mu(Z).$$

(3) *If* $Y_1 \subset \cdots \subset Y_n \subset \cdots$ *is an increasing*[3] *sequence of measurable sets, then*

$$\mu\Big(\bigcup_n Y_n\Big) = \lim_{n\to+\infty} \mu(Y_n) = \sup_{n\geqslant 1} \mu(Y_n).$$

(4) *If* $Y_1 \supset \cdots \supset Y_n \supset \cdots$ *is a decreasing sequence of measurable sets, and if furtherfore* $\mu(Y_1) < +\infty$*, then we have*

$$\mu\Big(\bigcap_n Y_n\Big) = \lim_{n\to+\infty} \mu(Y_n) = \inf_{n\geqslant 1} \mu(Y_n).$$

(5) *For any countable family* (Y_n) *of measurable sets, we have*

$$\mu\Big(\bigcup_n Y_n\Big) \leqslant \sum_n \mu(Y_n).$$

Proof. In each case, a quick drawing or diagram is likely to be more insightful than the formal proof that we give.

(1): we have a disjoint union

$$Z = Y \cup (Z - Y)$$

and hence $\mu(Z) = \mu(Y) + \mu(Z - Y) \geqslant \mu(Y)$, since μ takes non-negative values.

(2): similarly, we note the disjoint unions

$$Y \cup Z = Y \cup (Z - Y) \text{ and } Z = (Z - Y) \cup (Z \cap Y),$$

hence $\mu(Y \cup Z) = \mu(Y) + \mu(Z - Y)$ and $\mu(Z) = \mu(Z - Y) + \mu(Z \cap Y)$. This gives $\mu(Y \cup Z) = \mu(Y) + \mu(Z) - \mu(Z \cap Y)$, as claimed.

(3): let $Z_1 = Y_1$ and $Z_n = Y_n - Y_{n-1}$ for $n \geqslant 2$. These are all measurable, and we have

$$Y = Y_1 \cup Y_2 \cup \cdots = Z_1 \cup Z_2 \cup \cdots.$$

Moreover, the Z_i are now disjoint (because the original sequence (Y_n) was increasing) for $j > 0$ – indeed, note that

$$Z_{i+j} \cap Z_i \subset Z_{i+j} \cap Y_i \subset Z_{i+j} \cap Y_{i+j-1} = \emptyset.$$

Consequently, using (1.10) we get

$$\mu(Y) = \sum_n \mu(Z_n) = \lim_{k\to+\infty} \sum_{1\leqslant n\leqslant k} \mu(Z_n);$$

and since the same argument also gives

$$\mu(Y_k) = \sum_{1\leqslant n\leqslant k} \mu(Z_n),$$

we get the result.

(4): this result complements the previous one, but one must be careful that the additional condition $\mu(Y_1) < +\infty$ is necessary (a counterexample is obtained by taking $Y_n = \{k \geqslant n\}$ in $X = \{n \geqslant 1\}$ with the counting measure described below; then the intersection of all Y_n is empty, although the measure of each Y_n is infinite).

Let $Z_n = Y_1 - Y_n$; the measurable sets Z_n form an increasing sequence, and hence, by the previous result, we have

$$\mu\Big(\bigcup_n Z_n\Big) = \lim_n \mu(Y_1 - Y_n).$$

[3] For inclusion.

Now, on the left-hand side, we have

$$\bigcup_n Z_n = Y_1 - \bigcap_n Y_n,$$

and on the right-hand side, we get $\mu(Y_1 - Y_n) = \mu(Y_1) - \mu(Y_n)$ (because $Y_n \subset Y_1$). Since $\mu(Y_1)$ is not $+\infty$, we can subtract it from both sides to get the result.

(5): The union Y of the Y_n can be seen as the disjoint union of the measurable sets Z_n defined inductively by $Z_1 = Y_1$ and

$$Z_{n+1} = Y_{n+1} - \bigcup_{1 \leqslant i \leqslant n} Y_i$$

for $n \geqslant 1$. We have $\mu(Z_n) \leqslant \mu(Y_n)$ by (1), and from (1.10), it follows that

$$\mu(Y) = \sum_n \mu(Z_n) \leqslant \sum_n \mu(Y_n).$$

$\square$

Before giving some first examples of measures, we introduce one notion that is very important. Let $(X, \mathcal{M}, \mu)$ be a measured space. The measurable sets of measure 0 play a particular role in integration theory, because they are "invisible" to the process of integration.

DEFINITION 1.2.4. Let $(X, \mathcal{M}, \mu)$ be a measured space. A subset $Y \subset X$ is said to be μ-negligible if there exists $Z \in \mathcal{M}$ such that

$$Y \subset Z, \text{ and } \mu(Z) = 0$$

(for instance, any set in $\mathcal{M}$ of measure 0).

If $\mathcal{P}(x)$ is a mathematical property parametrized by $x \in X$, then one says that $\mathcal{P}$ *is true μ-almost everywhere* if

$$\{x \in X \mid \mathcal{P}(x) \text{ is not true}\}$$

is μ-negligible.

REMARK 1.2.5. (1) By the monotonicity formula for countable unions (Proposition 1.2.3, (5)), we see immediately that any countable union of negligible sets is still negligible. Also, of course, the intersection of any collection of negligible sets remains so (as it is contained in any of them).

(2) The definition is more complicated than one might think necessary because although the intuition naturally expects that a subset of a negligible set is negligible, it is not true that any subset of a measurable set of measure 0 is itself measurable (one can show, for instance, that this property is not true for the Lebesgue measure defined on Borel sets of $\mathbf{R}$ – see Section 1.3 for the definition).

If all μ-negligible sets are measurable, the measured space is said to be *complete* (this has nothing to do with completeness in the sense of Cauchy sequences having a limit). The following procedure can be used to construct a natural complete "extension" of a measured space $(X, \mathcal{M}, \mu)$.

Let $\mathcal{M}_0$ denote the collection of μ-negligible sets. Then define

$$\mathcal{M}' = \{Y \subset X \mid Y = Y_1 \cup Y_0 \text{ with } Y_0 \in \mathcal{M}_0 \text{ and } Y_1 \in \mathcal{M}\},$$

the collection of sets which, intuively, "differ" from a measurable set only by a μ-negligible set. Define then $\mu'(Y) = \mu(Y_1)$ if $Y = Y_0 \cup Y_1 \in \mathcal{M}'$ with Y_0 negligible and Y_1 measurable.

Proposition 1.2.6. *The triple* $(X, \mathcal{M}', \mu')$ *is a complete mesured space; the* σ*-algebra* $\mathcal{M}'$ *contains* $\mathcal{M}$*, and* $\mu' = \mu$ *on* $\mathcal{M}$*.*

This proposition is intuively clear, but the verification is somewhat tedious. For instance, to check that the measure μ' is well-defined, independently of the choice of Y_0 and Y_1, assume that $Y = Y_1 \cup Y_0 = Y_1' \cup Y_0'$, with Y_0, Y_0' both negligible. Say that $Y_0' \subset Z_0'$, where Z_0' is measurable of measure 0. Then $Y_1 \subset Y_1' \cup Z_0' \in \mathcal{M}$, hence

$$\mu(Y_1) \leqslant \mu(Y_1') + \mu(Z_0') = \mu(Y_1')$$

and similarly one gets $\mu(Y_1') \leqslant \mu(Y_1)$.

It is a good exercise to check all the remaining points needed to prove the proposition.

Here are now the simplest examples of measures. The most interesting example, the Lebesgue measure, is introduced in Section 1.3, though its rigorous construction will come only later.

Example 1.2.7. (1) For any set X, $\mathcal{M}_{max}$ the σ-algebra of all subsets of X. Then we obtain a measure, called the *counting measure*, on $(X, \mathcal{M}_{max})$ by defining

$$\mu(Y) = |Y|.$$

It is obvious that (1.10) holds in that case. Moreover, only $\emptyset$ is μ-negligible.

(2) Let again X be an arbitrary set and $\mathcal{M}_{max}$ the σ-algebra of all subsets of X. For a fixed $x_0 \in X$, the *Dirac measure at* x_0 is defined by

$$\delta_{x_0}(Y) = \begin{cases} 1 & \text{if } x_0 \in Y \\ 0 & \text{otherwise.} \end{cases}$$

The formula (1.10) is also obvious here: indeed, at most one of a collection of disjoint sets can contain x_0. Here, a set is negligible if and only if it does not contain x_0.

(3) For any *finite* set X, the measure on $\mathcal{M}_{max}$ defined by

$$\mu(Y) = \frac{|Y|}{|X|}$$

is a probability measure.

This measure is the basic tool in "discrete" probability; since each singleton set $\{x\}$ has the same measure

$$\mu(\{x\}) = \frac{1}{|X|},$$

we see that, relative to this measure, all experiences have equal probability of occuring.

(4) Let $X = \mathbf{N}$, with the σ-algebra $\mathcal{M}_{max}$. There is no "uniform" probability measure defined on X, i.e., there is no measure P on $\mathcal{M}$ such that $P(\mathbf{N}) = 1$ and $P(n + A) = P(A)$ for any $A \subset \mathbf{N}$, where $n + A = \{n + a \mid a \in A\}$ is the set obtained by translating A by the amount n.

Indeed, if such a measure P existed, it would follow that $P(\{n\}) = P(n + \{0\}) = P(0)$ for all n, so that, by additivity, we would get

$$P(A) = \sum_{a \in A} P(\{a\}) = |A| P(0)$$

for any finite set A. If $P(0) \neq 0$, it would follow that $P(A) > 1$ if $|A|$ is large enough, which is impossible for a probability measure. If, on the other hand, we had $P(0) = 0$, it would then follow that $P(A) = 0$ for all A using countable additivity.

However, there do exist many probability measures on $(\mathbf{N}, \mathcal{M}_{max})$. More precisely, writing each set as the disjoint union, at most countable, of its one-element sets, we see that it is equivalent to give such a measure P or to give a sequence $(p_k)_{k\in\mathbf{N}}$ of real numbers $p_k \in [0,1]$ such that

$$\sum_{k\geqslant 0} p_k = 1,$$

the correspondance being given by $p_k = P(\{k\})$. If some p_k are equal to 0, the corresponding one-element sets are P-negligible. Indeed, the P-negligible sets are exactly all the subsets of $\{k \in \mathbf{N} \mid p_k = 0\}$.

For instance, the *Poisson measure* with parameter $\lambda > 0$ is defined by

$$P(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}$$

for $k \geqslant 0$.

We now describe some important ways to operate on measures, and to construct new ones from old ones.

PROPOSITION 1.2.8. *Let $(X, \mathcal{M}, \mu)$ be a measured space.*

(1) *For any finite collection $\mu_1, \ldots, \mu_n$ of measures on $(X, \mathcal{M})$, and any choice of real numbers $\alpha_i \in [0, +\infty[$, the measure $\mu = \sum \alpha_i \mu_i$ is defined by*

$$\mu(Y) = \sum_{1\leqslant i\leqslant n} \alpha_i \mu_i(Y)$$

for $Y \in \mathcal{M}$; it is a measure on $(X, \mathcal{M})$.

(2) *Let $f : (X, \mathcal{M}) \to (X', \mathcal{M}')$ be a measurable map. Let*

$$f_*(\mu)(Y) = \mu(f^{-1}(Y)) \text{ for } Y \in \mathcal{M}'.$$

Then $f_*(\mu)$ is a measure one X', called the image measure *of μ under f. It is also sometimes denoted $f(\mu)$.*

If we have another measurable map $g : (X', \mathcal{M}') \to (X'', \mathcal{M}'')$, then we have

$$(g \circ f)_*(\mu) = g_*(f_*(\mu)). \tag{1.11}$$

(3) *For any measurable subset $Y \subset X$, the restriction of μ to the σ-algebra of measurable subsets of Y is a measure on Y for this σ-algebra.*

PROOF. All these facts are immediate. For the second part, use again (1.3) to check (1.10). The formula (1.11) is due to the simple fact that

$$(g \circ f)^{-1}(Y'') = f^{-1}(g^{-1}(Y''))$$

for $Y'' \subset X''$. □

REMARK 1.2.9. In probabilistic language, given a random variable $X : \Omega \to \mathbf{C}$, the image measure $X(P)$ is called the *probability law* of the variable X. Note that it is a probability measure because

$$X(P)(\mathbf{C}) = P(X \in \mathbf{C}) = P(\Omega) = 1.$$

This is a crucial notion, because the underlying space Ω is often left unspecified, and the data for a given probability problem is some collection of random variables (which, for instance, are supposed to give a model of some natural phenomenon), together with some assumptions on their probability laws. These are measures on $\mathbf{C}$, a much more concrete set, and the main point is that knowing the law μ_i of X_i is sufficient to answer all questions relative to probabilities of values of X_i: indeed, by definition, we have

$$P(X_i \text{ satisfies a property } \mathcal{P}) = \mu_i(\{z \in \mathbf{C} \mid \mathcal{P}(z) \text{ is true}\})$$

for any property $\mathcal{P}(z)$ parametrized by $z \in \mathbf{C}$ (assuming the set it defines is measurable, of course).

We now introduce a further definition which is purely probabilistic.

Definition 1.2.10. Let (Ω, Σ, P) be a probability space, A and $B \in \Sigma$ events.

(1) The *conditional probability of A knowing B* is the quantity

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

which is well-defined if $P(B) \neq 0$.

(2) The events A, B are said to be *independent* if

$$P(A \cap B) = P(A)P(B).$$

(3) Let X_1, X_2 be complex-valued random variables on Ω. Then X_1 and X_2 are said to be *independent* if the events

$$\{X_1 \in C\} = X_1^{-1}(C), \quad \text{and} \quad \{X_2 \in D\} = X_2^{-1}(D)$$

are independent for any choice of C, $D \in \mathcal{B}_{\mathbf{C}}$.

(4) More generally, the events in an arbitrary collection $(A_i)_{i \in I}$ of events are *independent* if, for any finite family $A_{i(1)}, \ldots, A_{i(n)}$, we have

$$P(A_{i(1)} \cap \ldots \cap A_{i(n)}) = P(A_{i(1)}) \cdots P(A_{i(n)}),$$

and similarly, the random variables $(X_i)_{i \in I}$ in an arbitrary collection are independent if the events $(\{X_i \in C_i\})_i$ are all independant, for arbitrary Borel sets $(C_i)_{i \in I}$.

Remark 1.2.11. If $P(B) = 0$, we have $P(A \cap B) = 0$ by monotony for any A, hence a negligible set is independent of any other measurable set A.

The following elementary result provides a way to check independence of random variables in a number of cases.

Proposition 1.2.12. (1) *Let (Ω, Σ, P) be a probability space, and $\mathcal{A} \subset \Sigma$ a collection of subsets such that $\sigma(\mathcal{A}) \supset \Sigma$. In the definition of independence, it is enough to check the stated conditions for events chosen in $\mathcal{A}$.*

(2) *Let (X_n), $n \leqslant N$, be independent random variables, and let $\varphi_n : \mathbf{C} \to \mathbf{C}$ be measurable maps. Then the random variables $Y_n = \varphi(X_n)$ are independent.*

Proof. Part (1) is elementary; for instance, for two random variables X and Y, consider the set of pairs (C, D) of events such that $\{X \in C\}$ and $\{Y \in D\}$ are independent. For a fixed C (resp. a fixed D), one can check that the corresponding D's form a σ-algebra containing $\mathcal{A}$, hence containing $\sigma(\mathcal{A}) \supset \Sigma$. The result follows easily from this, and the other cases are simply more complicated in terms of notation.

For (2), note that for arbitrary choices of events C_n, we have

$$(Y_1, \ldots, Y_N) \in C_1 \times \cdots \times C_N$$

if and only if

$$(X_1, \ldots, X_N) \in \varphi_1^{-1}(C_1) \times \cdots \times \varphi_N^{-1}(C_N)$$

hence the assumption that the X_n are independent gives the result for the Y_n. ☐

EXERCISE 1.2.13. Generalize the last result to situations like the following:

(1) Let $X_1, \ldots, X_4$ be independent random variables. Show that

$$P((X_1, X_2) \in C \text{ and } (X_3, X_4) \in D) = P((X_1, X_2) \in C)P((X_3, X_4) \in D)$$

for any C and D is the product σ-algebra $\Sigma \otimes \Sigma$. (Hint: use an argument similar to that of Proposition 1.2.12, (1)).

(2) Let $\varphi_i : \mathbf{C}^2 \to \mathbf{C}$, $i = 1, 2$, be measurable maps. Show that $\varphi_1(X_1, X_2)$ and $\varphi_2(X_3, X_4)$ are independant.

Note that these results are intuitively natural: starting with four "independent" quantities, in the intuitive sense of the word, "having nothing to do with each other", if we perform separate operations on the first two and the last two, the results should naturally themselves be independent...

1.3. The Lebesgue measure

The examples of measures in the previous section are too elementary to justify the amount of formalism involved. Indeed, the theory of integration with respect to a measure only makes good sense when brought together with the following fundamental theorem:

THEOREM 1.3.1. *There exists a unique measure λ on the σ-algebra of Borel subsets of $\mathbf{R}$ with the property that*

$$\lambda([a, b]) = b - a$$

for any real numbers $a \leqslant b$. This measure is called the Lebesgue measure.

REMARK 1.3.2. One can show that the Lebesgue measure, defined on the Borel σ-algebra, is not complete. The completed measure (see Proposition 1.2.6) is also called the Lebesgue measure, and the corresponding complete σ-algebra is called the Lebesgue σ-algebra.

We will prove this theorem only later, since it is quite technical, and we will also refine the statement. It will be interesting to see how a fully satisfactory theory of integration for $\mathbf{R}^n$ can be built by using this theorem as a "black box": the details of the construction are only useful for the study of rather fine properties of functions.

Note that λ is σ-finite, although it is not finite, since

$$\mathbf{R} = \bigcup_{n \geqslant 1} [-n, n].$$

Note also that the Lebesgue measure, restricted to the interval $[0, 1]$ gives an example of probability measure, which is also fundamental.

EXAMPLE 1.3.3. (1) We have $\lambda(\mathbf{N}) = \lambda(\mathbf{Q}) = 0$. Indeed, by definition

$$\lambda(\{x\}) = \lambda([x, x]) = 0 \text{ for any } x \in \mathbf{R},$$

and hence, by countable additivity, we get

$$\lambda(\mathbf{N}) = \sum_{n \in \mathbf{N}} \lambda(\{n\}) = 0,$$

the same property holding for $\mathbf{Q}$ because $\mathbf{Q}$ is also countable. In fact, any countable set is λ-negligible.

(2) There are many λ-negligible sets besides those that are countable. Here is a well-known example: the Cantor set C.

Let $X = [0,1]$. For any fixed integer $b \geqslant 2$, and any $x \in [0,1]$, one can expand x in "base b": there exists a sequence of "digits" $d_i(x) \in \{0, \ldots, b-1\}$ such that

$$x = \sum_{i=1}^{+\infty} d_i(x) b^{-i},$$

which is also denoted $x = 0.d_1 d_2 \ldots$ (the base b being understood from context; the case $b = 10$ corresponds to the customary expansion in decimal digits).

This expansion is not entirely unique (for instance, in base 10, we have

$$0.1 = 0.09999\ldots$$

as one can check by a geometric series computation). However, the set of those $x \in X$ for which there exist more than one base b expansion is countable, and therefore is λ-negligible.

The Cantor set C is defined as follows: take $b = 3$, so the digits are $\{0,1,2\}$, and let

$$C = \{x \in [0,1] \mid d_i(x) \in \{0,2\} \text{ for all } i\},$$

the set of those x which have no digit 1 in their base 3 expansion.

We then claim that

$$\lambda(C) = 0,$$

but C is not countable.

The last property is not difficult to check: to each

$$x = d_1 d_2 \ldots$$

in C (base 3 expansion), we can associate

$$y = e_1 e_2 \ldots, \qquad e_i = \begin{cases} 0 & \text{if } d_i = 0 \\ 1 & \text{if } d_i = 2, \end{cases}$$

seen as a base 2 expansion. Up to countable sets of exceptions, the values of y cover all $[0,1]$, and hence C is in bijection with $[0,1]$, and is not countable.

To check that $\lambda(C) = 0$, we observe that

$$C = \bigcap_n C_n \text{ where } C_n = \{x \in [0,1] \mid d_i(x) \neq 1 \text{ for all } i \leqslant n\}.$$

Each C_n is a Borel subset of $[0,1]$ – indeed, we have for instance

$$C_1 = [0,1/3] \cup [2/3,1], \quad C_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] ;$$

more generally, C_n is seen to be a disjoint union of 2^n intervals of length 3^{-n}. Thus, by additivity, we have

$$\lambda(C_n) = 2^n \times 3^{-n} = (2/3)^n,$$

and since $C_n \supset C_{n+1}$ and $\lambda(C_1) < +\infty$, we see by Proposition 1.2.3, (4) that

$$\lambda(C) = \lim_{n \to +\infty} \lambda(C_n) = 0.$$

One can phrase this result in a probabilistic way. Fix again an arbitrary base $b \geqslant 2$. Seeing $(X, \mathcal{B}, \lambda)$ as a probability space, the maps

$$X_i \quad \begin{cases} [0,1] \to \{0, \ldots, b-1\} \\ x \mapsto d_i(x) \end{cases}$$

are random variables (they are measurable because the inverse image under X_i of $d \in \{0, \ldots, b-1\}$ is the union of b^{i-1} intervals of length b^{-i}).

A crucial fact is the following:

Lemma 1.3.4. *The* (X_i) *are independent random variables on* $([0,1], \mathcal{B}, \lambda)$*, and the probability law of each is the same measure on* $\{0, \ldots, b-1\}$*, namely the normalized counting measure* P*:*

$$\lambda(X_i = d) = P(d) = \frac{1}{b}, \qquad \textit{for all digits } d.$$

It is, again, an instructive exercise to check this directly. Note that, knowing this, the computation of $\lambda(C_n)$ above can be rewritten as

$$\lambda(C_n) = \lambda(X_1 \neq 1, \ldots, X_n \neq 1) = \lambda(A_1 \cap \cdots \cap A_n)$$

where A_i is the event $\{X_i \neq 1\}$. By the definition of independence, this implies that

$$\lambda(A_1 \cap \cdots \cap A_n) = \prod_j \lambda(A_j) = \prod_j P(\{0,2\}) = (2/3)^n.$$

The probabilistic phrasing would be: "the probability that a uniformly, randomly chosen real number $x \in [0,1]$ has no digit 1 in its base 3 expansion is zero".

1.4. Borel measures and regularity properties

This short section gives a preliminary definition of some properties of measures for the Borel σ-algebra, which lead to a better intuitive understanding of these measures, and in particular of the Lebesgue measure.

Definition 1.4.1. (1) Let X be a topological space. A *Borel measure* on X is a measure μ for the Borel σ-algebra on X.

(2) A Borel measure μ on a topological space X is said to be *regular* if, for any Borel set $Y \subset X$, we have

$$\mu(Y) = \inf\{\mu(U) \mid U \text{ is an open set containing } Y\} \tag{1.12}$$

$$\mu(Y) = \sup\{\mu(K) \mid K \text{ is a compact subset contained in } Y\}. \tag{1.13}$$

This property, when it holds, gives a link between general Borel sets and the more "regular" open or compact sets.

We will show in particular the following:

Theorem 1.4.2. *The Lebesgue measure on* $\mathcal{B}_{\mathbf{R}}$ *is regular.*

Remark 1.4.3. If we take this result for granted, one may recover a way to define the Lebesgue measure: first of all, for any open subset $U \subset \mathbf{R}$, there exists a unique decomposition (in connected components)

$$U = \bigcup_{i \geqslant 1}]a_i, b_i[$$

in disjoint open intervals, and thus

$$\lambda(U) = \sum_i (b_i - a_i) \ ;$$

furthermore, for any Borel set $Y \subset \mathbf{R}$, we can then recover $\lambda(Y)$ using (1.12). In other words, for a Borel set, this gives the expression

$$\lambda(Y) = \inf\{m \geqslant 0 \mid \text{there exist disjoint intervals }]a_i, b_i[\\ \text{with } Y \subset \bigcup_i]a_i, b_i[\text{ and } \sum_i (b_i - a_i) = m\}.$$

This suggests one approach to prove Theorem 1.3.1: define the measure using the right-hand side of this expression, and check that this defines a measure on the Borel sets. This, of course, is far from being obvious (in particular because it is not clear at all how to use the assumption that the measure is restricted to Borel sets).

The following criterion, which we will also prove later, shows that regularity can be achieved under quite general conditions:

Theorem 1.4.4. *Let X be a locally compact topological space, in which any open set is a countable union of compact subsets. Then any Borel measure μ such that $\mu(K) < +\infty$ for all compact sets $K \subset X$ is regular; in particular, any finite measure, including probability measures, is regular.*

It is not difficult to check that the special case of the Lebesgue measure follows from this general result (if $K \subset \mathbf{R}$ is compact, it is also bounded, so we have $K \subset [-M, M]$ for some $M \in [0, +\infty[$, and hence $\lambda(K) \leqslant \lambda([-M, M]) = 2M$).

CHAPTER 2

Integration with respect to a measure

We now start the process of constructing the procedure of integration with respect to a measure, for a fixed measured space $(X, \mathcal{M}, \mu)$. The idea is to follow the idea of Lebesgue described in the introduction; the first step is the interpretation of the measure of a set $Y \in \mathcal{M}$ as the value of the integral of the characteristic function of Y, from which we can compute the integral of functions taking finitely many values (by linearity). One then uses a limiting process – which turns out to be quite simple – to define integrals of non-negative, and then general functions. In the last step, a restriction to *integrable* functions is necessary.

In this chapter, any function $X \to \mathbf{C}$ which is introduced is assumed to be measurable for the fixed σ-algebra $\mathcal{M}$, sometimes without explicit notice.

2.1. Integrating step functions

Definition 2.1.1 (Step function). A *step function* $s : X \to \mathbf{C}$ on a measured space $(X, \mathcal{M}, \mu)$ is a measurable function which takes only finitely many values.

In other words, a step function $s : X \to \mathbf{C}$ is a function that may be expressed as a finite sum

$$s = \sum_{i=1}^{n} \alpha_i \chi_{Y_i} \tag{2.1}$$

where the $\alpha_i \in \mathbf{C}$ are the distinct values taken by s and

$$Y_i = \{x \in X \mid s(x) = \alpha_i\} \in \mathcal{M}$$

are disjoint subsets of X. This expression is clearly unique, and $s \geqslant 0$ if and only if $\alpha_i \geqslant 0$ for all i.

The set of step functions on X, denoted $S(X)$, is stable under sums and products (in algebraic terms, it is a $\mathbf{C}$-algebra; the word "algebra" has here a completely different meaning than in σ-algebra). Indeed, this boils down to the two formulas

$$\chi_Y \chi_Z = \chi_{Y \cap Z} \text{ and } \chi_Y + \chi_Z = \chi_{Y \cap Z} + \chi_{Y \cup Z} = 2\chi_{Y \cap Z} + \chi_{(Y \cup Z) - (Y \cap Z)}. \tag{2.2}$$

We now define the integral of a step function, following Lebesgue's procedure.

Definition 2.1.2 (Integral of a step function). Let $(X, \mathcal{M}, \mu)$ be a measured space, $s : X \to \mathbf{R}$ a step function given by (2.1) such that s is non-negative. For any measurable set $Y \in \mathcal{M}$, we define

$$\int_Y s(x) d\mu(x) = \sum_{i=1}^{n} \alpha_i \mu(Y_i \cap Y) \in [0, +\infty],$$

which is called *the integral of s on Y with respect to μ.*

There are a number of different notation for the integral in the literature; for instance, one may find any of the following:

$$\int_Y sd\mu, \quad \int_Y s(x)d\mu, \quad \int_Y s(x)\mu, \quad \int_Y s(x)\mu(dx), \text{ or even } \int_Y s(x)dx,$$

when there is no ambiguity concerning the choice of the measure μ (this is the most common notation when the Lebesgue measure on $\mathbf{R}$ is involved).

Note that we can also write (see (2.2))

$$\int_Y s(x)d\mu(x) = \int_X s(x)\chi_Y(x)d\mu(x),$$

replacing the constraint of restricting the range of integration Y by a multiplication by its characteristic function.

REMARK 2.1.3. Although this is only the first step in the construction, it is interesting to note that it is already more general than the Riemann integral in the case where $X = [a, b]$ and μ is the Lebesgue measure: many step functions fail to be Riemann-integrable. A typical example (which goes back to Dirichlet, who pointed out that it was an example of a function which was not Riemann-integrable) is the following: take $X = [0, 1]$ and let f be the characteristic function of $\mathbf{Q} \cap [0, 1]$. Because any interval (of non-zero length) in any subdivision of $[0, 1]$ contains both rational numbers and irrational numbers, one can for all subdivisions construct Riemann sums for f with value 0 and with value 1. Thus there is no common limit for the Riemann sums.

However, since f is measurable and in fact is a step function for the Borel σ-algebra, we have by definition

$$\int_{[0,1]} f(x)d\mu(x) = \mu(\mathbf{Q} \cap [0, 1]) = 0$$

(by Example 1.3.3, (1)).

PROPOSITION 2.1.4. (1) *The map* $S(X)_{\geqslant 0} \to [0, +\infty]$

$$\Lambda : s \mapsto \int_X s(x)d\mu(x)$$

satisfies

$$\Lambda(\alpha s + \beta t) = \alpha\Lambda(s) + \beta\Lambda(t)$$

for any step functions s, t *both non-negative and for* α, $\beta \geqslant 0$. *Moreover, we have* $\Lambda(s) \geqslant 0$ *and* $\Lambda(s) = 0$ *if and only if* s *is zero almost everywhere (with respect to* μ*).*

(2) *Assume* $s \geqslant 0$. *The map*

$$\mu_s : Y \mapsto \int_Y s(x)d\mu(x)$$

is a measure on $\mathcal{M}$, *often denoted* $\mu_s = sd\mu$. *Moreover, for any step functions* s, $t \geqslant 0$, *we have*

$$\mu_{s+t} = \mu_s + \mu_t \text{ i.e. } \int_Y (s + t)d\mu = \int_Y sd\mu + \int_Y td\mu.$$

In addition, we have $\mu_s(Y) = 0$ *for any measurable set* Y *with* $\mu(Y) = 0$, *and any* μ*-negligible set is* μ_s*-negligible.*

These facts will be easily checked and are quite simple. We highlight them because they will reappear later on, in generalized form, after integration is defined for all non-negative functions.

Proof. (1) Since $S(X)$ is generated additively from functions χ_Y with $Y \in \mathcal{M}$, it is enough to show that

$$\int_X (\alpha\chi_Y + \beta\chi_Z) d\mu = \alpha \int_X \chi_Y d\mu + \beta \int_X \chi_Z d\mu,$$

or in other words, that

$$(\alpha + \beta)\mu(Y \cap Z) + \alpha\mu(Y - Z) + \beta\mu(Z - Y) = \alpha\mu(Y) + \beta\mu(Z).$$

This follows from the additivity of measure, since we have disjoint expressions

$$Y = (Y \cap Z) \cup (Y - Z), \text{ and } Z = (Y \cap Z) \cup (Z - Y).$$

It is clear that $\Lambda(s) \geqslant 0$ if $s \geqslant 0$, and that $\Lambda(s) = 0$ if s is zero almost everywhere. Conversely, assuming $\Lambda(s) = 0$, we obtain

$$0 = \sum_i \alpha_i \mu(Y_i) \geqslant \alpha_j \mu(Y_j)$$

for any fixed j, and therefore $\mu(Y_j) = 0$ for all j such that $\alpha_j > 0$. Consequenly, we obtain

$$\mu(\{x \mid s(x) \neq 0\}) = \sum_{\alpha_j > 0} \mu(Y_j) = 0.$$

(2) It is clear that the map μ_s takes non-negative values with $\mu_s(\emptyset) = 0$. Now let (Z_k), $k \geqslant 1$, be a sequence of pairwise disjoint measurable sets in X, and let Z denote the union of the Z_k. Using the countable additivity of μ, we obtain

$$\begin{aligned}\mu_s(Z) &= \sum_{i=1}^n \alpha_i \mu(Z \cap Y_i) = \sum_{i=1}^n \alpha_i \sum_{k \geqslant 1} \mu(Z_k \cap Y_i) \\ &= \sum_{k \geqslant 1} \sum_{i=1}^n \alpha_i \mu(Z_k \cap Y_i) = \sum_{k \geqslant 1} \mu_s(Z_k) \in [0, +\infty],\end{aligned}$$

where it is permissible to change the order of summation because the sum over i involves finitely many terms only.

The formula $\mu_{s+t}(Y) = \mu_s(Y) + \mu_t(Y)$ is simply the formula $\Lambda(s+t) = \Lambda(s) + \Lambda(t)$ proved in (1), applied to Y instead of X.

Finally, if $\mu(Y) = 0$, we have quite simply

$$\mu_s(Y) = \sum_i \alpha_i \mu(Y \cap Y_i) = 0,$$

which concludes the proof. □

2.2. Integration of non-negative functions

We now consider a non-negative measurable function $f : X \to [0, +\infty]$. We will define the integral of f with respect to μ by approximation with non-negative step functions. More precisely, we define:

Definition 2.2.1 (Integral of a non-negative function). Let

$$f : X \to [0, +\infty]$$

be measurable.[1] For any $Y \in \mathcal{M}$, the integral of f with respect to μ on Y is defined by

$$\int_Y f d\mu = \sup\Big\{\int_Y s d\mu \mid s \text{ is a step function such that } s \leqslant f\Big\} \in [0, +\infty]. \tag{2.3}$$

According to this definition, the integral of *any* measurable non-negative function is defined, provided the value $+\infty$ be permitted. This is analogous to the fact that series with non-negative terms converge in $[0, +\infty]$.

We now start by establishing the most elementary properties of this definition.

PROPOSITION 2.2.2. *We have:*

(0) *If $f \geqslant 0$ is a step function, the definition above gives the same as the previous definition for a step function.*

(1) *We have $\int f d\mu = 0$ if and only $f(x) = 0$ almost everywhere, and $\int f d\mu < +\infty$ implies that $f(x) < +\infty$ almost everywhere.*

(2) *If $\mu(Y) = 0$, then*

$$\int_Y f d\mu = 0 \text{ even if } f = +\infty \text{ on } Y.$$

(3) *We have*

$$\int_Y f d\mu = \int_X f(x)\chi_Y(x) d\mu.$$

(4) *If $0 \leqslant f \leqslant g$, then*

$$\int_X f d\mu \leqslant \int_X g d\mu. \tag{2.4}$$

(5) *If $Y \subset Z$, then*

$$\int_Y f d\mu \leqslant \int_Z f d\mu.$$

(6) *If $\alpha \in [0, +\infty[$, then*

$$\int_Y \alpha f d\mu = \alpha \int_Y f d\mu.$$

PROOF. (0): For any step function $s \leqslant f$, the difference $f - s$ is a step function, and is non-negative; we have

$$\int_Y f d\mu = \int_Y s d\mu + \int_Y (f - s) d\mu \geqslant \int_Y s d\mu,$$

and since we can in fact take $s = f$, we obtain the result.

(1): If f is zero outside of a negligible set Z, then any step function $s \leqslant f$ is zero outside Z, and therefore satisfies

$$\int_Y s(x) d\mu(x) = 0,$$

which implies that the integral of f is zero.

Conversely, assume f is not zero almost everywhere; we want to show that the integral of f is > 0. Consider the measurable sets

$$Y_k = \{x \mid f(x) \geqslant 1/k\}$$

[1] Recall from Lemma 1.1.9 that this is equivalent with $f^{-1}(]-\infty, a])$ being measurable for all real a.

where $k \geqslant 1$. Note that $Y_k \subset Y_{k+1}$ (this is an increasing sequence), and that

$$\bigcup_k Y_k = \{x \in X \mid f(x) > 0\}.$$

By Proposition 1.2.3), we have

$$0 < \mu(\{x \in X \mid f(x) > 0\}) = \lim_{k \to +\infty} \mu(Y_k),$$

and therefore there exists some k such that $\mu(Y_k) > 0$. Then the function

$$s = \frac{1}{k}\chi_{Y_k}$$

is a step function such that $s \leqslant f$, and such that

$$\int s d\mu = \frac{1}{k}\mu(Y_k) > 0,$$

which implies that the integral of f is > 0.

If $f(x) = +\infty$ for all $x \in Y$, where $\mu(Y) > 0$, the step functions

$$s_n = n\chi_Y$$

satisfy $0 \leqslant s_n \leqslant f$ and $\int s_n d\mu = n\mu(Y) \to +\infty$ hence, if $\int f d\mu < +\infty$, it must be that $\mu(\{x \mid f(x) = +\infty\}) = 0$.

Part (2) is clear, remembering that $0 \cdot +\infty = 0$.

For Part (3), notice that, for any step function $s \leqslant f$, we have

$$\int_Y s d\mu = \int_X s\chi_Y d\mu,$$

by definition, and moreover we can see that any step function $t \leqslant f\chi_Y$ is of this form $t = s\chi_Y$ where $s \leqslant f$ (it suffices to take $s = t\chi_Y$). Hence the result follows.

For Part (4), we just notice that $s \leqslant f$ implies $s \leqslant g$, and for (5), we just apply (3) and (4) to the inequality $0 \leqslant f\chi_Y \leqslant f\chi_Z$.

Finally, for (6), we observe first that the result holds when $\alpha = 0$, and for $\alpha > 0$, we have

$$s \leqslant f \text{ if and only if } \alpha s \leqslant \alpha f$$

which, together with

$$\int_Y (\alpha s) d\mu = \alpha \int_Y s d\mu$$

(Proposition 2.1.4, (1)) leads to the conclusion. □

We now come to the first really important result in the theory of integration: Beppo Levi's *monotone convergence theorem*. This shows that, for non-decreasing sequences of functions, *one can always exchange a limit and an integral.*

THEOREM 2.2.3 (Monotone convergence). *Let (f_n), $n \geqslant 1$, be a* non-decreasing *sequence of non-negative measurable functions. Define*

$$f(x) = \lim f_n(x) = \sup_n f_n(x) \in [0, +\infty]$$

for $x \in X$. Then $f \geqslant 0$ is measurable and we have

$$\int_X f d\mu = \lim_{n \to +\infty} \int_X f_n d\mu.$$

Proof. We have already explained in Lemma 1.1.12 that the limit function f is measurable.

To prove the formula for the integral, we combine inequalities in both directions. One is very easy: since

$$f_n \leqslant f_{n+1} \leqslant f$$

by assumption, Part (3) of the previous proposition shows that

$$\int_X f_n d\mu \leqslant \int_X f_{n+1} d\mu \leqslant \int_X f d\mu.$$

This means that the sequence of integrals $(\int f_n d\mu)$ is itself non-decreasing, and that its limit in $[0,+\infty]$ satisfies

$$\lim_{n\to+\infty} \int_X f_n d\mu \leqslant \int_X f d\mu. \tag{2.5}$$

It is of course the converse that is crucial. Consider a step function $s \leqslant f$. We must show that we have

$$\int s d\mu \leqslant \lim_{n\to+\infty} \int_X f_n d\mu.$$

This would be easy if we had $s \leqslant f_n$ for some n, but that is of course not necessarily the case. However, if we make s a little bit smaller, the increasing convergence $f_n(x) \to f(x)$ implies that something like this is true. More precisely, fix a parameter $\varepsilon \in]0,1]$ and consider the sets

$$X_n = \{x \in X \mid f_n(x) \geqslant (1-\varepsilon)s(x)\}.$$

for $n \geqslant 1$. We see that (X_n) is an increasing sequence of measurable sets (increasing because (f_n) is non-decreasing) whose union is equal to X, because $f_n(x) \to f(x)$ for all x (if $f(x) = 0$, we have $x \in X_1$, and otherwise $(1-\delta)s(x) < s(x) \leqslant f(x)$ shows that for some n, we have $f_n(x) > (1-\delta)s(x)$).

Using notation introduced in the previous section and the elementary properties above, we deduce that

$$\int_X f_n d\mu \geqslant \int_{X_n} f_n d\mu \geqslant (1-\varepsilon) \int_{X_n} s d\mu = (1-\varepsilon)\mu_s(X_n)$$

for all n.

We have seen that μ_s is a measure; since (X_n) is an increasing sequence, it follows that

$$\mu_s(X) = \lim_n \mu_s(X_n)$$

by Proposition 1.2.3, (3), and hence

$$(1-\varepsilon)\mu_s(X) = (1-\varepsilon)\lim_n \mu_s(X_n) \leqslant \lim_{n\to+\infty} \int_X f_n d\mu.$$

This holds for all ε, and since the right-hand side is independent of ε, we can let $\varepsilon \to 0$, and deduce

$$\mu_s(X) = \int_X s d\mu \leqslant \lim_{n\to+\infty} \int_X f_n d\mu.$$

This holds for all step functions $s \leqslant f$, and therefore implies the converse inequality to (2.5). Thus the monotone convergence theorem is proved. $\square$

Using this important theorem, we can describe a much more flexible way to approximate the integral of a non-negative function.

Proposition 2.2.4. *Let $f : X \to [0, +\infty]$ be a non-negative measurable function.*

(1) *There exists a sequence (s_n) of non-negative step functions, such that $s_n \leqslant s_{n+1}$ for $n \geqslant 1$, and moreover*

$$f(x) = \lim_{n \to +\infty} s_n(x) = \sup_n s_n(x)$$

for all $x \in X$.

(2) *For any such sequence (s_n), the integral of f can be recovered by*

$$\int_Y f d\mu = \lim_{n \to +\infty} \int_Y s_n d\mu$$

for any $Y \in \mathcal{M}$.

Proof. Part (2) is, in fact, merely a direct application of the monotone convergence theorem.

To prove part (1), we can construct a suitable sequence (s_n) explicitly.[2] For $n \geqslant 1$, we define

$$s_n(x) = \begin{cases} n & \text{if } f(x) \geqslant n \\ \frac{i-1}{2^n} & \text{if } \frac{i-1}{2^n} \leqslant f(x) < \frac{i}{2^n} \text{ where } 1 \leqslant i \leqslant n2^n. \end{cases}$$

The construction shows immediately that s_n is a non-negative step function, and that it is measurable (because f is):

$$f^{-1}([n, +\infty[) \in \mathcal{M} \text{ and } f^{-1}\Big(\Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[\Big) \in \mathcal{M}.$$

It is also immediate that $s_n \leqslant s_{n+1}$ for $n \geqslant 1$. Finally, to prove the convergence, we first see that if $f(x) = +\infty$, we will have

$$s_n(x) = n \to +\infty,$$

for $n \geqslant 1$, and otherwise, the inequality

$$0 \leqslant f(x) - s_n(x) \leqslant \frac{1}{2^n}$$

holds for $n > f(x)$, and implies $s_n(x) \to f(x)$. ☐

It might seem better to define the integral using the combination of these two facts. This is indeed possible, but the difficulty is to prove that the resulting definition is consistent; in other words, it is not clear at first that the limit in (2) is independent of the choice of (s_n) converging to f. However, now that the agreement of these two possible approaches is established, the following corollaries, which would be quite tricky to derive directly from the definition as a supremum over all $s \leqslant g$, follow quite simply.

Corollary 2.2.5. (1) *Let f and g be non-negative measurable functions on X. We then have*

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu. \tag{2.6}$$

(2) *Let (f_n), $n \geqslant 1$, be a sequence of non-negative measurable functions, and let*

$$g(x) = \sum_{n \geqslant 1} f_n(x) \in [0, +\infty].$$

[2] The simplicity of the construction is remarkable; note how it depends essentially on being able to use step functions where each level set can be an arbitrary measurable set.

Then $g \geqslant 0$ is measurable, and we have

$$\int_X g d\mu = \sum_{n \geqslant 1} \int_X f_n d\mu.$$

(3) *Let f be a non-negative measurable function. Define*

$$\mu_f(Y) = \int_Y f d\mu$$

for $Y \in \mathcal{M}$. Then μ_f is a measure on $(X, \mathcal{M})$, such that any μ-negligible set is μ_f-negligible. Moreover, we can write

$$\int_Y g d\mu_f = \int_Y g f d\mu \tag{2.7}$$

for any $Y \in \mathcal{M}$ and any measurable $g \geqslant 0$.

(4) *Let $\varphi : X \to X'$ be a measurable map. For any $g \geqslant 0$ measurable on X', and any $Y \in \mathcal{M}'$, we have the* change of variable formula

$$\int_{\varphi^{-1}(Y)} (g \circ \varphi) d\mu = \int_{\varphi^{-1}(Y)} g(\varphi(x)) d\mu(x) = \int_Y g d\varphi_*(\mu). \tag{2.8}$$

Proof. (1): let f and g be as stated. According to the first part of the previous proposition, we can find increasing sequences (s_n) and (t_n) of non-negative step functions on X such that

$$s_n(x) \to f(x), \qquad t_n(x) \to g(x).$$

Then, obviously, the increasing sequence $u_n = s_n + t_n$ converges pointwise to $f + g$. Since s_n and t_n are step functions (hence also u_n), we know that

$$\int_X (s_n + t_n) d\mu = \int_X s_n d\mu + \int_X t_n d\mu$$

for $n \geqslant 1$, by Proposition 2.1.4, (2). Applying the monotone convergence theorem three times by letting n go to infinity, we get

$$\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

(2): By induction, we can prove

$$\int_X \sum_{n=1}^{N} f_n(x) d\mu(x) = \sum_{n=1}^{N} \int_X f_n(x) d\mu(x)$$

for any $N \geqslant 1$ and any family of non-negative measurable functions $(f_n)_{n \leqslant N}$.

Now, since the terms in the series are $\geqslant 0$, the sequence of partial sums

$$g_n = \sum_{j=1}^{n} f_j(x)$$

is increasing and converges pointwise to $g(x)$. Applying once more the monotone convergence theorem, we obtain

$$\int_X g(x) d\mu(x) = \sum_{n \geqslant 1} \int_X f_n(x) d\mu(x),$$

as desired.

For part (3), we know already that the map μ_f takes non-negative values, and that $\mu_f(\emptyset) = 0$ (Proposition 1.2.3). There remains to check that μ_f is countable additive. But for any sequence (Y_n) of pairwise disjoint measurable sets, with union Y, we can write the formula

$$f(x)\chi_Y(x) = \sum_{n\geqslant 1} f(x)\chi_{Y_n}(x) \geqslant 0$$

for all $x \in X$ (note that since the Y_n are disjoint, at most one term in the sum is non-zero, for a given x). Hence the previous formula gives

$$\mu_f(Y) = \int_Y f d\mu = \int_X f(x)\chi_Y(x) d\mu(x) = \sum_{n\geqslant 1} \int_X f(x)\chi_{Y_n} d\mu(x) = \sum_{n\geqslant 1} \mu_f(Y_n),$$

which confirms that μ_f is a measure.

We also can see that if $\mu(Y) = 0$, we have $\mu_s(Y) = 0$ for any step function $s \leqslant f$, and hence taking the supremum, we derive $\mu_f(Y) = \sup \mu_s(Y) = 0$.

The formula (2.7) is valid, by definition, if g is a step function. Then, note that if (s_n) is an increasing sequence of step functions converging pointwise to g, the sequence given by $t_n = s_n f \chi_Y$ is also non-decreasing and converges pointwise to $fg\chi_Y$. Applying the monotone convergence theorem twice (for integration with respect to μ and to μ_f), we obtain

$$\int_Y g d\mu_f = \lim_{n\to+\infty} \int_Y s_n d\mu_f = \lim_{n\to+\infty} \int_Y s_n f d\mu = \int_Y g f d\mu.$$

Finally, we prove (4) by checking it for more and more general functions g. First of all, for $g = \chi_Z$, the characteristic function of a set $Z \in \mathcal{M}'$, we have

$$g \circ \varphi = \chi_{\varphi^{-1}(Z)}$$

(since the left-hand side takes values 0 and 1, and is equal to 1 if and only if $\varphi(x) \in Z$), and the formula becomes the definition

$$(\varphi_*(\mu))(Y \cap Z) = \mu(\varphi^{-1}(Y \cap Z)).$$

Next, observe that if (2.8) holds for two functions, it holds for their sum, by additivity of the integral on both sides. This means that the formula also holds for all step functions. And then, finally, if $g \geqslant 0$ is the pointwise non-decreasing limit of a sequence (s_n) of step functions, we have

$$(s_n \circ \varphi)(x) = s_n(\varphi(x)) \leqslant s_{n+1}(\varphi(x)) \to (g \circ \varphi)(x),$$

and

$$(s_n \circ \varphi)(x)\chi_Y(x) \to (g \circ \varphi)(x)\chi_Y(x).$$

Consequently, the monotone convergence theorem shows that

$$\int_Y s_n(\varphi(x)) d\mu(x) \to \int_Y g(\varphi(x)) d\mu(x),$$

which is what we wanted to show. □

Example 2.2.6. The change of variable formula is sometimes quite useful to replace the problem of proving results for integrals of complicated functions on a complicated space X to those of simple functions on a simple space, but with respect to a possibly complicated measure... For instance, for any measurable $g : X \to \mathbf{R}$, denote by $\nu = g_*(\mu)$ the image of μ, which is a Borel measure on $\mathbf{R}$. We have by (2.8), applied to $|g|$, that

$$\int_X |g(x)| d\mu(x) = \int_{\mathbf{R}} |x| d\nu(x).$$

2.3. Integrable functions

Finally, we can define what are the functions which are integrable with respect to a measure. Although there are other approaches, we consider the simplest, which uses the order structure on the set of real numbers for real-valued functions, and linearity for complex-valued ones.

For this, we recall the decomposition

$$f = f^+ - f^-, \qquad f^+(x) = \max(0, f(x)), \quad f^-(x) = \max(0, -f(x)),$$

and we observe that $|f| = f^+ + f^-$.

Definition 2.3.1. Let $(X, \mathcal{M}, \mu)$ be a measured space.

(1) A measurable function $f : X \to \mathbf{R}$ is said to be *integrable on* Y *with respect to* μ, for $Y \in \mathcal{M}$, if the non-negative function $|f| = f^+ + f^-$ satisfies

$$\int_Y |f| d\mu < +\infty,$$

and its integral on Y is defined by

$$\int_Y f d\mu = \int_Y f^+ d\mu - \int_Y f^- d\mu \in \mathbf{R}.$$

(2) A measurable function $f : X \to \mathbf{C}$ is said to be integrable on $Y \in \mathcal{M}$ if $|f| = \sqrt{\mathrm{Re}(f)^2 + \mathrm{Im}(f)^2} \geqslant 0$ satisfies

$$\int_Y |f| d\mu < +\infty,$$

and its integral on Y is defined by

$$\int_Y f d\mu = \int_Y \mathrm{Re}(f) d\mu + i \int_Y \mathrm{Im}(f) d\mu \in \mathbf{C},$$

so that, by definition, we have

$$\mathrm{Re}\Bigl(\int_Y f d\mu\Bigr) = \int_Y \mathrm{Re}(f) d\mu \text{ and } \mathrm{Im}\Bigl(\int_Y f d\mu\Bigr) = \int_Y \mathrm{Im}(f) d\mu. \tag{2.9}$$

(3) We denote by $L^1(X, \mu)$, or sometimes simply $L^1(\mu)$, the set of all μ-integrable complex-valued functions defined on X.

The first thing to notice here is that this definition, which is extremely general (without assumption on the structure of X or boundedness of f, etc), has the nature of *"absolutely convergent"* integral; this may seem like a restriction, but it is essential for the resulting process to behave reasonably. In particular, note that if f is μ-integrable on X, then it follows that f is μ-integrable on Y for any measurable subset $Y \subset X$, because

$$\int_Y |f| d\mu = \int_X |f| \chi_Y d\mu,$$

and $|f|\chi_Y \leqslant |f|$. This innocuous property would *fail* for most definitions of a non-absolutely convergent integral.

We first observe that, as claimed, these definitions lead to well-defined real (or complex) numbers for the values of the integrals. For instance, since we have inequalities

$$0 \leqslant f^\pm \leqslant |f|,$$

we can see by monotony (2.4) that

$$\int_X f^{\pm} d\mu \leqslant \int_X |f| d\mu < +\infty,$$

if f is μ-integrable. Similarly, if f is complex-valued, we have

$$|\operatorname{Re}(f)| \leqslant |f|, \text{ and } |\operatorname{Im}(f)| \leqslant |f|,$$

which implies that the real and imaginary parts of f are themselves μ-integrable.

One may also remark immediately that

$$\int_X \bar{f} d\mu = \overline{\int_X f d\mu} \tag{2.10}$$

if f (equivalently, $\bar{f}$) is integrable.

Remark 2.3.2. Let (Ω, Σ, P) be a probability space, and let X be a complex-valued random variable defined on Ω. The integral of X on Ω is then customarily called the *expectation of* X, and is denoted $E(X)$.

Note the following important formula: denoting by $\mu = X(P)$ the measure on $\mathbf{C}$ which is the probability law of X (see Remark 1.2.9), we have (see 2.8)

$$E(|X|) = \int_{\mathbf{C}} |x| f d\mu, \qquad E(X) = \int_{\mathbf{C}} x d\mu$$

As we did in the previous section, we collect immediately a few simple facts before giving some interesting examples.

Proposition 2.3.3. (1) *The set* $L^1(X, \mu)$ *is a* $\mathbf{C}$*-vector space. Moreover, consider the map*

$$\begin{cases} L^1(\mu) \to [0, +\infty[\\ f \mapsto \|f\|_1 = \int_X |f| d\mu. \end{cases}$$

This map is a semi-norm *on* $L^1(X, \mu)$*, i.e., we have*

$$\|af\|_1 = |a| \|f\|_1$$

for $a \in \mathbf{C}$ *and* $f \in L^1(X, \mu)$*, and*

$$\|f + g\|_1 \leqslant \|f\|_1 + \|g\|_1,$$

for f, $g \in L^1(X, \mu)$*. Moreover,* $\|f\|_1 = 0$ *if and only if* f *is zero* μ*-almost everywhere.*

(2) *The map*

$$\begin{cases} L^1(\mu) \to \mathbf{C} \\ f \mapsto \int_X f d\mu \end{cases}$$

is a linear map, it is non-negative in the sense that $f \geqslant 0$ *implies* $\int f d\mu \geqslant 0$*, and it satisfies*

$$\Bigl| \int_X f d\mu \Bigr| \leqslant \int_X |f| d\mu = \|f\|_1. \tag{2.11}$$

(3) *For any measurable map* $\varphi : (X, \mathcal{M}) \to (X', \mathcal{M}')$*, and any measurable function* g *on* X'*, we have*

$$\int_{\varphi^{-1}(Y)} g(\varphi(x)) d\mu(x) = \int_Y g(y) d\varphi_*(\mu)(y), \tag{2.12}$$

in the sense that *if either of the two integrals written are defined, i.e., the corresponding function is integrable with respect to the relevant measure, then the other is also integrable, and their respective integrals are equal.*

In concrete terms, note that (2) means that

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu$$

for f, $g \in L^1(\mu)$ and α, $\beta \in \mathbf{C}$; the first part ensures that the left-hand side is defined (i.e., $\alpha f + \beta g$ is itself integrable).

The interpretation of (3) that we spelled out is very common for identities involving various integrals, and sometimes we will not explicitly mention the implied statement that both sides of such a formula are either simultaneously integrable or non-integrable (they both make sense at the same time), and that they coincide in the first case.

Proof. Notice first that

$$|\alpha f + \beta g| \leqslant |\alpha||f| + |\beta||g|$$

immediately shows (using the additivity and monotony of integrals of non-negative functions) that

$$\int_X |\alpha f + \beta g| d\mu \leqslant \int_X |\alpha||f| d\mu + \int_X |\beta||g| d\mu < +\infty$$

for any f, $g \in L^1(\mu)$ and α, $\beta \in \mathbf{C}$. This proves both $\alpha f + \beta g \in L^1(\mu)$ and the triangle inequality (take $\alpha = \beta = 1$). Moreover, if $g = 0$, we have an identity $|\alpha f| = |\alpha||f|$, and hence also $\|\alpha f\|_1 = |\alpha|\|f\|_1$ by Proposition 2.2.2, (6).

To conclude the proof of (1), we note that those f for which $\|f\|_1 = 0$ are such that $|f|$ is almost everywhere zero, by Proposition 2.2.2, (1), which is the same as saying that f itself is almost everywhere zero.

We now look at (2), and first show that the integral is linear with respect to f. This requires some care because the operations that send f to its positive and negative parts $f \mapsto f^{\pm}$ are not linear themselves.

We prove the following separately:

$$\int \alpha f d\mu = \alpha \int f d\mu, \tag{2.13}$$

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu \tag{2.14}$$

where $\alpha \in \mathbf{C}$ and f, $g \in L^1(\mu)$. Of course, by (1), all these integrals make sense.

For the first identity, we start with $\alpha \in \mathbf{R}$ and a real-valued function f. If $\varepsilon \in \{-1, 1\}$ is the sign of α (with sign 1 for $\alpha = 0$), we see that

$$(\alpha f)^+ = (\varepsilon\alpha) f^{\varepsilon +}, \text{ and } (\alpha f)^- = (\varepsilon\alpha) f^{\varepsilon -},$$

where the notation on the right-hand sides should have obvious meaning. Then (2.13) follows in this case: by definition, we get

$$\int_X \alpha f d\mu = \int_X (\varepsilon\alpha) f^{\varepsilon +} d\mu - \int_X (\varepsilon\alpha) f^{\varepsilon -} d\mu,$$

hence, by Proposition 2.2.2, (6) this gives

$$\int_X \alpha f d\mu = \varepsilon\alpha \int_X f^{\varepsilon +} d\mu - \varepsilon\alpha \int_X f^{\varepsilon -} d\mu = \varepsilon\alpha \int_X (f^{\varepsilon +} - f^{\varepsilon -}) d\mu = \alpha \int_X f d\mu,$$

as claimed.

Next (still in the real case) we come to (2.14). Let $h = f + g$, so that

$$h = h^+ - h^- = f^+ - f^- + g^+ - g^-,$$

which means also

$$h^+ + f^- + g^- = f^+ + g^+ + h^-,$$

where each side is now a sum of non-negative functions, each with finite integral. By the additivity which we have proved for this situation (see (2.6)), we get

$$\int h^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int f^+ d\mu + \int g^+ d\mu + \int h^- d\mu,$$

hence, as desired, we get

$$\int f d\mu + \int g d\mu = \int h d\mu.$$

We leave to the reader to check the case of $\alpha \in \mathbf{C}$ and f taking complex values; this involves similar manipulations.

We now come to checking (2.11). This is obvious when f is real-valued because $|f| = f^+ + f^-$ in that case. For f taking complex values, we require a small trick to avoid getting entangled in the square root defining $|f|$. Let $\theta \in \mathbf{R}$ be a real number such that

$$\int_X f d\mu = e^{i\theta} \Big| \int_X f d\mu \Big| \in \mathbf{C}.$$

Using (2.13) and (2.9), we obtain

$$\Big| \int_X f d\mu \Big| = \int_X e^{-i\theta} f d\mu = \mathrm{Re}\Big(\int_X e^{-i\theta} f d\mu \Big) = \int_X \mathrm{Re}(e^{-i\theta} f) d\mu.$$

Now, since $\mathrm{Re}(z) \leqslant |z|$, this last integral is bounded easily by

$$\int_X \mathrm{Re}(e^{-i\theta} f) d\mu \leqslant \int_X |f| d\mu$$

using the monotony (2.4) (or positivity and $|f| - \mathrm{Re}(e^{-i\theta} f) \geqslant 0$.

As to the change of variable identity (2.12), it generalizes (2.8), and is an immediate consequence by linearity. □

These detailed checks are of course rather boring, but necessary to ensure that the integral has the properties we expect. However, having verified that this is the case, it will very rarely be necessary to use the definition of integrals using either (2.3) or the decomposition $f = f^+ - f^-$. The usual linearity property and the monotone convergence theorem will be used instead.

Example 2.3.4. We start with the simplest examples of measures in Example 1.2.7, keeping the important case of the Lebesgue measure to a separate section.

(1) Let X be an arbitrary set and μ the counting measure on X. Of course, we expect that

$$\int_X f(x) d\mu = \sum_{x \in X} f(x),$$

for a non-negative function f on X, but since there is no assumption on an ordering of the summation set, or that it be countable, we see that even this simple case offers something new.

One can then check that a function is integrable with respect to this measure if it is "absolutely summable", in the following sense: given a family $(f(x))_{x \in X}$, it is said to be *absolutely summable with sum* $S \in \mathbf{C}$ if and only if, for any $\varepsilon > 0$, there exists a finite subset $X_0 \subset X$ for which

$$\Big| \sum_{x \in X_1} f(x) - S \Big| < \varepsilon,$$

for any finite subset X_1 with $X_0 \subset X_1 \subset X$.

In particular, if $X = \mathbf{N}$ with the counting measure, we are considering series

$$\sum_{n \geqslant 1} a_n,$$

of complex numbers, and a sequence (a_n) is *integrable* if and only if the series *converges absolutely*. In particular, $a_n = (-1)^n/(n+1)$ does *not* define an integrable function.

Corollary 2.2.5 then shows that, provided each $a_{i,j}$ is $\geqslant 0$, we can exchange two series:

$$\sum_{i \geqslant 1} \sum_{j \geqslant 1} a_{i,j} = \sum_{j \geqslant 1} \sum_{i \geqslant 1} a_{i,j}.$$

This is not an obvious fact, and it is quite nice to recover this as part of the general theory.

(2) Let $\mu = \delta_{x_0}$ be the Dirac measure at $x_0 \in X$; then any function $f : X \to \mathbf{C}$ is μ-integrable and we have

$$\int_X f(x) d\delta_{x_0}(x) = f(x_0). \tag{2.15}$$

More generally, let $x_1, \ldots, x_n \in X$ be finitely many points in X; one can construct the probability measure

$$\delta = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} \delta_{x_i}$$

such that

$$\int_X f(x) d\delta(x) = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} f(x_i),$$

which is some kind of “sample sum” which is very useful in applications (since any “integral” which is really numerically computed is in fact a finite sum of a similar type).

Exercise 2.3.5. Let (Ω, Σ, P) be a probability space. We will show that if X and Y are *independent* complex-valued random variables (Definition 1.2.10), we have $XY \in L^1$ and

$$E(XY) = E(X)E(Y), \qquad \text{i.e.} \qquad \int_\Omega XY dP = \int_\Omega X dP \times \int_\Omega Y dP. \tag{2.16}$$

Note that, in general, it is certainly *not true* that the product of two integrable function is integrable (and even if that is the case, the integral of the product is certainly not usually given by the product of the integrals of the factors)! This formula depends essentially on the assumption of independence of X and Y.

(1) Show that if X and Y are non-negative step functions, the formula (2.16) holds, using the definition of independent random variables.

(2) Let X and Y be non-negative random variables, and let S_n and T_n be the step functions constructed in the proof of Proposition 2.2.4 to approximate X and Y. Show that S_n and T_n are independent for any fixed $n \geqslant 1$. (Warning: we only claim this for a given n, not the independence of two sequences).

(3) Deduce from this that (2.16) holds for non-negative random variables, using the monotone convergence theorem.

(4) Deduce the general case from this.

In a later chapter, we will see that product measures give a much quicker approach to this important property.

Our next task is to prove the most useful convergence theorem in the theory, *Lebesgue's dominated convergence theorem.* Recall from the introduction, specifically (0.3), that one can not hope to have

$$\int_X \Big(\lim_{n\to+\infty} f_n(x)\Big) d\mu(x) = \lim_{n\to+\infty} \int_X f_n(x) d\mu(x),$$

whenever $f_n(x) \to f(x)$ pointwise. Lebesgue's theorem adds a single, very flexible, condition, that ensures the validity of this conclusion.

THEOREM 2.3.6 (Dominated convergence theorem). *Let $(X, \mathcal{M}, \mu)$ be a measured space, (f_n) a sequence of complex-valued μ-integrable functions. Assume that, for all $x \in X$, we have*

$$f_n(x) \to f(x) \in \mathbf{C}$$

as $n \to +\infty$, so f is a complex-valued function on X.

Then f is measurable; moreover, if there exists $g \in L^1(\mu)$ such that

$$|f_n(x)| \leqslant g(x) \text{ for all } n \geqslant 1 \text{ and all } x \in X, \tag{2.17}$$

then the limit function f is μ-integrable, and it satisfies

$$\int_X f(x) d\mu(x) = \lim_{n\to+\infty} \int_X f_n(x) d\mu(x). \tag{2.18}$$

In addition, we have in fact

$$\int_X |f_n - f| d\mu \to 0 \text{ as } n \to +\infty, \tag{2.19}$$

or in other words, f_n converges to f in $L^1(\mu)$ for the distance given by $d(f, g) = \|f - g\|_1$.

REMARK 2.3.7. The reader will have no difficulty checking that the sequence given by (0.3) does not satisfy the *domination condition* (2.17) (where the measure is the Lebesgue measure).

This additional condition is not *necessary* to ensure that (2.18) holds (when $f_n \to f$ pointwise); however, experience shows that it is highly flexible: it is very often satisfied, and very often quite easy to check.

The idea of the proof is to *first* prove (2.19), which is enough because of the general inequality

$$\Big|\int_X f_n(x) d\mu(x) - \int_X f(x) d\mu(x)\Big| \leqslant \int_X |f_n - f| d\mu \ ;$$

the immediate advantage is that this reduces to a problem concerning integrals of non-negative functions. One needs another trick to bring up some monotonic sequence to which the monotone convergence theorem will be applicable. First we have another interesting result:

LEMMA 2.3.8 (Fatou's lemma). *Let (f_n) be a sequence of non-negative measurable functions $f_n : X \to [0, +\infty]$. We then have*

$$\int_X (\liminf_{n\to+\infty} f_n) d\mu \leqslant \liminf_{n\to+\infty} \int_X f_n d\mu,$$

and in particular, if $f_n(x) \to f(x)$ for all x, we have

$$\int_X f(x)d\mu(x) \leqslant \liminf_{n\to+\infty} \int_X f_n(x)d\mu(x).$$

PROOF. By definition, a liminf is a monotone limit:

$$\liminf_{n\to+\infty} f_n = \lim_{n\to+\infty} g_n,$$

where

$$g_n(x) = \inf_{k\geqslant n} f_k(x),$$

so that $0 \leqslant g_n \leqslant g_{n+1}$. The functions g_n are also measurable, and according to the monotone convergence theorem, we have

$$\int_X (\liminf_{n\to+\infty} f_n)d\mu = \lim_{n\to+\infty} \int_X g_n d\mu.$$

But since, in addition, we have $g_n \leqslant f_n$, it follows that

$$\int g_n d\mu \leqslant \int f_n d\mu$$

for all n, hence

$$\int_X (\liminf_{n\to+\infty} f_n)d\mu \leqslant \int_X f_n(x)d\mu(x).$$

Although the right-hand side may not converge, we can pass to the liminf and obtain the stated inequality. ☐

PROOF OF THE DOMINATED CONVERGENCE THEOREM. The first step is to check that f is integrable; this would be quite delicate in general, but the domination condition provides this with little difficulty: from

$$|f_n| \leqslant g$$

for all n, going to the pointwise limit, we get $|f| \leqslant g$, and since g is μ-integrable, the monotony of integral implies that f is also μ-integrable.

Now we are going to show that

$$\limsup_{n\to+\infty} \int_X |f_n - f|d\mu(x) = 0, \tag{2.20}$$

which implies (2.19) since the sequences involved are all $\geqslant 0$; as already observed, this also gives the exchange of limit and integral (2.18).

Fatou's lemma does not deal with limsups, but we can easily exchange to a liminf by considering $-|f_n - f|$; however, this is not $\geqslant 0$, so we shift by adding a sufficiently large function. So let $h_n = 2g - |f_n - f|$ for $n \geqslant 1$; we have

$$h_n \geqslant 0, \qquad h_n(x) \to 2g(x),$$

for all x. By Fatou's Lemma, it follows that

$$2\int_X g d\mu = \int_X (\lim_{n\to+\infty} h_n)d\mu \leqslant \liminf_{n\to+\infty} \int_X h_n d\mu. \tag{2.21}$$

By linearity, we compute the right-hand side

$$\int_X h_n d\mu = 2\int_X g d\mu - \int_X |f_n - f|d\mu,$$

and therefore

$$\liminf_{n\to+\infty} \int_X h_n d\mu = 2\int_X g d\mu - \limsup_{n\to+\infty} \int_X |f_n - f| d\mu.$$

By comparing with (2.21), we get

$$\limsup_{n\to+\infty} \int_X |f_n - f| d\mu \leqslant 0,$$

which is (2.20). ☐

EXAMPLE 2.3.9. Here is a first general example: assume that μ is a finite measure (for instance, a probability measure). Then the constant functions are in $L^1(\mu)$, and the domination condition holds, for instance, for any sequence of functions (f_n) which are uniformly bounded (over n) on X.

EXAMPLE 2.3.10. Here is another simple example:

LEMMA 2.3.11. *Let $X_n \in \mathcal{M}$ be an increasing sequence of measurable sets such that*

$$X = \bigcup_{n\geqslant 1} X_n,$$

and let $Y_n = X - X_n$ be the sequence of complementary sets.

For any $f \in L^1(\mu)$, we have

$$\lim_{n\to+\infty} \int_{X_n} f(x)d\mu(x) = \int_X f(x)d\mu(x),$$

$$\lim_{n\to+\infty} \int_{Y_n} f(x)d\mu(x) = 0.$$

PROOF. We apply the dominated convergence theorem to the sequence

$$f_n(x) = f(x)\chi_{X_n}(x).$$

Since $X_n \subset X_{n+1}$, it follows that, for any fixed $x \in X$, we have $f_n(x) = f(x)$ for all n large enough (but depending on x!). Thus $f_n(x) \to f(x)$ pointwise. Moreover, the domination condition is very easy to achieve here: we have

$$|f_n(x)| = |f(x)|\chi_{X_n}(x) \leqslant |f(x)|$$

for all x and all n. Since, by assumption, $g = |f|$ is μ-integrable, this gives condition (2.17). Consequently, we obtain

$$\int_{X_n} f(x)d\mu(x) = \int_X f_n(x)d\mu(x) \to \int_X f(x)dx.$$

For the second statement, we just need to remark that

$$\int_{X_n} f(x)d\mu(x) + \int_{Y_n} f(x)d\mu(x) = \int_X f(x)d\mu(x)$$

for $n \geqslant 1$. ☐

For instance, consider $f \in L^1(X,\mu)$, and define

$$X_n = \{x \in X \mid |f(x)| \leqslant n\}$$

for $n \geqslant 1$. We have $X_n \subset X_{n+1}$, of course, and

$$\bigcup_{n\geqslant 1} X_n = X,$$

(since f only takes complex values). We therefore conclude that

$$\int_X f(x)d\mu(x) = \lim_{n\to+\infty} \int_{X_n} f(x)d\mu(x) = \lim_{n\to+\infty} \int_{\{|f|\leqslant n\}} f(x)d\mu(x).$$

This is often quite useful because it shows that one may often prove general properties of integrals by restricting first to situations where the function that is integrated is bounded.

2.4. Integrating with respect to the Lebesgue measure

In this section, we consider the special case of the measure space $(\mathbf{R}, \mathcal{B}, \lambda)$, where λ is the Lebesgue measure, or its subsets $([a,b], \mathcal{B}, \lambda)$. We will see basic examples of λ-integrable functions (and of functions which are not), and will explain the relation between this integral and the Riemann integral in the case of sufficiently smooth functions.

Example 2.4.1 (Integrable functions). In order to prove that a given measurable function f is integrable (with respect to the Lebesgue measure on a subset $X \subset \mathbf{R}$, or for a more general measure space), the most common technique is to find a “simple” comparison function g which is known to be integrable and for which it is known that

$$|f(x)| \leqslant g(x), \qquad x \in X.$$

Frequently, one uses more than one comparison function: for instance, one finds disjoint subsets X_1, X_2 such that $X = X_1 \cup X_2$, and functions g_1, g_2 integrable on X_1 and X_2 respectively, such that

$$|f(x)| \leqslant \begin{cases} g_1(x) & \text{if } x \in X_1, \\ g_2(x) & \text{if } x \in X_2. \end{cases}$$

With some care, this can be applied with infinitely many subsets. For instance, consider

$$X = [1, +\infty[, \qquad f(x) = x^{-\nu}$$

where $\nu \geqslant 0$. Then f is λ-integrable on X if (in fact, only if) $\nu > 1$. Indeed, note that

$$0 \leqslant f(x) \leqslant n^{-\nu}, \quad x \in [n, n+1[, \quad n \geqslant 1,$$

and therefore, using the monotone convergence theorem, we get

$$\int_X x^{-\nu} d\lambda(x) \leqslant \sum_{n\geqslant 1} n^{-\nu} < +\infty$$

if $\nu > 1$ (the converse is proved in the next example).

Note that the range of integration here, the unbounded interval $[0, +\infty[$, is treated just like any other; there is no distinction in principle between bounded and unbounded intervals in Lebesgue's theory, like there was in Riemann's integral. However, using Lemma 2.3.11, we *recover* the fact that

$$\int_{[0,+\infty[} f(x)d\lambda(x) = \lim_{n\to+\infty} \int_{[0,n]} f(x)d\lambda(x)$$

for any λ-integrable function on X.

Similarly, Lebesgue's integral deals uniformly with unbounded functions (either on bounded or unbounded intervals). Consider $X = [0,1]$ and

$$f(x) = \begin{cases} x^{-\nu} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

for some fixed $\nu \geqslant 0$.

We now note that

$$0 \leqslant f(x) \leqslant \Big(\frac{1}{n+1}\Big)^{-\nu} = (n+1)^{\nu}, \quad x \in](n+1)^{-1}, n^{-1}], \quad n \geqslant 1,$$

and summing over $n \geqslant 1$, keeping in mind the size

$$\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} \leqslant \frac{2}{(n+1)^2}$$

of each interval in the subdivision, we obtain

$$\int_{[0,1]} x^{-\nu} d\lambda(x) \leqslant 2 \sum_{n\geqslant 1} (n+1)^{\nu-2}$$

which is finite for $\nu - 2 < -1$, i.e., for $\nu < 1$.

One can also deal easily with functions with singularities located at arbitrary points of $\mathbf{R}$, without requiring to place them at the extremities of various intervals.

Another very common comparison function, that may be applied to any set with finite measure, is a constant function: for $X = [a, b]$ for instance, any measurable bounded function is λ-integrable on X, because the constant function 1 has integral $\leqslant (b - a)$.

Example 2.4.2 (Non-integrable functions). Except for the requirement of measurability, which is very rarely an issue in practice, the condition of integrability of a function f on a measurable subset $X \subset \mathbf{R}$ with respect to λ, is analogue to the condition of *absolute convergence* for series

$$\sum_{n\geqslant 1} a_n, \qquad a_n \in \mathbf{C}.$$

As such, corresponding to the existence of series like

$$\sum_{n\geqslant 1} \frac{(-1)^{n+1}}{n}$$

which are convergent but not absolutely so, there are many examples of functions f which are not-integrable although some specific approximating limits may exist. For instance, consider

$$f(x) = \begin{cases} \frac{\sin(x)}{x} & \text{if } x > 0, \\ 1 & \text{if } x = 0, \end{cases}$$

defined for $x \in [0, +\infty[$. It is well-known that

$$\lim_{N\to+\infty} \int_{[0,N]} f(x) d\lambda(x) = \frac{\pi}{2}, \tag{2.22}$$

which means that the integral of f on $[0, +\infty[$ exists in the sense of "improper" Riemann integrals (see also the next example; f is obviously integrable on each interval $[0, N]$ because it is bounded by 1 there). However, f is *not* in $L^1(\mathbf{R}_{\geqslant 0}, \lambda)$. Indeed, note that

$$|f(x)| \geqslant \frac{\sqrt{3}}{2x} \geqslant \frac{\sqrt{3}}{2(n+1)\pi}$$

whenever $x \in [\frac{\pi}{2} + n\pi - \frac{\pi}{3}, \frac{\pi}{2} + n\pi + \frac{\pi}{3}]$, $n \geqslant 1$, so that

$$\int_{[0,+\infty[} |f(x)| d\lambda(x) \geqslant \frac{\sqrt{3}}{2\pi} \sum_{n\geqslant 1} \frac{1}{n+1} = +\infty,$$

which proves the result.

In fact, note that if we restrict f to the set Y which is the union over $n \geqslant 1$ of the intervals

$$I_n = [\tfrac{\pi}{2} + 2n\pi - \tfrac{\pi}{3}, \tfrac{\pi}{2} + 2n\pi + \tfrac{\pi}{3}],$$

we have

$$f(x) \geqslant \frac{\sqrt{3}}{2x} \geqslant \frac{\sqrt{3}}{2(2n+1)\pi}$$

on I_n (without sign changes), hence also

$$\int_Y f(x) d\lambda(x) = +\infty,$$

and this should be valid with whatever definition of integral is used. So if one managed to change the definition of integral further so that f becomes integrable on $[0, +\infty[$ with integral $\pi/2$, we would have to draw the undesirable conclusion that the restriction of an integrable function to a subset of its domain of definition may fail to be integrable. Avoiding this is one of the main reasons for using a definition related to absolute convergence. This is not to say that formulas like (2.22) have no place in analysis: simply, they have to be studied separately, and should not be thought of as being truly analogous to a statement of integrability.

Example 2.4.3 (Comparison with the Riemann integral). We have already seen examples of λ-integrable step-functions which are not Riemann-integrable. However, we now want to discuss the relation between the two notions in the case of sufficiently regular functions. For clarity, we denote here

$$\int_a^b f(x) dx$$

the Riemann integral, and use

$$\int_{[a,b]} f(x) d\lambda(x)$$

for the Lebesgue integral.

<u>1st case</u>: Let $I = [a, b]$ be a compact interval, and let

$$f : I \to \mathbf{R}$$

be a measurable function which is Riemann-integrable, for instance a continuous function. Then f is λ-integrable, and we have the equality

$$\int_a^b f(x) dx = \int_I f(x) d\lambda(x). \tag{2.23}$$

This already allows us to compute many Lebesgue integrals by using known Riemann-integral identities.

The proof of this fact is very simple: for any subdivision

$$a = y_0 < y_1 < \cdots < y_n = b$$

of the interval of integration, let

$$S_+(f) = \sum_{i=1}^{n-1} (y_i - y_{i-1}) \sup_{y_{i-1} \leqslant x \leqslant y_i} f(x),$$

$$S_-(f) = \sum_{i=1}^{n-1} (y_i - y_{i-1}) \min_{y_{i-1} \leqslant x \leqslant y_i} f(x)$$

be the upper and lower Riemann sums for the integral of f. We can immediately see that

$$S_-(f) \leqslant \sum_{i=1}^{n-1} \int_{[y_{i-1}, y_i]} f(x) d\lambda(x) = \int_{[a,b]} f(x) d\lambda(x) \leqslant S_+(f),$$

and hence (2.23) is an immediate consequence of the definition of the Riemann integral. This applies, in particular, to any function f which is either continuous or piecewise continuous.

One can also study general Riemann-integrable functions, although this requires a bit of care with measurability. The following result indicates clearly the restriction that Riemann's condition imposes:

THEOREM 2.4.4. *Let $f : [a, b] \to \mathbf{C}$ be any function. Then f is Riemann-integrable if and only if f is bounded, and the set of points where f is not continuous is negligible with respect to the Lebesgue measure.*

(The points concerned are those x where any of the four limits

$$\liminf_{y \to x} \mathrm{Re}(f(x)), \quad \limsup_{y \to x} \mathrm{Re}(f(x)), \quad \liminf_{y \to x} \mathrm{Im}(f(x)), \quad \limsup_{y \to x} \mathrm{Im}(f(x)),$$

is not equal to $\mathrm{Re}(f(x))$ or $\mathrm{Im}(f(x))$, respectively). We do not prove this, since this is not particularly enlightening or useful for the rest of the book.

<u>2nd case</u>: Let $I = [a, +\infty[$ and let $f : I \to \mathbf{C}$ be such that the Riemann integral

$$\int_a^{+\infty} f(x) dx$$

converges *absolutely*. Then $f \in L^1(I, \lambda)$ and

$$\int_a^{+\infty} f(x) dx = \int_I f(x) d\lambda(x)$$

(we have already seen in Example 2.4.2 that the restriction to absolutely convergent integrals is necessary).

Indeed, as we have already remarked in Example 2.4.1, we let

$$X_n = [a, a+n[$$

for $n \geqslant 1$ and $f_n = f\chi_{X_n}$; as in Lemma 2.3.11, the bounds

$$|f_n| \leqslant |f_{n+1}| \to |f|,$$

imply that

$$\int_I |f(x)| d\lambda(x) = \lim_n \int_I |f_n| d\lambda(x) = \lim_n \int_a^{a+n} |f(x)| dx = \int_a^{+\infty} |f(x)| dx < +\infty$$

where we used the monotone convergence theorem, and the earlier comparison (2.23) together with the assumption on the Riemann-integrability of f. It follows that f is Lebesgue-integrable, and from Lemma 2.3.11, we get

$$\int_I f(x)d\lambda(x) = \int_a^{+\infty} f(x)dx$$

using the definition of the Riemann integral again.

A similar argument applies to an unbounded function defined on a closed interval $[a, b[$ such that the Riemann integral

$$\int_a^b f(x)dx$$

converges absolutely at a and (or) b.

CHAPTER 3

First applications of the integral

Having constructed the integral, and proved the most important convergence theorems, we can now develop some of the most important applications of the integration process. The first one is the well-known use of integration to construct *functions* – here, the most important gain compared with Riemann integrals is the much greater flexibility and generality of the results, which is well illustrated by the special case of the Fourier transform. The second application depends much more on the use of Lebesgue's integral: it is the construction of *spaces of functions* with good analytic properties, in particular, completeness. After these two applications, we give some of a probabilistic nature.

3.1. Functions defined by an integral

One of the most common use of integration is an averaging process which can be used to create functions by integrating over a fixed set a function of two (or more) variables, one of which is seen as a parameter. This procedure has a regularizing effect – even very rough functions, after integration, may become quite nice. This provides analysis with one of its most important technical tools.

Regularity properties of a function typically make sense when they are defined on a space with a topological structure. For simplicity, we assume that we have such a space which is a *metric* space (X, d). For instance, X could be an interval $[a, b]$ or $\mathbf{R}$ with the usual metric $d(x, y) = |x - y|$. Furthermore, let $(Y, \mathcal{M}, \mu)$ be a measured space. Given a function

$$h : X \times Y \to \mathbf{C}$$

such that each "slice" restriction

$$h_x : y \mapsto h(x, y)$$

is measurable and μ-integrable, we can form a function f on X by integrating over y, i.e., we define

$$f(x) = \int_Y h(x, y) d\mu(y), \qquad \text{for all } x \in X.$$

In order to prove regularity properties of f, it is necessary to strengthen the minimal assumption of integrability of each h_x, in a way similar to the use of the domination condition in Lebesgue's dominated convergence theorem. We assume that there exists $g \in L^1(\mu)$ such that

$$|h(x, y)| \leqslant g(y) \tag{3.1}$$

for all $(x, y) \in X \times Y$.

With these conditions, we first establish that if h is continuous *with respect to* $x \in X$ for all y, then f is also continuous. In fact, we prove a more precise result:

Proposition 3.1.1 (Continuity under the integral sign). *Let X, Y, h and f be as described above, and let $x_0 \in X$ be given. Assume that for μ-*almost all *$y \in Y$, the* *function*

$$h(\cdot, y) : x \mapsto h(x, y)$$

defined on the metric space X is continuous at x_0. Then f is continuous at x_0.

In particular, if these functions are all continuous at all $x_0 \in X$ for $y \in Y$, then f is continuous on all of X.

Proof. Since X is a metric space, continuity can be detected using convergent sequences, and therefore it is enough to prove that, for any sequence (x_n) in X such that

$$\lim_{n \to +\infty} x_n = x_0,$$

the sequence $(f(x_n))$ is also convergent, and

$$\lim_{n \to +\infty} f(x_n) = f(x_0).$$

But this almost writes itself using the dominated convergence theorem! Indeed, consider the sequence of functions

$$u_n : \begin{cases} Y & \to \mathbf{C} \\ y \mapsto h(x_n, y). \end{cases}$$

Writing down $f(x_n)$ and $f(x_0)$ as integrals, our goal is to prove

$$\int_Y h(x_n, y) d\mu(y) \longrightarrow \int_Y h(x_0, y) d\mu(y),$$

and this is clearly a job for the dominated convergence theorem.

Precisely, the assumption of continuity implies that $u_n(y) = h(x_n, y) \to h(x_0, y)$ for almost all $y \in Y$. Let $Y_0 \subset Y$ be the exceptional set where this is not true, and redefine $u_n(y) = 0$ for $y \in Y_0$: then the limit holds for *all* $y \in Y$ with limit

$$\tilde{h} : y \mapsto \begin{cases} 0 & \text{if } y \in Y_0, \\ h(x_0, y) & \text{otherwise.} \end{cases}$$

Moreover, we have also

$$|u_n(y)| \leqslant g(y)$$

with $g \in L^1(\mu)$. Hence, the dominated convergence theorme applies to prove

$$\lim_{n \to +\infty} \int_Y u_n(y) d\mu(y) = \int_Y \tilde{h}(y) d\mu(y) = \int_Y h(x_0, y) d\mu(y),$$

since $\tilde{h}(y) = h(x_0, y)$ almost everywhere. □

Remark 3.1.2. Note how we dealt in pitiless detail with the fact that we assumed continuity at x_0 only for almost all y. More generally, we have showed that, in the dominated convergence theorem, it is not necessary that the sequence converge pointwise to the limiting function for *all* points: it is enough that this happen μ-almost everywhere. We will use this type of easy property from now on without always mentioning it explicitly.

Example 3.1.3. Here is an example where the additional flexibility is useful. Consider $Y = [0, 1]$ with the Lebesgue measure (denoted simply dy); for $f \in L^1(Y)$, let $g : [0, 1] \to \mathbf{C}$ be the "primitive" of f, defined by

$$g(x) = \int_0^x f(y) dy = \int_{[0,x]} f(y) dy.$$

Then we claim that g is continuous – this independently of any other assumption on f, which may well be unbounded and everywhere discontinuous on Y (so that bounding $g(x+h)-g(x)$ by bounding f is not obvious at all!)

The point is that we can write

$$g(x) = \int_Y h(x,y)dy,$$

with $h(x,y) = f(y)\chi_{[0,x]}(y)$, which incorporates the condition on the range of integration in a two-variables function. This function h satisfies

$$|h(x,y)| \leqslant |f(y)| \text{ for all } x \in [0,1],$$

and hence, since $f \in L^1([0,1])$, we see that (3.1) is satisfied. Moreover, for any *fixed* y, we have

$$h(x,y) = f(y)\chi_{[0,x]}(y) = \begin{cases} f(y) & \text{if } 0 \leqslant y \leqslant x \\ 0 & \text{otherwise,} \end{cases}$$

which is a step function, with a single discontinuity at $x_0 = y$. Hence, for any fixed x_0, these functions are almost all continuous at x_0, and the proposition above implies the continuity of g on $[0,1]$.

Beyond continuity, we can investigate additional regularity properties, like differentiability, when this makes sense. We consider the simplest case where $X = I$ is an open interval in $\mathbf{R}$. In addition to the earlier assumption (3.1), which is always assumed to hold, another similar extra assumption is needed.

Proposition 3.1.4 (Differentiability under the integral sign). *Let $X = I \subset \mathbf{R}$ be a non-empty interval, let Y, h and f be as above. Assume that, for almost all $y \in Y$, the function*

$$h(\cdot, y) \,:\, I \to \mathbf{C}$$

is differentiable, and that its derivative satisfies a bound

$$\left|\frac{d}{dx}h(x,y)\right| \leqslant g_1(y). \tag{3.2}$$

for all $(x,y) \in X \times Y$ and some integrable function $g_1 \in L^1(Y,\mu)$

Then f is differentiable on I, and we have

$$f'(x) = \int_Y \frac{d}{dx}h(x,y)d\mu(y).$$

Proof. Let

$$f_1(x) = \int_Y \frac{d}{dx}h(x,y)d\mu(y),$$

which is well-defined according to our additional assumption (3.2).

Now fix $x_0 \in I$, and let us prove that f is differentiable at x_0 with $f'(x_0) = f_1(x_0)$. By definition, this means proving that

$$\lim_{\substack{\delta\to 0\\ \delta\neq 0}} \frac{f(x_0+\delta)-f(x_0)}{\delta} = f_1(x_0).$$

First of all, we observe that since I is open, there exists some $\alpha > 0$ such that $x_0+\delta \in I$ if $|\delta| < \alpha$. Now observe that

$$\frac{f(x_0+\delta)-f(x_0)}{\delta} = \int_Y \frac{h(x_0+\delta,y)-h(x_0,y)}{\delta}d\mu(y),$$

so that the problems almost becomes that of applying the previous proposition. Precisely, define

$$\psi(\delta, y) = \begin{cases} \dfrac{h(x_0+\delta, h) - h(x_0, y)}{\delta} & \text{if } \delta \neq 0 \\ \dfrac{d}{dx} h(x_0, y) & \text{if } \delta = 0, \end{cases}$$

for $|\delta| < \alpha$ and $y \in Y$. The assumption of differentiability on $h(x, y)$ for almost all y implies that $\delta \mapsto \psi(\delta, y)$ is continuous at 0 for almost all $y \in Y$. We now check that the assumption (3.1) holds for this new function of δ and y. First, if $\delta = 0$, we have

$$|\psi(0, y)| \leqslant g_1(y),$$

by (3.1); for $\delta \neq 0$, we use the mean-value theorem: there exists some $\eta \in [0, 1]$ such that

$$\begin{aligned} |\psi(\delta, y)| &= \Big| \frac{h(x_0+\delta, y) - h(x_0, y)}{\delta} \Big| \\ &= \Big| \frac{d}{dx} h(x_0 + \eta\delta, y) \Big| \leqslant g_1(y). \end{aligned}$$

It follows that (3.2) is valid for ψ instead of h with g_1 taking place of g

Hence, the previous proposition of continuity applies, and we deduce that the function

$$\delta \mapsto \int_Y \psi(\delta, y) d\mu(y)$$

is (defined and) continuous at 0, which is exactly the desired statement since

$$\int_Y \frac{d}{dx} h(x_0, y) d\mu(y) = f_1(x_0).$$

□

3.2. An example: the Fourier transform

The results of the previous section can very often be applied extremely easily. We use them here to present the definition and the simplest properties of the *Fourier transform* of an integrable function defined on $\mathbf{R}$ (with respect to the Lebesgue measure). This very important operation will also be considered in greater detail in later chapters.

First of all, we define a function $e : \mathbf{C} \to \mathbf{C}$ by

$$e(z) = e^{2i\pi z}. \tag{3.3}$$

This close cousin of the exponential function satisfies

$$e(x+y) = e(x)e(y),\ e(x+1) = e(x) \text{ and } |e(x)| = 1 \text{ if } x \in \mathbf{R}$$

(the reason the factor $2i\pi$ was introduced is to obtain a 1-periodic function). Note also that

$$e(x) = 1 \text{ if and only if } x \in \mathbf{Z}.$$

Definition 3.2.1 (Fourier transform of an integrable function). Let $f \in L^1(\mathbf{R})$ be a complex-valued function integrable with respect to the Lebesgue measure. The *Fourier transform* of f is the function $\hat{f} : \mathbf{R} \to \mathbf{C}$ defined by the integral

$$\hat{f}(t) = \int_{\mathbf{R}} f(y) e(-yt) dy$$

for all $t \in \mathbf{R}$.

Note that this is well-defined precisely because $|f(y)e(-yt)| = |f(y)|$ and f is assumed to be integrable.

REMARK 3.2.2. Many other normalizations of the Fourier transform occur in the literature, such as

$$\int_{\mathbf{R}} f(x)e^{ixt}dx, \quad \int_{\mathbf{R}} f(x)e^{-ixt}dx, \quad \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} f(x)e^{-ixt}dx, \text{ etc...}$$

The normalization we have chosen has the advantage of leading to a very simple Fourier Inversion Formula (as we will see in a later chapter). Of course, all definitions are related through elementary transformations, for instance the three functions above are, respectively, equal to

$$\hat{f}(-t/2\pi), \quad \hat{f}(t/2\pi), \quad \frac{1}{\sqrt{2\pi}}\hat{f}(t/2\pi).$$

However, one should always check which definition is used before using a formula found in a book (e.g., a table of Fourier transforms).

PROPOSITION 3.2.3 (Elementary regularity properties). (1) *Let $f \in L^1(\mathbf{R})$. Then $\hat{f}$ is a bounded continuous function on $\mathbf{R}$ such that*

$$\sup_{x\in\mathbf{R}} |\hat{f}| \leqslant \|f\|_1. \tag{3.4}$$

(2) *If the function $g(x) = xf(x)$ is also integrable on $\mathbf{R}$ with respect to the Lebesgue measure, then the Fourier transform $\hat{f}$ of f is of C^1 class on $\mathbf{R}$, and its derivative is given by*

$$\hat{f}'(t) = -2i\pi\hat{g}(t)$$

for $t \in \mathbf{R}$.

(3) *If f is of C^1 class, and is such that $f' \in L^1(\mathbf{R})$ and moreover*

$$\lim_{x\to\pm\infty} f(x) = 0,$$

then we have

$$\widehat{f'}(t) = 2i\pi t\hat{f}(t).$$

PROOF. The point is of course that we can express the Fourier transform as a function defined by integration using the two-variable function $h(t, y) = f(y)e(-yt)$. Since $|h(t, y)| = |f(t)|$ for all $y \in \mathbf{R}$, condition (3.1) holds. Moreover, for any fixed y, the function $t \mapsto h(t, y)$ on $\mathbf{R}$ is a constant times an exponential, and is therefore continuous everywhere. Therefore Proposition 3.1.1 shows that $\hat{f}$ is continuous on $\mathbf{R}$. The property (3.4) is also immediate since $|e(x)| = 1$.

For Part (2), we note that h is also (indefinitely) differentiable on $\mathbf{R}$ for any fixed y, in particular differentiable with

$$\frac{d}{dt}h(t, y) = -2i\pi yf(y)e(yt) \text{ hence } \left|\frac{d}{dt}h(t, y)\right| \leqslant 2\pi|yf(y)| = 2\pi|g(y)|.$$

If $g \in L^1(\mathbf{R})$, we see that we can now apply directly Proposition 3.1.4, and deduce that $\hat{f}$ is differentiable and satisfies

$$\hat{f}'(t) = \int_{\mathbf{R}} 2i\pi yf(y)e(yt)dy = -2i\pi\hat{g}(t),$$

for $t \in \mathbf{R}$. Since this derivative is itself the Fourier transform of an integrable function, it is also continuous on $\mathbf{R}$ by the previous result.

As for (3), assume that f' exists and is integrable, and that $f \to 0$ at infinity. Then, one can integrate by parts in the definition of $\widehat{f'}$ and find

$$\widehat{f'}(t) = \int_{\mathbf{R}} f'(y)e(-yt)dy = 2i\pi t \int_{\mathbf{R}} f(y)e(-yt)dy$$

(the boundary terms disappear because of the decay at infinity; since f is here differentiable, hence continuous, we can justify the integration by parts by comparison with a Riemann integral, but we will give a more general version later on). $\square$

Example 3.2.4. Let $f : \mathbf{R} \to \mathbf{C}$ be a continuous function with *compact support* (i.e., such that there exists $M \geqslant 0$ for which $f(x) = 0$ whenever $|x| \geqslant M$). Then f is integrable (being bounded and zero outside $[-M, M]$), and moreover the functions

$$x \mapsto x^k f(x) \tag{3.5}$$

are integrable for all $k \geqslant 1$ (they have the same compact support). Using proposition 3.2.3, it follows by induction on $k \geqslant 1$ that $\hat{f}$ is indefinitely differentiable on $\mathbf{R}$, and is given by

$$\hat{f}^{(k)}(t) = (-2i\pi)^k \int_{\mathbf{R}} y^k f(y)e(-yt)dy.$$

The same argument applies more generally to any f such that all the functions (3.5) are integrable, even if they are not compactly supported. Examples of these are given by $f(y) = e^{-|y|}$ or $f(y) = e^{-y^2}$.

Remark 3.2.5. In the language of normed vector spaces, the inequality (3.4) means that the "Fourier transform" map

$$\begin{cases} L^1(\mathbf{R}) \to C_b(\mathbf{R}) \\ f \mapsto \hat{f} \end{cases}$$

is continuous, and has norm $\leqslant 1$, where $C_b(\mathbf{R})$ is the space of bounded continuous functions on $\mathbf{R}$, which is a normed vector space with the norm

$$\|f\|_\infty = \sup_{x \in \mathbf{R}} |f(x)|.$$

This continuity means, in particular, that if (f_n) is a sequence of integrable functions converging to f_0 in the L^1-norm, i.e., such that

$$\|f_n - f_0\|_1 = \int_{\mathbf{R}} |f_n(x) - f_0(x)|dx \longrightarrow 0,$$

then the Fourier transforms $\hat{f}_n$ converge *uniformly* to $\hat{f}_0$ on $\mathbf{R}$ (since this is the meaning of $\|f_n - f_0\|_\infty \to 0$).

Remark 3.2.6. Let (Ω, Σ, P) be a probability space, and let X be a real-valued random variable on Ω. Probabilists define the *characteristic function* of X to be the function $\varphi : \mathbf{R} \to \mathbf{C}$ defined by

$$\varphi(t) = \int_\Omega e^{itX} dP = E(e^{itX}).$$

Because $|e^{itX}| = 1$ (since X is real-valued), and since the measure P is a probability measure, this map φ is well-defined. Applying Proposition 3.1.1, we see that it is always a continuous function.

Of course, this is a variant of the Fourier transform. Indeed, if we denote by $\mu = X(P)$ the probability measure on $\mathbf{R}$ which is the law of X, we have

$$\varphi(t) = \int_{\mathbf{R}} e^{ity} d\mu(y)$$

by the change of variable formula (2.12) for image measures. In particular, if the law μ is of the type

$$\mu = f(x)dx$$

for some function f on $\mathbf{R}$ (where dx is again the Lebesgue measure), we have

$$\varphi(t) = \int_{\mathbf{R}} e^{ity} f(y)dy = \hat{f}\Big(-\frac{t}{2\pi}\Big)$$

(note that f is necessarily integrable because it must be $\geqslant 0$ and

$$\int_{\mathbf{R}} f d\mu = \int_{\Omega} dP = 1$$

by (2.12)).

In a later chapter, we will study the probabilistic aspects of the characteristic function and its many fundamental applications.

3.3. L^p-spaces

The second important use of integration theory – in fact, maybe the most important – is the construction of *function spaces* with excellent analytic properties, in particular *completeness* (i.e., convergence of Cauchy sequences). Indeed, we have the following property, which we will prove below: given a measure space $(X, \mathcal{M}, \mu)$, and a sequence (f_n) of integrable functions on X such that, for any $\varepsilon > 0$, we have

$$\|f_n - f_m\|_1 < \varepsilon$$

for all n, m large enough (a Cauchy sequence), there exists a limit function f, unique except that it may be changed arbitrarily on any μ-negligible set, such that

$$\lim_{n \to +\infty} \|f_n - f\| = 0$$

(but f is not usually a pointwise limit of the (f_n)).

Before considering this, we need to be involved in a bit of abstract manipulations in order to have structures which coincide exactly with the expected topological notions. In particular, note that the norm $\|f\|_1$, as defined, does not have the property that $\|f\|_1 = 0$ implies that $f = 0$. The solution is to consider systematically functions defined "up to a perturbation defined on a set of measure zero".

Definition 3.3.1. Let $(X, \mathcal{M}, \mu)$ be a measured space. The space $L^1(X, \mu) = L^1(\mu)$ is defined to be the quotient vector space

$$\{f : X \to \mathbf{C} \mid f \text{ is integrable}\}/N$$

where

(3.6)

$$N = \{f \mid f \text{ is measurable and } f \text{ is zero } \mu\text{-almost everywhere}\} = \Big\{f \mid \int_X |f| d\mu = 0\Big\}.$$

This is a normed vector space for the norm

$$\|f\|_1 = \int_X |f| d\mu,$$

in particular $\|f\|_1 = 0$ if and only if $f = 0$ in $L^1(X, \mu)$.

From now on, the notation $L^1(\mu)$ and $L^1(X, \mu)$ refer exclusively to this definition

Concretely, although this means that an element of $L^1(X, \mu)$ can not properly be thought of as a function defined everywhere on X, the equivalence relation does not induce much difficulty. One usually works with "actual" functions f, and one remembers to say that $f = g$ if it is shown that f and g coincide except on a set of measure zero. Most importantly, in order to define a map

$$L^1(X, \mu) \xrightarrow{\phi} M,$$

for any set M, one usually defines a map on actual (integrable) functions, before checking that the value of $\phi(f)$ is the same as that of $\phi(g)$ if $f - g$ is zero almost everywhere. This is, for instance, exactly why the norm is well-defined on $L^1(\mu)$.

On the other hand, fix a point x_0 and consider the map

$$\phi : f \mapsto f(x_0)$$

defined on all integrable functions on X. If $\mu(\{x_0\}) = 0$, any two functions which differ only at x_0 are equal almost everywhere, and therefore the value of $\phi(f)$ is not well-defined on the equivalence classes which are the elements of $L^1(X, \mu)$. Hence, under this condition, one *can not* speak sensibly of the value of a function in $L^1(\mu)$ at x_0.

It will be quickly seen that manipulating elements of $L^1(\mu)$ is quite easy, and it is frequent to abuse notation by stating that they are "integrable functions on X". Also, one often has a function f which is constructed in a way which makes sense almost everywhere, say outside Z_0, and is integrable on $X - Z_0$; then defining $f(z) = 0$ if $z \in Z_0$, one obtains a well-defined element of $L^1(\mu)$. The following exercise is a good way to get some familiarity with this type of reasoning.

Exercise 3.3.2. Let L be the set of all pairs (Y, f) where $Y \in \mathcal{M}$ is such that $\mu(X - Y) = 0$, and $f : Y \to \mathbf{C}$ is a measurable function such that

$$\int_Y |f| d\mu < +\infty.$$

Let L_1 be the quotient set $L/\sim$ where the equivalence relation $\sim$ is given by

$$(Y, f) \sim (Y', f') \text{ if } f(x) = f'(x) \text{ for all } x \in Y \cap Y'.$$

(1) Show that this is indeed an equivalence relation.

(2) Show that L_1 is a normed vector space with addition and multiplication obtained from

$$(Y, f) + (Y', f') = (Y \cap Y', f + f'), \text{ and } \lambda(Y, f) = (Y, \lambda f),$$

by passing to the quotient, and for the norm defined by

$$\|(Y, f)\| = \int_Y |f| d\mu.$$

(3) Show that the map

$$\begin{cases} L^1(\mu) \to L_1 \\ f \mapsto (X, f) \end{cases}$$

is a linear isometry of normed vector spaces, with inverse given by

$$(Y, f) \mapsto g \text{ such that } g(x) = \begin{cases} f(x) & \text{if } x \in Y \\ 0 & \text{otherwise.} \end{cases}$$

(4) Show that the dominated convergence theorem implies the following result: if (f_n) is a sequence of elements $f_n \in L_1$ such that $f_n \to f$ almost everywhere, and if $|f_n| \leqslant g$ with $g \in L_1$, then f_n converges to f in L_1 and

$$\int_X f d\mu = \lim_{n \to +\infty} \int_X f_n d\mu.$$

Here is the first main result concerning the space L^1, which will lead quickly to its completeness property.

Proposition 3.3.3. *Let $(X, \mathcal{M}, \mu)$ be a measured space, and let (f_n) be a function of elements $f_n \in L^1(\mu)$. Assume that the series $\sum f_n$ is* normally convergent *in $L^1(\mu)$, i.e., that*

$$\sum_{n \geqslant 1} \|f_n\|_1 < +\infty.$$

Then the series

$$\sum_{n \geqslant 1} f_n(x)$$

converges almost everywhere in X, and if $f(x)$ denotes its limit, well-defined almost everywhere, we have $f \in L^1(\mu)$ and

$$\int_X f d\mu = \sum_{n \geqslant 1} \int_X f_n d\mu. \tag{3.7}$$

Moreover the convergence is also valid in the norm $\|\cdot\|_1$, i.e., we have $\|f_n - f\| \to 0$ as $n \to +\infty$.

For this proof only, we will be quite precise and careful about distinguishing functions and equivalence classes modulo those which are zero almost everywhere; the type of boring formal details involved will not usually be repeated afterwards.

Proof. Define

$$h(x) = \sum_{n \geqslant 1} |f_n(x)| \in [0, +\infty]$$

for $x \in X$; if the values of f_n are modified on a set Y_n of measure zero, then h is changed, at worse, on the set

$$Y = \bigcup_{n \geqslant 1} Y_n$$

which is still of measure zero. This indicates that h is well-defined almost everywhere, and it is measurable. It is also non-negative, and hence its integral is well-defined in $[0, +\infty]$. By Corollary 2.2.5 to the monotone convergence theorem, we have

$$\int_X h d\mu = \sum_{n \geqslant 1} \int_X |f_n| d\mu = \sum_{n \geqslant 1} \|f_n\|_1 < +\infty \text{ by assumption,}$$

and from this we derive in particular that h is *finite almost everywhere* (see Proposition 2.2.2, (1)).

For any x such that $h(x) < \infty$, the numeric series $\sum f_n(x)$ converges absolutely, and hence has a sum $f(x)$ such that $|f(x)| \leqslant h(x)$. We can extend f to X by defining, for instance, $f(x) = 0$ on the negligible set (say, Z) of those x where $h(x) = +\infty$.

In any case, we have $|f| \leqslant h$, and therefore the above inequality implies that $f \in L^1(\mu)$. We can now apply the dominated convergence theorem to the sequence of partial sums

$$u_n(x) = \sum_{1 \leqslant i \leqslant n} f_i(x) \text{ if } x \notin Z, \quad u_n(x) = 0 \text{ otherwise}$$

which satisfy $u_n(x) \to f(x)$ for all x and $|u_n(x)| \leqslant h(x)\chi_Z(x)$ for all n and x. We conclude that

$$\int_X f d\mu = \lim_n \int_X u_n d\mu = \lim_n \sum_{1 \leqslant i \leqslant n} \int_X f_i d\mu = \sum_{n \geqslant 1} \int_X f_n d\mu.$$

□

Proposition 3.3.4 (Completeness of the space $L^1(X, \mu)$). (1) *Let $(X, \mathcal{M}, \mu)$ be a measured space. Then any Cauchy sequence in $L^1(X, \mu)$ has a limit in $L^1(X, \mu)$, or in other words, the normed vector space $L^1(\mu)$ is complete.*[1]

(2) *More precisely, or concretely, we have the following: if (f_n) is a Cauchy sequence in $L^1(\mu)$, there exists a unique $f \in L^1(\mu)$ such that $f_n \to f$ in $L^1(\mu)$, i.e., such that*

$$\lim_{n \to +\infty} \|f_n - f\|_1 = 0.$$

Moreover, there exists a subsequence (f_{n_k}) of (f_n) such that

$$\lim_{k \to +\infty} f_{n_k}(x) = f(x),$$

for μ-almost all x.

Remark 3.3.5. Part (2) can not be improved: it may be that the sequence $(f_n(x))$ itself does not converge, for any x (an example is given below in Exercise 3.3.6), or that it diverges at some points. Thus, it is really an L^1-type statement, and it would be rather inelegant to express it purely in terms of functions.

Proof. We recall the Cauchy condition: for any $\varepsilon > 0$, there exists $N(\varepsilon) \geqslant 1$ such that

$$\|f_n - f_m\|_1 < \varepsilon \tag{3.8}$$

for all $n \geqslant N$ and $m \geqslant N$.

Contrary to many arguments involving Cauchy sequences, we can not claim that $(f_n(x))$ is itself a Cauchy sequence for some $x \in X$. But we will first construct the limit function by appealing to Proposition 3.3.3, proving Part (2) in effect: the Cauchy condition states that, in L^1-norm, all terms of the sequence are very close for n large enough, and this provides a series of differences which converges normally. Precisely, consider successively $\varepsilon = \varepsilon_k = 2^{-k}$ for $k \geqslant 1$.[2] Then, using induction on k, we see that there exists a strictly increasing sequence (n_k) such that

$$\|f_{n_{k+1}} - f_{n_k}\|_1 < 2^{-k}$$

for $k \geqslant 1$. We claim that (f_{n_k}) converges almost everywhere. Indeed, the series

$$\sum_{k \geqslant 1} (f_{n_{k+1}} - f_{n_k})$$

[1] It is therefore what is called a Banach space.

[2] Any other sequence of ε_k's decreasing to 0 with $\sum \varepsilon_k < +\infty$ would do as well.

is a series with terms $g_k = f_{n_{k+1}} - f_{n_k} \in L^1(\mu)$ such that

$$\sum_{k \geqslant 1} \|g_k\|_1 \leqslant \sum_{k \geqslant 1} 2^{-k} < +\infty.$$

By Proposition 3.3.3, this series converges almost everywhere, and in $L^1(\mu)$, to a function $g \in L^1(\mu)$. But since the partial sums are given by the telescoping sums

$$\sum_{k=1}^{K} g_k = (f_{n_{K+1}} - f_{n_K}) + (f_{n_K} - f_{n_{K-1}}) + \cdots + (f_{n_2} - f_{n_1}) = f_{n_{K+1}} - f_{n_1},$$

this means that the subsequence (f_{n_k}) converges almost everywhere, with limit $f = g + f_{n_1}$.

Moreover, since the convergence is valid in L^1, we have

$$\lim_{k \to +\infty} \|f - f_{n_k}\|_1 = 0, \tag{3.9}$$

which means that f is a limit point of the Cauchy sequence (f_n). However, it is well-known that a Cauchy sequence with a limit point, in a metric space, converges necessarily to this limit point. Here is the argument for completeness: for all n and k, we have

$$\|f - f_n\|_1 \leqslant \|f - f_{n_k}\|_1 + \|f_{n_k} - f_n\|_1,$$

and by (3.8), we know that $\|f_n - f_{n_k}\|_1 < \varepsilon$ if n and n_k are both $\geqslant N(\varepsilon)$. Moreover, by (3.9), we have $\|f - f_{n_k}\|_1 < \varepsilon$ for $k > K(\varepsilon)$.

Now fix some $k > K(\varepsilon)$ such that, in addition, we have $n_k > N(\varepsilon)$ (this is possible since n_k is strictly increasing). Then, for all $n > N(\varepsilon)$, we obtain $\|f - f_n\|_1 < 2\varepsilon$, and this proves the convergence of (f_n) towards f in $L^1(\mu)$. $\square$

EXERCISE 3.3.6. Consider the function

$$f : [1, +\infty[\to \mathbf{R}$$

defined as follows: given $x \geqslant 1$, let $n \geqslant 1$ be the integer with $n \leqslant x < n+1$, and write $n = 2^k + j$ for some $k \geqslant 0$ and $0 \leqslant j < 2^k$; then define

$$f(x) = \begin{cases} 1 & \text{if } j2^{-k} \leqslant x - n < (j+1)2^{-k} \\ 0 & \text{otherwise.} \end{cases}$$

(1) Sketch the graph of f; what does it look like?

(2) For $n \geqslant 1$, we define $f_n(x) = f(x+n)$ for $x \in X = [0,1]$. Show that $f_n \in L^1(X, d\lambda)$ and that $f_n \to 0$ in $L^1(X)$.

(3) For any $x \in [0,1]$, show that the sequence $(f_n(x))$ does not converge.

(4) Find an explicit subsequence (f_{n_k}) which converges almost everywhere to 0.

The arguments used above to prove Theorem 3.3.10 can be adapted quite easily to the L^p-spaces for $p \geqslant 1$, which are defined as follows (there is also an L^∞-space, which has a slightly different definition found below).

DEFINITION 3.3.7 (L^p-spaces, $p < +\infty$). Let $(X, \mathcal{M}, \mu)$ be a measure space and let $p \geqslant 1$ be a real number. The $L^p(X, \mu) = L^p(\mu)$ is defined to be the quotient vector space

$$\{f : X \to \mathbf{C} \mid |f|^p \text{ is integrable}\}/N$$

where N is defined as in *(3.6)*. This space is equiped with the norm

$$\|f\|_p = \Big(\int_X |f|^p d\mu\Big)^{1/p}.$$

The first thing to do is to check that the definition actually makes sense, i.e., that the set of functions f where $|f|^p$ is integrable *is* a vector space, and then that $\|\cdot\|_p$ defines a norm on $L^p(\mu)$, which is not entirely obvious (in contrast with the case $p = 1$). For this, we need some inequalities which are also quite useful for other purposes.

We thus start a small digression. First, we recall that a function $\varphi : I \to [0, +\infty[$ defined on an interval I is said to be *convex* if it is such that

$$\varphi(ra + sb) \leqslant r\varphi(a) + s\varphi(b) \tag{3.10}$$

for any a, b such that $a \leqslant b$, a, $b \in I$ and any real numbers r, $s \geqslant 0$ such that $r + s = 1$. If φ is twice-differentiable with continuous derivatives, it is not difficult to show that φ is convex on an open interval if and only if $\varphi'' \geqslant 0$.

PROPOSITION 3.3.8 (Jensen, Hölder and Minkowski inequalities). *Let $(X, \mathcal{M}, \mu)$ be a fixed measure space.*

(1) *Assume μ is a probability measure. Then, for any function*

$$\varphi : [0, +\infty[\to [0, +\infty[$$

which is non-decreasing, continuous and convex, and any measurable function

$$f : X \to [0, +\infty[,$$

we have Jensen's inequality*:*

$$\varphi\Big(\int_X f(x) d\mu(x)\Big) \leqslant \int_X \varphi(f(x)) d\mu(x), \tag{3.11}$$

with the convention $\varphi(+\infty) = +\infty$.

(2) *Let $p > 1$ be a real number and let $q > 1$ be the "dual" real number such that $p^{-1} + q^{-1} = 1$. Then for any measurable functions*

$$f,\ g : X \to [0, +\infty],$$

we have Hölder's inequality*:*

$$\int_X fg d\mu \leqslant \Big(\int_X f^p d\mu\Big)^{1/p} \Big(\int_X g^q d\mu\Big)^{1/q} = \|f\|_p \|g\|_q. \tag{3.12}$$

(3) *Let $p > 1$ be any real number. Then for any measurable functions*

$$f,\ g : X \to [0, +\infty],$$

we have Minkowski's inequality*:*

$$\begin{aligned} \|f + g\|_p &= \Big(\int_X (f + g)^p d\mu\Big)^{1/p} \\ &\leqslant \Big(\int_X f^p d\mu\Big)^{1/p} + \Big(\int_X g^p d\mu\Big)^{1/p} = \|f\|_p + \|g\|_p. \end{aligned} \tag{3.13}$$

PROOF. (1) The basic point is that the defining inequality (3.10) is itself a version of (3.11) for a step function f such that $f(X) = \{a, b\}$ and

$$r = \mu\{x \in X \mid f(x) = a\}, \quad s = \mu\{x \in X \mid f(x) = b\}$$

(since $r + s = 1$). It follows easily by induction that (3.11) is true for any step function $s \geqslant 0$. Next, for an arbitrary $f \geqslant 0$, we select a sequence of step functions (s_n) which is non-decreasing and converges pointwise to f. By the monotone convergence theorem, we have

$$\int s_n d\mu \to \int f d\mu$$

and similarly, since φ is itself non-decreasing, we have

$$\varphi(s_n(x)) \to \varphi(f(x))$$

and $\varphi \circ s_n \leqslant \varphi \circ s_{n+1}$, so that

$$\int \varphi(s_n(x))d\mu \to \int \varphi(f(x))d\mu$$

by the monotone convergence theorem again.

Finally, the continuity of φ and Jensen's inequality for step functions implies that

$$\varphi\Bigl(\int_X fd\mu\Bigr) = \lim_{n\to+\infty} \varphi\Bigl(\int_X s_n d\mu\Bigr) \leqslant \lim_{n\to+\infty} \int_X \varphi(s_n(x))d\mu = \int_X \varphi(f(x))d\mu.$$

(2) We claim that if $1/p + 1/q = 1$, we have *Young's inequality*

$$xy \leqslant \frac{x^p}{q} + \frac{y^q}{q}, \text{ for } x \geqslant 0,\ y \geqslant 0. \tag{3.14}$$

Assuming this (to be checked quickly below), we continue as follows to prove (3.12): it is enough to consider the cases when $|f|^p$ and $|g|^q$ are integrable (otherwise, the result is trivial), and then by replacing f with $f/\|f\|_p$ and g with $g/\|g\|_q$, homogeneity shows that it is enough to prove the inequality when $\|f\|_p = \|g\|_q = 1$. But in that case, we use the pointwise Young inequality

$$f(x)g(x) \leqslant \frac{f(x)^p}{p} + \frac{g(x)^q}{q},$$

and we integrate to obtain

$$\int_X f(x)g(x)d\mu \leqslant \frac{\|f\|_p}{p} + \frac{\|f\|_q}{q} = 1,$$

To show (3.14), we use the well-known convexity of the exponential function on $\mathbf{R}$: since $1/p + 1/q = 1$, we have

$$e^{a/p+b/q} \leqslant \frac{e^a}{p} + \frac{e^b}{q},$$

for a, $b \in \mathbf{R}$. Taking $a = p\log x$, $b = p\log y$, we get exactly the desired statement.

(3) Once more, the inequality is obvious if either $\|f\|_p = +\infty$ or $\|g\|_p = +\infty$. In the remaining case, we first note that

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

so that we also get $\|f+g\|_p < +\infty$. Now, we consider the auxiliary function

$$h = (f+g)^{p-1},$$

which has the property that $(f+g)^p = fh + gh$, so that by applying Hölder's inequality (3.12), we derive

$$\int_X (f+g)^p d\mu = \int_X fhd\mu + \int_X ghd\mu \leqslant \|f\|_p\|h\|_q + \|g\|_p\|h\|_q.$$

But

$$\|h\|_q = \Bigl(\int_X (f+g)^{q(p-1)}d\mu\Bigr)^{1/q} = \|f+g\|_p^{p/q} \text{ since } q(p-1) = p.$$

Moreover, $1 - 1/q = 1/p$, hence (considering separately the case where h vanishes almost everywhere) the last inequality gives

$$\|f+g\|_p = \Big(\int_X (f+g)^p d\mu\Big)^{p(1-1/q)} \leqslant \|f\|_p + \|g\|_p$$

after dividing by $\|h\|_q = \|f+g\|_p^{p/q} \in]0,+\infty[$. $\square$

Example 3.3.9. Consider the space $(\mathbf{R}, \mathcal{B}, \lambda)$; then the function f defined by

$$f(x) = \frac{\sin(x)}{x} \text{ if } x \neq 0 \text{ and } f(0) = 1$$

is in $L^2(\mathbf{R})$ (because it decays faster than $1/x^2$ for $|x| \geqslant 1$ and is bounded for $|x| \leqslant 1$), although it is *not* in $L^1(\mathbf{R})$ (as we already observed in the previous chapter). In particular, note that a function in L^p, $p \neq 1$, may well have the property that it is *not* integrable. This is one of the main ways to go around restriction to absolute convergence in Lebesgue's definition: many properties valid in L^1 are still valid in L^p, and can be applied to functions which are not integrable.

The next theorem, often known as the Riesz-Fisher Theorem, generalizes Proposition 3.3.3.

Theorem 3.3.10 (Completeness of L^p-spaces). *Let $(X, \mathcal{M}, \mu)$ be a fixed measure space.*

(1) *Let $p \geqslant 1$ be a real number, (f_n) a sequence of elements $f_n \in L^p(\mu)$. If*

$$\sum_{n \geqslant 1} \|f_n\|_p < +\infty,$$

the series

$$\sum_{n\geqslant 1} f_n$$

converges almost everywhere, and in L^p-norm, to a function $g \in L^p(\mu)$.

(2) *For any $p \geqslant 1$, the space $L^p(\mu)$ is a Banach space for the norm $\|\cdot\|_p$; more precisely, for any Cauchy sequence (f_n) in $L^p(\mu)$, there exists $f \in L^p(\mu)$ such that $f_n \to f$ in $L^p(\mu)$, and in addition there exists a subsequence (f_{n_k}), $k \geqslant 1$, such that*

$$\lim_{k\to+\infty} f_{n_k}(x) = f(x) \text{ for almost all } x \in X.$$

(3) *In particular, for $p = 2$, $L^2(\mu)$ is a Hilbert space with respect to the inner product*

$$\langle f, g\rangle = \int_X f(x)\overline{g(x)} d\mu.$$

Proof. Assuming (1) is proved, the same argument used to prove Proposition 3.3.4 starting from Proposition 3.3.3 can be copied line by line, replacing every occurence of $\|\cdot\|_1$ with $\|\cdot\|_p$, proving (2).

In order to show (1), we also follow the proof of Proposition 3.3.3: we define

$$h(x) = \sum |f_n(x)| \in [0, +\infty]$$

for $x \in X$, and observe that (by continuity of $y \mapsto y^p$ on $[0, +\infty]$) we also have

$$h(x)^p = \lim_{N\to+\infty} \Big(\sum_{n\leqslant N} |f_n(x)|\Big)^p.$$

This expresses h^p as a non-decreasing limit of non-negative functions, and therefore by the monotone convergence theorem, and the triangle inequality, we derive

$$\Big(\int_X h(x)^p d\mu(x)\Big)^{1/p} = \lim_{N\to+\infty}\Big(\int_X \Big(\sum_{n\leqslant N}|f_n(x)|\Big)^p d\mu(x)\Big)^{1/p}$$
$$= \lim_{N\to+\infty}\Big\|\sum_{n\leqslant N}|f_n|\Big\|_p \leqslant \sum_{n\geqslant 1}\|f_n\|_p < +\infty,$$

by assumption. It follows that h^p, and also h, is finite almost everywhere. This implies (as was the case for $p=1$) that the series

$$f(x) = \sum_{n\geqslant 1} f_n(x)$$

converges absolutely almost everywhere. Since $|f(x)| \leqslant h(x)$, we have $f \in L^p(\mu)$, and since

$$\Big\|f - \sum_{n\leqslant N} f_n\Big\|_p = \Big\|\sum_{n>N} f_n\Big\|_p \leqslant \sum_{n>N}\|f_n\|_p \to 0,$$

the convergence is also valid in L^p. □

REMARK 3.3.11. If (Ω,Σ,P) is a probability space, and X a random variable on Ω, then

$$\sigma^2(X) = V(X) = E(|X-E(X)|^2) = E(|X|^2) - |E(X)|^2$$

is called the *variance* of X. It is well-defined when $X \in L^2(\Omega)$ (see below for the easy check). Intuitively, it measures the average difference between X and its mean value; a "small" variance (in a sense that may depend on context) will indicate that the random variable is quite concentrated around its mean.

We now check the second formula for the variance given above: we have

$$E(|X-E(X)|^2) = E(|X|^2 - XE(\bar{X}) - \bar{X}E(X) + |E(X)|^2) = E(|X|^2) - 2|E(X)|^2 + |E(X)|^2$$

since $E(1) = 1$ is a probability space. Notice also that $V(aX) = |a|^2 V(X)$.

The square root $\sigma(X) = \sqrt{V(X)} = \|X - E(X)\|_2$ of the variance is called the *standard deviation* of X.(If, for a physical application, X corresponds to a quantity given in some unit u (length, weight, speed, etc), the variance corresponds to units u^2, and the standard deviation has the same unit as X.)

The variance and standard deviation of X only depend on the probability law μ of X, and one may speak of the variance or standard deviation of μ. Indeed, by (2.8), we have

$$E(X) = \int_{\mathbf{C}} x\,d\mu(x) \text{ and } V(X) = \int_{\mathbf{C}} (x - E(X))^2 d\mu(x).$$

In general, the variance of a sum $X+Y$ of random variables can not be computed directly in terms of the laws of X and Y. However, this is possible if X and Y are *independent*:

PROPOSITION 3.3.12. *Let (Ω,Σ,P) be a probability space, let X and Y be independent L^2 random variables on X. Then we have*

$$V(X+Y) = V(X) + V(Y). \tag{3.15}$$

Proof. We have $E(X+Y) = E(X)+E(Y)$ and therefore, using the second formula for the variance, we obtain

$$\begin{aligned} V(X+Y) &= E(|X|^2 + 2\operatorname{Re}(XY) + |Y|^2) - (|E(X)|^2 + 2\operatorname{Re}(E(X)E(Y)) + |E(Y)|^2) \\ &= V(X) + V(Y) + 2\operatorname{Re}(E(XY) - E(X)E(Y)) = V(X) + V(Y) \end{aligned}$$

by (2.16). □

This property of additivity for independent summands is of course false for the standard deviation. It is also false in general for the seemingly more natural quantity

$$E(|X - E(X)|) = \int_{\mathbf{C}} |x - E(X)| d\mu(x),$$

and this is one of the main reason why the variance is more useful as a measure of spread around the mean.

We now come to the last space in the L^p family, the one denoted L^∞. This, although rather different in some respects, plays an important role as a "limit" of L^p for $p \to +\infty$. Roughly speaking, its elements are measurable bounded functions, but the correct definition requires some care since one must (as usual) disregard values which are only attained on a set of measure zero.

Definition 3.3.13 (Essentially bounded functions). A measurable function

$$f : X \to \mathbf{C}$$

is said to be *essentially bounded* by $M \geqslant 0$ if

$$\mu(\{x \mid |f(x)| > M\}) = 0. \tag{3.16}$$

The numbers for which f is essentially bounded by M are used to define the "right" space of bounded measurable functions.

Proposition 3.3.14 (L^∞ space). *Let f be a measurable function on X and let*

$$\|f\|_\infty = \inf\{M \mid f \text{ is essentially bounded by } M\} \in [0, +\infty].$$

Then f is essentially bounded by $\|f\|_\infty$, or in other words the infimum is attained. Moreover, the quotient vector space

$$L^\infty(\mu) = \{f \mid \|f\|_\infty < +\infty\}/N,$$

where N is the subspace (3.6) *of measurable functions vanishing almost everywhere, is a normed vector space with norm $\|\cdot\|_\infty$.*

If $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$, we have $fg \in L^1(\mu)$ and

$$\int_X |fg| d\mu \leqslant \|f\|_1 \|g\|_\infty. \tag{3.17}$$

The last inequality should be thought of as the analogue of Hölder's inequality for the case $p = 1$, $q = +\infty$.

Note that the obvious inequality $\|f\|_\infty \leqslant \sup\{f(x)\}$ is not, in general, an equality, even if the supremum is attained. For instance, if we consider $([0,1], \mathcal{B}, \lambda)$, and take $f(x) = x\chi_{\mathbf{Q}}(x)$, we find that $\|f\|_\infty = 0$, although the function f takes all rational values in $[0,1]$, and in particular has maximum equal to 1 on $[0,1]$.

The definition of the L^∞ norm is most commonly used as follows: we have

$$|f(x)| \leqslant M$$

μ-almost everywhere, if and only if

$$M \geqslant \|f\|_\infty.$$

PROOF OF PROPOSITION 3.3.14. We must check that $M = \|f\|_\infty$ satisfies the condition (3.16), when $M < +\infty$ (the case $M = +\infty$ being obvious). For this, we note that there exists by definition a sequence (M_n) of real numbers such that

$$M_{n+1} \leqslant M_n, \qquad M_n \to M,$$

and M_n satisfies the condition (3.16). We then can write

$$\mu(\{x \mid |f(x)| > M\}) = \mu\Big(\bigcup_n \{x \mid |f(x)| > M_n\}\Big) = 0,$$

as the measure of a countable union of sets of measure zero.

It is also immediate that $\|f\|_\infty = 0$ is equivalent with f being zero μ-almost everywhere, since the definition becomes

$$\mu(\{x \mid f(x) \neq 0\}) = \mu(\{x \mid |f(x)| > 0\}) = 0.$$

Since all the other axioms defining a normed vector space can be checked very easily, there only remains to prove (3.17). However, this is clear by monotony by integrating the upper bound

$$|fg| \leqslant \|g\|_\infty f$$

which, by the above, is valid except on a set of measure zero. $\square$

The space L^∞ is still complete, but the proof of this analogue of the Riesz-Fisher theorem is naturally somewhat different.

PROPOSITION 3.3.15 (Completeness of L^∞). (1) *Let $(X, \mathcal{M}, \mu)$ be a measure space and (f_n) a sequence of measurable functions on X with $f_n \in L^\infty(\mu)$. If*

$$\sum \|f_n\|_\infty < +\infty,$$

the series

$$\sum_{n\geqslant 1} f_n$$

converges μ-almost everywhere, and in L^∞, to function $g \in L^\infty(\mu)$.

(2) *The space $L^\infty(\mu)$ is a complete normed vector space. More precisely, for any Cauchy sequence (f_n) in $L^\infty(\mu)$, there exists $f \in L^\infty(\mu)$ such that $f_n \to f$ in $L^\infty(\mu)$, and the convergence also holds almost everywhere.*

In the last part, note that (in contrast with the other L^p spaces) there is no need to use a subsequence to obtain convergence almost everywhere.

PROOF. (1) The method is the same as the one used before, so we will be brief. Let

$$h(x) = \sum_{n\geqslant 1} |f_n(x)|, \quad g(x) = \sum_{n\geqslant 1} f_n(x)$$

so that the assumption quickly implies that both series converge almost everywhere. Since

$$|g(x)| \leqslant h(x) \leqslant \sum_{n\geqslant 1} \|f_n\|_\infty,$$

almost everywhere, it follows that $g \in L^\infty(\mu)$. Finally

$$\sum f_n = g$$

with convergence in the space L^∞, since

$$\Big\| g - \sum_{n \leqslant N} f_n \Big\|_\infty = \Big\| \sum_{n>N} f_n \Big\|_\infty \leqslant \sum_{n>N} \|f_n\|_\infty \to 0.$$

(2) The proof is in fact more elementary than the one for L^p, $p < \infty$. Indeed, if (f_n) is a Cauchy sequence in $L^\infty(\mu)$, we obtain

$$|f_n(x) - f_m(x)| \leqslant \|f_n - f_m\|_\infty$$

for any fixed n and m, and for almost all $x \in X$; say $A_{n,m}$ is the exceptional set (of measure zero) such that the inequality holds outside $A_{n,m}$; let A be the union over n, m of the $A_{n,m}$. Since this is a countable union, we still have $\mu(A) = 0$.

Now, for all $x \notin A$, the sequence $(f_n(x))$ is a Cauchy sequence in $\mathbf{C}$, and hence it converges, say to an element $f(x) \in \mathbf{C}$. The resulting function f (extended to be zero on A, for instance) is of course measurable. We now check that f is in L^∞.

For this, note that (by the triangle inequality) the sequence $(\|f_n\|_\infty)$ of the norms of f_n is itself a Cauchy sequence in $\mathbf{R}$. Let $M \geqslant 0$ be its limit, and let B_n be the set of x for which $|f_n(x)| > \|f_n\|_\infty$, which has measure zero, Again, the union B of all B_n has measure zero, and so has $A \cup B$.

Now, for $x \notin A \cup B$, we have

$$|f_n(x)| \leqslant \|f_n\|_\infty$$

for all $n \geqslant 1$, and

$$f_n(x) \to f(x).$$

For such x, it follows by letting $n \to +\infty$ that $|f(x)| \leqslant M$, and consequently we have shown that $|f(x)| \leqslant M$ almost everywhere. This gives $\|f\|_\infty \leqslant M < +\infty$.

Finally, we show that (f_n) converges to f in L^∞ (convergence almost everywhere is already established). Fix $\varepsilon > 0$, and then let N be such that

$$\|f_n - f_m\|_\infty < \varepsilon$$

when n, $m \geqslant N$. We obtain $|f_n(x) - f_m(x)| < \varepsilon$ for all $x \notin A \cup B$ and all n, $m \geqslant N$. Taking any $m \geqslant N$, and letting $n \to +\infty$, we get

$$|f(x) - f_m(x)| < \varepsilon \text{ for almost all } x\ x,$$

and this means that $\|f - f_m\|_\infty < \varepsilon$ when $m > N$. Consequently, we have shown that $f_n \to f$ in $L^\infty(\mu)$. ☐

Remark 3.3.16. Usually, there is no obvious relation between the various spaces $L^p(\mu)$, $p \geqslant 1$. For instance, consider $X = \mathbf{R}$ with the Lebesgue measure. The function $f(x) = \inf(1, 1/|x|)$ is in L^2, but not in L^1, whereas $g(x) = x^{-1/2}\chi_{[0,1]}$ is in L^1, but not in L^2. Other examples can be given for any choices of p_1 and p_2: we never have $L^{p_1}(\mathbf{R}) \subset L^{p_2}(\mathbf{R})$.

However, it is true that *if the measure μ is finite* (i.e., if $\mu(X) < +\infty$, for instance if μ is a probability measure), the spaces $L^p(X)$ form a "decreasing" family: we have

$$L^{p_1}(X) \subset L^{p_2}(X)$$

whenever $1 \leqslant p_2 \leqslant p_1 \leqslant +\infty$. In particular, a bounded functions is then in every L^p-space.

More precisely, we have the following, which shows that the inclusion above is also continuous, with respect to the respective norms:

Proposition 3.3.17 (Comparison of L^p spaces for finite measure). *Let $(X,\mathcal{M},\mu)$ be a measured space with $\mu(X)<+\infty$. For any p_1, p_2 with*

$$1\leqslant p_2\leqslant p_1\leqslant +\infty,$$

there is a continuous inclusion map

$$L^{p_1}(\mu)\hookrightarrow L^{p_2}(\mu),$$

and indeed

$$\|f\|_{p_2}\leqslant \mu(X)^{1/p_2-1/p_1}\|f\|_{p_1}.$$

Proof. The last inequality is clearly stronger than the claimed continuity (since it obviously implies that the inclusion maps convergent sequences to convergent sequences), so we need only prove the inequality as stated for $f\geqslant 0$. We use Hölder's inequality for this purpose: for any $p\in[1,+\infty]$, with dual exponent q, we have

$$\int_X f^{p_2}d\mu=\int_X f^{p_2}\cdot 1\ d\mu\leqslant\Bigl(\int_X f^{pp_2}d\mu\Bigr)^{1/p}\Bigl(\int_X d\mu(x)\Bigr)^{1/q}$$

and if we pick $p=p_1/p_2\geqslant 1$, with $q^{-1}=p_2/p_1-1$, we obtain

$$\int_X f^{p_2}d\mu\leqslant \mu(X)^{p_2/p_1-1}\|f\|_{p_1}^{p_2},$$

and taking the $1/p_2$-th power, this gives

$$\|f\|^{p_2}\leqslant \mu(X)^{1/p_1-1/p_2}\|f\|_{p_1},$$

as claimed. □

The next proposition is one of the justifications of the notation L^∞.

Proposition 3.3.18. *If $(X,\mathcal{M},\mu)$ is a measure space with $\mu(X)<+\infty$, we have*

$$\lim_{p\rightarrow+\infty}\|f\|_p=\|f\|_\infty.$$

for any $f\in L^\infty\subset L^p$.

Proof. This is obvious if $f=0$ in some L^p (or equivalently in L^∞). Otherwise, by replacing f by $f/\|f\|_\infty$, we can assume that $\|f\|_\infty=1$. Moreover, changing f on a set of measure zero if needed, we may as well assume that we have a function

$$f\,:\,X\rightarrow\mathbf{C}$$

such that $|f(x)|\leqslant 1$ for all $x\in X$.

The idea is that, although $|f(x)|^p\rightarrow 0$ as $p\rightarrow+\infty$ if $0\leqslant|f(x)|<1$, the function f takes values very close to one quite often, and the power $1/p$ in the norm compensates for the decay. Precisely, fix $\varepsilon>0$ and let

$$Y_\varepsilon=\{x\ \mid\ |f(x)|\geqslant 1-\varepsilon\}.$$

By monotony, we have

$$\int_X |f|^p d\mu\geqslant (1-\varepsilon)^p\mu(Y_\varepsilon)$$

for any $\varepsilon>0$, and hence

$$\|f\|_p\geqslant (1-\varepsilon)\mu(Y_\varepsilon)^{1/p}.$$

Since $\|f\|_\infty=1$, we know that $\mu(Y_\varepsilon)>0$ for any $\varepsilon>0$, and therefore

$$\lim_{p\rightarrow+\infty}\mu(Y_\varepsilon)^{1/p}=1$$

for fixed ε. It follows by taking the liminf that

$$\liminf_{p\to+\infty} \|f\|_p \geqslant 1-\varepsilon,$$

and if we now let $\varepsilon \to 0$, we obtain $\liminf \|f\|_p \geqslant 1$. But of course, under the assumption $\|f\|_\infty \leqslant 1$, we have $\|f\|_p \leqslant 1$ for all p, and hence this gives the desired limit. □

Exercise 3.3.19. Let μ be a finite measure and let $f \in L^\infty(\mu)$ be such that $\|f\|_\infty \leqslant 1$. Show that

$$\lim_{p\to 0} \int_X |f|^p d\mu = \mu(\{x \mid |f(x)| > 0\}).$$

We conclude this section with the statement of a result which is one justification of the importance of L^p spaces. This is a special case of much more general results, and we will prove it later.

Proposition 3.3.20 (Continuous functions are dense in L^p spaces). *Let $X = \mathbf{R}$ with the Lebesgue measure, and let p be such that*

$$1 \leqslant p < +\infty.$$

Then the space $C_c(\mathbf{R})$ of continuous functions on $\mathbf{R}$ which are zero outside a finite interval $[-B, B]$, $B \geqslant 0$, is dense *in $L^p(\mathbf{R})$. In other words, for any $f \in L^p(\mathbf{R})$, there exists a sequence (f_n) of continuous functions with compact support such that*

$$\lim_{n\to+\infty} \Bigl(\int_{\mathbf{R}} |f_n(t) - f(t)|^p dt\Bigr)^{1/p} = 0.$$

Remark 3.3.21. Note that this is definitely false if $p = +\infty$; indeed, for bounded continuous functions, the L^∞ norm corresponds to the "uniform convergence" norm, and hence a limit of continuous functions in L^∞ is always a continuous function.

Remark 3.3.22. Note also that any continuous function with compact support is automatically in L^p for any $p \geqslant 1$, because if $f(x) = 0$ for $|x| \geqslant B$, it follows that f is bounded (say by M) on $\mathbf{R}$, as a continuous function on $[-B, B]$ is, and we have

$$\int_{\mathbf{R}} |f|^p dt = \int_{[-B,B]} |f|^p dt \leqslant 2B\|f\|_\infty^p < +\infty.$$

3.4. Probabilistic examples: the Borel-Cantelli lemma and the law of large numbers

In this section, we present some elementary purely probabilistic applications of integration theory. For this, we fix a probability space (Ω, Σ, P).

The first result is the Borel-Cantelli Lemma. Its purpose is to answer the following type of questions: suppose we have a sequence of events A_n, $n \geqslant 1$, which are independent. How can we determine if there is a positive probability that infinity many A_n "happen"?

More precisely, the corresponding event

$$A = \{\omega \in \Omega \mid \omega \text{ is in infinitely many of the } A_n\}$$

can be expressed set-theoretically as the following combination of unions and intersections:

$$A = \bigcap_{N\geqslant 1} \bigcup_{n\geqslant N} A_n \in \Sigma,$$

and it follows in particular that A is measurable. (To check this formula, define

$$N(\omega) = \sup\{n \geqslant 1 \mid \omega \in A_n\},$$

for $\omega \in \Omega$, so that ω belongs to infinitely many A_n if and only if $N(\omega) = +\infty$; on the other hand, ω belongs to the set on the right-hand side if and only if, for all $N \geqslant 1$, we can find some $n \geqslant N$ such that $\omega \in A_n$, and this is also equivalent with $N(\omega) = +\infty$).

The Borel-Cantelli Lemma gives a quantitative expression of the intuitive idea that the probability is big is, roughly, the A_n are "big enough".

PROPOSITION 3.4.1 (Borel-Cantelli Lemma). *Let A_n, $n \geqslant 1$, be events, and let A be defined as above. Let*

$$p = \sum_{n\geqslant 1} P(A_n) \in [0, +\infty]. \tag{3.18}$$

(1) *If $p < +\infty$, we have $P(A) = 0$, independently of any assumptions on A_n.*
(2) If *the A_n are independent, and $p = +\infty$, then $P(A) = 1$.*

PROOF. The first part is very easy: by monotony, we have

$$P(A) \leqslant P\Bigl(\bigcup_{n\geqslant N} A_n\Bigr) \leqslant \sum_{n\geqslant N} P(A_n),$$

for any $N \geqslant 1$, and the assumption $p < +\infty$ shows that this quantity goes to 0 as $N \to +\infty$, and therefore that $P(A) = 0$.

For (2), one must be more carefuly (because the assumption of independence, or something similar, can not be dispensed with as the example $A_n = A_0$ for all n shows if $P(A_0) < 1$). We notice that $\omega \in A$ if and only if

$$\sum_{n\geqslant 1} \chi_{A_n}(\omega) = +\infty, \text{ if and only if } \exp\Bigl(-\sum_{n\geqslant 1} \chi_{A_n}(\omega)\Bigr) = 0. \tag{3.19}$$

Consider now the integral

$$\int \exp\Bigl(-\sum_{n\leqslant N} \chi_{A_n}\Bigr) dP = \int \prod_{n\leqslant N} \exp(-\chi_{A_n}) dP$$

for some $N \geqslant 1$. Since the A_n, $n \leqslant N$, are independent, we have immediately the relation

$$\int \prod_{n\leqslant N} \exp(-\chi_{A_n}) dP = \prod_{n\leqslant N} \int \exp(-\chi_{A_n}) dP = \prod_{n\leqslant N} (e^{-1}P(A_n) + 1 - P(A_n)).$$

Let $c = 1 - e^{-1} \in]0, 1[$; we obtain now

$$\log \prod_{n\leqslant N} \int \exp(-\chi_{A_n}) dP = \sum_{n\leqslant N} \log(1 - cP(A_n)),$$

and since $\log(1 - x) \leqslant -x$ for $x \geqslant 0$, we derive

$$\log \prod_{n\leqslant N} \int \exp(-\chi_{A_n}) dP \leqslant -c \sum_{n\leqslant N} P(A_n) \to -\infty \text{ as } N \to +\infty,$$

according to the hypothesis $p = +\infty$. Going backwards, this means (after applying the dominated convergence theorem) that

$$\int \exp\Bigl(-\sum_{n\geqslant 1} \chi_{A_n}\Bigr) dP = 0,$$

which implies that (3.19) holds almost surely. ☐

Our second probabilistic result is a simple (non-trivial) version of the (strong) law of large numbers. The situation here is the following: we have a sequence (X_n), $n \geqslant 1$, of random variables which are *independent* and *identically distributed*, i.e., the laws $X_n(P)$ of the variables X_n are all the same, say μ. This implies in particular that, if X_n is integrable (resp. square-integrable), we have

$$E(X_n) = E(X_1) = \int x d\mu \text{ and } V(X_n) = V(X_1) = \int (x - E(X_1))^2 d\mu \text{ for all } n \geqslant 1,$$

so the expectation (and the variance) of the X_n, when it makes sense, is independent of n.

A mathematical model of this situation is given by the example of the digits of the base b expansion of a real number $x \in [0,1] = \Omega$ (see Example 1.3.3, (2)). An intuitive model is that of repeating some experimental measurement arbitrarily many times, in such a way that each experiment behaves independently of the previous ones. A classical example is that of throwing "infinitely many times" the same coin, and checking if it falls on Heads or Tails. In the case, one would take X_n taking only two values $X_n \in \{\mathsf{h}, \mathsf{t}\}$, and in the (usual) case of a "fair" coin, one assumes that the law of X_n is unbiased:

$$P(X_n = \mathsf{h}) = P(X_n = \mathsf{t}) = \frac{1}{2}. \tag{3.20}$$

Of course, the case of laws like

$$P(X_n = \mathsf{p}) = p, \quad P(X_n = \mathsf{f}) = 1 - p$$

for a fixed $p \in]0,1[$ ("biased" coins) is also very interesting.

One should note that, in the case of a fair coin, there is a strong resemblance with the base 2 digits of a real number in $[0,1]$; indeed, in some sense, they are identical mathematically, as one can decide, e.g., that h corresponds to the digit 0, t to the digit 1, and one can construct "random" real numbers by performing the infinite coin-throwing experiment to decide which binary digits to use...

Now, in the general situation we have described, one expects intuitively that, when a large number of identical and independent measurements are made, the "empirical average"

$$\frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$$

should be close to the "true" average, which is simply the common expectation

$$E(X_n) = E(X_1) = \int x d\mu$$

of the X_n (assuming integrability).

This is all the more convincing that, on the one hand, we have

$$E(S_n/n) = E(X_1),$$

by linearity, and on the other hand, by independance (see (3.15)), the variance of S_n/n is

$$V\Big(\frac{S_n}{n}\Big) = \frac{V(S_n)}{n^2} = \frac{1}{n^2}(V(X_1) + \cdots + V(X_n)) = \frac{V(X_1)}{n} \tag{3.21}$$

which therefore converges to 0, something which implies that S_n/n tends to be quite narrowly concentrated around its mean when n is large.

This intuition, often called the "law of large numbers", can be interpreted in various rigorous mathematical senses. We start with the easiest one, the so-called "weak" law of large numbers. This allows us to introduce the notion of convergence in probability.

Definition 3.4.2 (Convergence in probability). Let (Ω, Σ, P) be a probability space, let (X_n), $n \geqslant 1$, be a sequence of random variables on Ω, and X another random variable. Then X_n *converges to* X *in probability* if, for any $\varepsilon > 0$, we have

$$\lim_{n \to +\infty} P(|X_n - X| > \varepsilon) = 0.$$

Remark 3.4.3. More generally, if (f_n) is a sequence of measurable functions on a measure space $(X, \mathcal{M}, \mu)$, we say that (f_n) converges in measure to f if

$$\mu(\{x \in X \mid |f_n(x) - f(x)| > \varepsilon\}) \to 0$$

as $n \to +\infty$, for every fixed $\varepsilon > 0$.

Proposition 3.4.4 (Weak law of large numbers). *Let (Ω, Σ, P) be a probability space, let (X_n), $n \geqslant 1$, be a sequence of independent, identically distributed, random variables, such that $X_n \in L^2$.*

Then S_n/n converges in probability to the constant random variable $E(X_1)$, i.e., we have

$$P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - E(X_1)\Big| > \varepsilon\Big) \to 0$$

as $n \to +\infty$, for any fixed $\varepsilon > 0$.

The proof is very simple, but it uses a very useful tool, the Chebychev–Markov inequality, which is often used to bound from above the probability of some random variable being "far from its average".

Proposition 3.4.5 (Chebychev–Markov inequality). (1) *Let (Ω, Σ, P) be a probability space, and let X be a random variable on Ω. If $X \in L^p(\Omega)$, where $p \geqslant 1$, we have*

$$P(|X| \geqslant \varepsilon) \leqslant \varepsilon^{-p} E(|X|^p) \tag{3.22}$$

for any $\varepsilon > 0$. In particular, if $X \in L^2(\Omega)$, we have

$$P(|X - E(X)| \geqslant \varepsilon) \leqslant \frac{V(X)}{\varepsilon^2}. \tag{3.23}$$

(2) *Let $(X, \mathcal{M}, \mu)$ be a measure space, $p \geqslant 1$ and $f \in L^p(\mu)$. Then we have*

$$\mu(\{x \mid |f(x)| \geqslant \varepsilon\}) \leqslant \varepsilon^{-p} \int_X |f|^p d\mu = \varepsilon^{-p} \|f\|_p^p \tag{3.24}$$

for all $\varepsilon > 0$

Proof. The second part is proved exactly like the first one; for the latter, consider the set

$$A = \{|X| > \varepsilon\},$$

and note that by positivity and monotony, we have

$$E(|X|^p) \geqslant \int_A |X|^p dP \geqslant \varepsilon^p \int_A dP = \varepsilon^p P(A).$$

The special case (3.23) corresponds to replacing X by $X - E(X)$, for which the mean-square is the variance. $\square$

Remark 3.4.6. From this inequality, we see in particular that when $f_n \to f$ in L^p, for some $p \in [1, +\infty[$, the sequence (f_n) also converges to f in measure:

$$\mu(\{x \mid |f_n(x) - f(x)| > \varepsilon\}) \leqslant \varepsilon^{-p} \|f_n - f\|_p^p \to 0.$$

On the other hand, if (f_n) converges to f almost everywhere, it is not always the case that (f_n) converges to f in measure. A counterexample is given by

$$f_n = \chi_{]n,n+1[},$$

a sequence of functions on $\mathbf{R}$ which converges to 0 pointwise, but also satisfies

$$\lambda(\{x \mid |f_n(x)| \geqslant 1\}) = 1$$

for all n.

However, if μ is a finite measure (in particular if it is a probability measure), convergence almost everywhere does imply convergence in measure. To see this, fix $\varepsilon > 0$ and consider the sets

$$A_n = \{x \mid |f_n(x) - f(x)| > \varepsilon\}$$
$$B_k = \bigcup_{n \geqslant k} A_n \qquad B = \bigcap_{k \geqslant 1} B_k.$$

We have $B_{k+1} \subset B_k$ and the assumption implies that $\mu(B) = 0$ (compare with (1.9)). Since $\mu(X) < +\infty$, we have

$$\lim_{k \to +\infty} \mu(B_k) = 0$$

by Proposition 1.2.3, (4), and since $\mu(A_n) \leqslant \mu(B_n)$ by monotony, it follows that $\mu(A_n) \to 0$, which means precisely that (f_n) converges to f in measure.

Proof of Proposition 3.4.4. Since $X_n \in L^2$, we have also $S_n \in L^2$, and we can apply (3.23): for any $\varepsilon > 0$, we get

$$\begin{aligned} P\Big(\Big|\frac{S_n}{n} - E(X_1)|\Big| > \varepsilon\Big) &\leqslant \varepsilon^{-2} E\Big(\Big|\frac{S_n}{n} - E(X_1)\Big|^2\Big) \\ &= \varepsilon^{-2} E\Big(\Big|\frac{S_n}{n} - E\Big(\frac{S_n}{n}\Big)\Big|^2\Big) \\ &= \varepsilon^{-2} V(S_n/n) = \varepsilon^{-2} n^{-1} V(X_1) \to 0 \text{ as } n \to +\infty, \end{aligned}$$

by (3.21). ☐

In general, convergence in probability is a much weaker statement that convergence almost everywhere. For instance, the sequence (f_n) of Exercise 3.3.6 converges in measure to 0 (since $P(|f_n| \geqslant \varepsilon) \leqslant 2^{-k}$ for $n \geqslant 2^k$), but it does not converge almost everywhere. So it is natural to ask what happens in our setting of the law of large numbers. It turns out that, in that case, S_n/n converges almost everywhere to the constant $E(X_1)$ under the only condition that X_n is integrable (this is the "Strong" law of large numbers of Kolmogorov). The proof of this result is somewhat technical, but one can give a much simpler argument under slightly stronger assumptions.

Theorem 3.4.7 (Strong law of large numbers). *Let (X_n) be a sequence of independent, identically distributed random variables with common probability law μ. Assume that X_n is almost surely bounded, or in other words that $X_n \in L^\infty(\Omega)$. Then we have*

$$\frac{X_1 + \cdots + X_n}{n} \to E(X_1) = \int_{\mathbf{C}} x d\mu$$

almost surely.

Proof. In fact, we will prove this under the assumption that $X_n \in L^4$ for all n. Of course, if $|X_n(\omega)| \leqslant M$ for almost all ω, we have $E(|X_n|^4) \leqslant M^4$, so this is really a weaker assumption. For simplicity, we also assume that (X_n) is real-valued; the general case can be dealt with simply by considering the real and imaginary parts separately.

The idea for the proof is similar to that used in the "difficult" part of the Borel-Cantelli Lemma. We consider the series

$$\sum_{n\geqslant 1}\Bigl(\frac{X_1+\cdots+X_n}{n}-E(X_1)\Bigr)^4$$

as a random variable with values in $[0,+\infty]$. If we can show that this series converges almost surely, the desired result will follow: indeed, in that case, we must have

$$\Bigl|\frac{X_1+\cdots+X_n}{n}-E(X_1)\Bigr|\rightarrow 0$$

almost surely. Similarly, to show the convergence almost everywhere, it is enough to show that

(3.25)

$$\int\sum_{n\geqslant 1}\Bigl(\frac{X_1+\cdots+X_n}{n}-E(X_1)\Bigr)^4 dP=\sum_{n\geqslant 1}\int\Bigl(\frac{X_1+\cdots+X_n}{n}-E(X_1)\Bigr)^4 dP<+\infty.$$

We denote $Y_n = X_n - E(X_n) = X_n - E(X_1)$, so that the (Y_n) are independent and have $E(Y_n) = 0$. We have obviously

$$E\Bigl(\frac{X_1+\cdots+X_n}{n}-E(X_1)\Bigr)^4=\frac{1}{n^4}E((Y_1+\cdots+Y_n)^4),$$

and we proceed to expand the fourth power, getting

$$\frac{1}{n^4}E((Y_1+\cdots+Y_n)^4)=\frac{1}{n^4}\sum_{1\leqslant p,q,r,s\leqslant n}E(Y_pY_qY_rY_s).$$

The idea now is that, while each term is bounded (because of the assumption that $X_n \in L^4(\Omega)$, most of them are also, in fact, exactly equal to zero. This leads to an upper bound which is good enough to obtain convergence of the series.

Precisely, we use the property that, for independent integrable random variables X and Y, we have

$$E(XY) = E(X)E(Y)$$

(Exercise 2.3.5), and moreover we use the fact that X^a and Y^b are also independent in that case, for any $a \geqslant 1$, $b \geqslant 1$ (Proposition 1.2.12, (2)). Similarly, Y_pY_q and $\bar{Y}_r\bar{Y}_s$ are independent if $\{p, q\} \cap \{r, s\} = \emptyset$ (Exercise 1.2.13).

So, for instance, we have

$$E(Y_pY_qY_rY_s) = E(Y_p)E(Y_q)E(Y_r)E(Y_s) = 0,$$

if p, q, r and s are distinct. Similarly if $\{p, q, r, s\}$ has three elements, say $r = s$ and no other equality, we have

$$E(Y_pY_qY_rY_s) = E(Y_p)E(Y_q)E(Y_r^2) = 0.$$

This means that, in the sum, there only remain the terms where $\{p, q, r, s\}$ has two elements, which are of the form

$$E(Y_p^2Y_q^2) = E(Y_p^2)E(Y_q^2),$$

and the "diagonal" ones

$$E(Y_p^4).$$

Each of these is bounded easily and uniformly, since the (Y_n) are identically distributed; for instance

$$|E(Y_p^2Y_q^2)| \leqslant \sqrt{E(Y_p^4)E(Y_q^4)} \leqslant E(Y_1^4).$$

The number of non-zero terms is clearly

$$n + 6\frac{n(n-1)}{2} \leqslant 3n^2$$

(the second term counts first the number of two-elements sets in $\{1,\ldots,n\}$, then the number of ways they can occur in the ordered family (p,q,r,s)). Hence we have

$$\frac{1}{n^4}E((Y_1+\cdots+Y_n)^4) \leqslant \frac{3E(Y_1^4)}{n^2},$$

which implies that the series over n converges, as desired. □

EXAMPLE 3.4.8. (1) We start with a somewhat artificial example of the Borel-Cantelli Lemma. Consider the game of Heads or Tails described above, where we assume that the corresponding sequence (X_n) of random variables has common probability distribution given by (3.20) (such distributions are called Bernoulli laws). Now partition the sequence

$$X_1,\ X_2,\ldots,X_n,\ldots,$$

of experiments (or throw of the coin) in successive blocks of length $2k$, where $k \geqslant 1$ increases fixed. Thus the first block is

$$X_1,\ X_2,$$

the second is

$$X_3,\ X_4,\ X_5,\ X_6,$$

and so on.

Now consider the event

$$A_k = \{\text{There appear as many Heads as Tails during the } k\text{-th block of experiments}\}$$

These A_k are independent, because they are defined entirely based on data concerning sets of indices n which are disjoint when k varies. We ask whether, almost surely, infinitely many of these events will occur. Of course, we must apply the Borel-Cantelli Lemma.

What is the probability of A_k? Consider the sequence

$$(Y_1,\ldots,Y_{2k})$$

of random variables (taken from the original (X_n)) which are used to define the k-th block; we have $\omega \in A_k$ if and only if

$$|\{i \leqslant 2k \mid Y_k(\omega) = \mathsf{h}\}| = k.$$

Since each sequence of $2k$ symbols taken from $\{\mathsf{h},\mathsf{t}\}$ occurs with the same probability (by independence and the "fair coin" assumption), we have therefore

$$P(A_k) = 2^{-2k}\binom{2k}{k}.$$

It is easy to bound this from below: since

$$2^{2k} = (1+1)^{2k} = \sum_{i=0}^{2k}\binom{2k}{i}$$

and $\binom{2k}{i} \leqslant \binom{2k}{k}$ for $0 \leqslant i \leqslant 2k$, we get

$$2^{2k} \leqslant 2k\binom{2k}{k}$$

so that

$$P(A_k) \geqslant (2k)^{-1}.$$

We deduce from this that the series

$$\sum_{k\geqslant 1} P(A_k)$$

diverges. By the Borel-Cantelli Lemma, it follows that, almost surely, infinitely many A_k's do occur.

(2) Let $b \geqslant 2$ be a fixed integer, and consider the sequence (X_n) of random variables on $[0,1]$ (with the Lebesgue measure) giving the base b digits of x. The X_n are of course bounded, and hence the Strong Law of Large Numbers (Theorem 3.4.7) is applicable. But consider first a fixed digit

$$d_0 \in \{0, 1, \ldots, b-1\},$$

and let χ be the characteristic function of $d_0 \in \mathbf{R}$. The random variables

$$Y_n = \chi(X_n)$$

take values in $\{0, 1\}$ and are still independent (Proposition 1.2.12), with common distribution law given by

$$P(Y_n = 1) = \frac{1}{b}, \text{ and } P(Y_n = 0) = \frac{b-1}{b}$$

(there are again Bernoulli laws).

In that situation, we have

$$Y_1 + \cdots + Y_n = |\{k \leqslant n \mid X_n = d_0\}|$$

and $E(Y_n) = b^{-1}$. Hence the strong law of large numbers implies that, for almost all $x \in [0,1]$ (with respect to the Lebesgue measure), we have

$$\lim_{n\to+\infty} \frac{|\{k \leqslant n \mid X_n = d_0\}|}{n} = \frac{1}{b}.$$

In concrete terms, for almost all $x \in [0,1]$, the asymptotic proportion of digits of x in base b equal to d_0 always has a limit which is $1/b$: "all digits occur equally often in almost all x". Of course, it is not hard to find examples where this is not true; for instance,

$$x = 0.111111\ldots,$$

or indeed (if $b \geqslant 3$) any element of the Cantor-like set where all digits are among $\{0, b-1\}$.

Moreover, since the intersection of countable many events which are almost sure is still almost sure (the complement being a countable union of sets of measure zero), we can say that almost all x has the stated property *with respect to every base* $b \geqslant 2$. One says that such an $x \in [0,1]$ is *normal in every base* b.

Here also, the complementary set is quite complicated and "big" in a certain intuitive sense. For instance, now all rationals are exceptional! (In a suitable base of the type $b = 10^m$, m large enough, the base b expansion of a rational contains a single digit from a certain point on).

It is in fact quite interesting that presenting a single *explicit* example of a normal number is very difficult (indeed, none is known!), although we have proved rather easily that such numbers exist in overwhelming abundance. One may expect, for instance, that $\pi - 3$ should work, but this is not known at the moment.

There are many questions arising naturally from the case of the Strong Law of Large Numbers we have proved. As already mentioned, one can weaken the assumption on (X_n) to ask only that X_n be integrable (which is a minimal hypothesis if one expects that $S_n/n \to E(X_1)$ almost surely!)

Another very important question is the following: what is the speed of convergence of the sums

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$$

to their limit $E(X_1)$? In other words, assuming $E(X_n) = 0$ (as one can do after replacing X_n by $X_n - E(X_n)$ as in the proof above), the question is: what is the "correct" order of magnitude of S_n/n as $n \to +\infty$? We know that $S_n/n \to 0$ almost surely; is this convergence very fast? Very slow?

We see first from the proof of Theorem 3.4.7 that we have established something stronger than what is stated, and which gives some information in that direction. Indeed, if $E(X_n) = 0$, it follows from the proof that

$$\frac{S_n}{n^\alpha} \to 0 \text{ almost surely}$$

for any $\alpha > 1/2$; indeed, it is only needed that $4\alpha > 2$ to obtain a convergent series

$$\sum_{n \geqslant 1} \frac{1}{n^{4\alpha} E((Y_1 + \cdots + Y_n)^4)}.$$

Therefore, almost surely, S_n is of order of magnitude smaller than n^α for any $\alpha > 1/2$. This result turns out to be close to the truth: the "correct" order of magnitude of S_n is $\sqrt{n}$. The precise meaning of this is quite subtle: it is not the case that $S_n/\sqrt{n}$ goes to 0, or to any other fixed value, but only that $S_n/\sqrt{n}$, as a real random variable, behaves according to a well-defined distribution when n gets large.

Theorem 3.4.9 (Fundamental Limit Theorem). *Let (X_n) be a sequence of real-valued, independent, identically distributed random variables, such that $X_n \in L^2(\Omega)$ for all n and with $E(X_n) = 0$. Let $\sigma^2 = V(X_n)$, $\sigma > 0$, be the common variance of the (X_n).*

Then, for any $a \in \mathbf{R}$, we have

$$P\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \leqslant a\Big) \to \frac{1}{\sqrt{2\pi\sigma^2}} \int_{[-\infty, a]} e^{-\frac{t^2}{2\sigma^2}} dt$$

as $n \to +\infty$.

We will give a proof of this result in the last Section of this book.

Remark 3.4.10. The classical terminology is "Central Limit Theorem"; this has led to some misunderstanding, where "central" was taken to mean "centered" (average zero), instead of "fundamental", as was the intended meaning when G. Pólya, from Zürich, coined the term ("Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem", Math. Zeitschrift VIII, 1920).

Remark 3.4.11. The probability measure μ_{0,σ^2} on $\mathbf{R}$ given by

$$\mu_{0,\sigma^2} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{t^2}{2\sigma}} d\lambda$$

(for $\sigma > 0$) is called the *centered normal distribution with variance* σ^2, or the *centered gaussian law with variance* σ^2. It is not obvious that $\int d\mu_{0,\sigma} = 1$, but we will prove this in the next chapter.

The type of convergence given by this theorem is called *convergence in law.* This is a weaker notion than convergence in probability, or convergence almost everywhere (in the situation of the fundamental limit theorem, one can show, that $(S_n - nE(X_1))/\sqrt{n}$ never converges in probability). We will come back to this notion of convergence in Section 5.7.

CHAPTER 4

Measure and integration on product spaces

Among the problems of Riemann's definition of integrals that we mentioned in the introduction, we can now claim that three are satisfactorily solved: exchanging limits and integrals can be handled very efficiently using the monotone and dominated convergence theorems, integrating over bounded or unbounded intervals (with respect to Lebesgue measure) is done using a unified process, and we have a robust notion of "probability". One important point still remains, however: integrating over subsets of $\mathbf{R}^m$, and in particular questions concerning iterated integrals of functions of more than one variable. Since the theory is developed in full generality, this can be thought of as only a special case of a problem concerning the definition of suitable integrals on a product space $X \times Y$, when measures on X and Y are given.

In this chapter, we present the solution of this problem, which is quite elegant...

4.1. Product measures

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two measured spaces. The product space $X \times Y$ can be equipped with the product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ generated by the *(measurable) rectangles* $A \times B$ where $(A, B) \in \mathcal{M} \times \mathcal{N}$ (see Definition 1.1.6).

The motivating goal of this section is to find conditions that ensure that the formula of "exchange of order of integration" is valid, or in other words, so that we have

$$\int_X \Bigl(\int_Y f(x, y) d\nu(y) \Bigr) d\mu(x) = \int_Y \Bigl(\int_X f(x, y) d\mu(x) \Bigr) d\nu(y). \tag{4.1}$$

However, some assumptions on the measures μ and ν are required for this to be valid.

Example 4.1.1. Let $X = [0, 1]$ with the Borel σ-algebra and the Lebesgue measure μ, and let $Y = [0, 1]$, but with the counting measure (on the maximal σ-algebra consisting of all subsets of Y).

Consider the set $D = \{(x, x) \mid x \in [0, 1]\} \subset X \times Y$, i.e., the diagonal of the square $X \times Y \subset \mathbf{R}^2$. This set is measurable for the product σ-algebra (because it is closed in the plane, hence measurable for the Borel σ-algebra of $\mathbf{R}^2$, and we have seen in Remark 1.1.7, (3), that this is the same as the product σ-algebra of the Borel σ-algebra on $[0, 1]$ on each factor). Now take f to be the characteristic function of D; we claim that (4.1) fails.

Indeed, on the left-hand side, the inner integral is $\nu(\{(x, x)\}] = 1$ for all x, hence this side is equal to 1. On the right-hand side, however, the inner integral is $\mu(\{(y, y)\}) = 0$ for all y, so that side is 0...

The suitable assumptions will be that μ and ν be σ-finite measures; recall (see Definition 1.2.1) that a measure μ on X is σ-finite if there exists a sequence (X_n) of measurable sets in X, each with finite measure $\mu(X_n) < +\infty$, and with union equal to X. For instance, the Lebesgue measure is σ-finite (and so is any probability measure) but the counting measure on an uncountable set is not (this is the problem in the example above). Under this assumption, we will in fact define a measure (called the *product measure*) $\mu \times \nu$

on $(X \times Y, \mathcal{M} \otimes \mathcal{N})$ in such a way that, for $f : X \times Y \to \mathbf{C}$ integrable, each of the two expressions above are equal to the integral of the two-variable function $f(x, y)$ with respect to $\mu \times \nu$: we have

$$\int_X \Big(\int_Y f(x,y) d\nu(y) \Big) d\mu(x) = \int_{X \times Y} f(x,y) d(\mu \times \nu)(x,y) \tag{4.2}$$
$$= \int_Y \Big(\int_X f(x,y) d\mu(x) \Big) d\nu(y).$$

If we apply this expected formula to the characteristic function f of a measurable set $C \in \mathcal{M} \otimes \mathcal{N}$, we see that we should have the formulas

$$(\mu \otimes \nu)(C) = \int_X \nu(t_x(C)) d\mu(x) = \int_Y \mu(t^y(C)) d\nu(y) \tag{4.3}$$

where

$$t_x(C) = \{y \mid (x,y) \in C\} = C \cap (\{x\} \times Y) \subset Y \tag{4.4}$$
$$t^y(C) = \{x \mid (x,y) \in C\} = C \cap (X \times \{y\}) \subset X \tag{4.5}$$

are the horizontal and vertical "slices" of C; indeed, we have the expressions

$$\int_Y f(x,y) d\nu(y) = \nu(t_x(C)) \text{ for all } x \in X$$
$$\int_X f(x,y) d\mu(x) = \mu(t^y(C)) \text{ for all } y \in Y,$$

for the integrals of $f = \chi_C$ when x or y is fixed.

Thus (4.3) gives, a priori, two definitions of the desired measure. The main part of the work is now to make sure that these actually make sense: it is not clear, a priori, that the non-negative functions

$$x \mapsto \nu(t_x(C)), \quad y \mapsto \mu(t^y(C)),$$

are measurable with respect to $\mathcal{M}$ and $\mathcal{N}$, respectively. Even once this is known, it is not obvious (of course) that the two definitions coincide!

The basic reason one may hope for this strategy to work is that the desired properties can be easily checked directly in the special (initial) case where $C = A \times B$ is a rectangle, and we know that, by definition, these rectangles generate the whole σ-algebra $\mathcal{M} \otimes \mathcal{N}$. Indeed we have

$$\int_X \int_Y \chi_C(x,y) d\nu(y) d\mu(x) = \int_Y \int_X \chi_C(x,y) d\mu(x) d\nu(y) = \mu(A)\nu(B)$$

for a rectangle, where the inner integrals are constant in each case, and therefore measurable. However, constructing measures from generating sets is not a simple algebraic process (as the counter-example above indicates), so one must be careful.

We first state some elementary properties of the "slicing" operations, which may be checked directly, or by applying (1.3) after noticing that

$$t_x(C) = i_x^{-1}(C), \qquad t^y(C) = j_y^{-1}(C)$$

where $i_x : Y \to X \times Y$ is the map given by $y \mapsto (x, y)$ and $j_y : X \to X \times Y$ is given by $x \mapsto (x, y)$.

LEMMA 4.1.2. *For any fixed $x \in X$, we have*

$$t_x((X \times Y) - C) = Y - t_x(C) \tag{4.6}$$

$$t_x\Big(\bigcup_{i\in I} C_i\Big) = \bigcup_{i\in I} t_x(C_i) \tag{4.7}$$

$$t_x\Big(\bigcap_{i\in I} C_i\Big) = \bigcap_{i\in I} t_x(C_i). \tag{4.8}$$

We can now state the first main result.

PROPOSITION 4.1.3. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measured spaces.*

(1) *For any $C \in \mathcal{M} \otimes \mathcal{N}$, all the slices $t_x(C) \subset Y$, $x \in X$, belong to the σ-algebra $\mathcal{N}$. In other words, for any fixed x, the map $i_x : Y \to X \times Y$ is mesurable, the same holds true for j_y.*

(2) *If we* assume also *that Y is σ-finite, the non-negative function*

$$x \mapsto \nu(t_x(C))$$

is $\mathcal{M}$-measurable for all $C \in \mathcal{M} \otimes \mathcal{N}$.

(3) *Again if Y is σ-finite, the map*

$$\pi \begin{cases} \mathcal{M} \otimes \mathcal{N} \to [0, +\infty] \\ C \mapsto \displaystyle\int_X \nu(t_x(C)) d\mu(x) \end{cases} \tag{4.9}$$

is a measure on $\mathcal{M} \otimes \mathcal{N}$.

PROOF. If we assume that (1) and (2) are proved, it is easy to check that (3) holds. Indeed, (2) ensures first that the definition makes sense. Then, we have $\pi(\emptyset) = 0$, obviously, and if (C_n), $n \geqslant 1$, is a countable family of disjoint sets in $\mathcal{M} \otimes \mathcal{N}$, with union C, we obtain

$$\nu(t_x(C)) = \sum_{n\geqslant 1} \nu(t_x(C_n))$$

for all x (by the lemma above), and hence by the monotone convergence theorem, we have

$$\pi(C) = \int_X \sum_{n\geqslant 1} \nu(t_x(C_n)) d\mu(x) = \sum_{n\geqslant 1} \int_X \nu(t_x(C_n)) d\mu(x) = \sum_{n\geqslant 1} \pi(C_n).$$

Part (1) is also easy: let x be fixed, so that we must show that i_x is measurable, i.e., that $i_x^{-1}(C) \in \mathcal{N}$ for all $C \in \mathcal{M} \otimes \mathcal{N}$. From Lemma 1.1.9, it is enough to verify that $i_x^{-1}(C) \in \mathcal{N}$ if $C = A \times B$ is a measurable rectangle. But we have

$$i_x^{-1}(C) = t_x(C) = \begin{cases} \emptyset & \text{if } x \notin A, \\ B & \text{if } x \in A, \end{cases} \tag{4.10}$$

and all of these sets belong of course to $\mathcal{N}$.

Part (2) requires a bit more work. The idea (compare with the proof of Lemma 1.1.9) is to consider the collection of all subsets $C \in \mathcal{M} \otimes \mathcal{N}$ which satisfy the required conclusion, and to prove that it is a σ-algebra containing the rectangles, and hence also the generated σ-algebra $\mathcal{M} \otimes \mathcal{N}$. Therefore, let

$$\mathcal{O} = \{C \in \mathcal{M} \otimes \mathcal{N} \mid x \mapsto \nu(t_x(C)) \text{ is } \mathcal{M}\text{-measurable}\}.$$

First of all, rectangles $C = A \times B$ are in $\mathcal{O}$ according to (4.10) since

$$\nu(t_x(C)) = \begin{cases} 0 & \text{if } x \notin A \\ \nu(B) & \text{if } x \in A, \end{cases}$$

which shows that $x \mapsto \nu(t_x(A \times B))$ is even a step function in that case.

Now we first consider the case where ν is a finite measure; the reason is that, otherwise, we can not even easily prove that $\mathcal{O}$ is stable under complement (because if $C \in \mathcal{O}$ and $D = X \times Y - C$ is the complement, the measure of $t_x(D) = Y - t_x(C)$ can possibly not be computed if $\nu(Y) = \nu(t_x(C)) = +\infty$.)

First step: Assume that $\nu(Y) < +\infty$. We obtain therefore $\nu(t_x(D)) = \nu(Y) - \nu(t_x(C))$, which is again measurable if $\nu(t_x(C))$ is, and we deduce that $\mathcal{O}$ is stable by complement. However, the other properties of a σ-algebra are not so easy to derive. What is fairly simple is to prove the following, where (1) and (2) are already known:

(1) $\mathcal{O}$ contains the rectangles;
(2) $\mathcal{O}$ is stable by complement;
(3) $\mathcal{O}$ is stable under countable *disjoint* union;
(4) $\mathcal{O}$ is stable under increasing (resp. decreasing) countable union (resp. intersection).

Indeed, (3) follows immediately from the formula

$$\nu\Big(t_x\Big(\bigcup_{n\geqslant 1} C_n\Big)\Big) = \sum_{n\geqslant 1} \nu(t_x(C_n)),$$

valid if the C_n are disjoint, and the measurability of a pointwise limit of measurable functions. Similarly, (4) comes from

$$\nu\Big(t_x\Big(\bigcup_{n\geqslant 1} C_n\Big)\Big) = \lim_{n\to+\infty} \nu(t_x(C_n))$$

valid if

$$C_1 \subset C_2 \subset \ldots \subset C_n \ldots,$$

the case of decreasing intersections being obtained from this by taking complements and using the assumption $\nu(Y) < +\infty$.

In particular, using the terminology of Lemma 4.1.5 below, we see that $\mathcal{O}$ is a *monotone class* that contains the collection $\mathcal{E}$ of finite union of (measurable) rectangles – for this last purpose, we use the fact that any finite union of rectangles may, using the various formulas below, be expressed as a finite disjoint union of rectangles. This last collection is an *algebra of sets* on $X \times Y$, meaning that it contains $\emptyset$ and $X \times Y$ and is stable under the operations of taking the complement, and of taking finite unions and intersections. Indeed, $C_1 = A_1 \times B_1$ and $C_2 = A_2 \times B_2$ are rectangles, we have

$$C_1 \cap C_2 = (A_1 \cap A_2) \times (B_1 \cap B_2) \in \mathcal{E} \tag{4.11}$$

$$\begin{aligned} X \times Y - C_1 = ((X - A_1) \times (Y - B_1)) \cup ((X - A_1) \times B_1) \\ \cup (A_1 \times (Y - B_1)) \in \mathcal{E} \end{aligned} \tag{4.12}$$

$$C_1 \cup C_2 = (X \times Y) - \{(X \times Y - C_1) \cap (X \times Y - C_2)\} \in \mathcal{E}, \tag{4.13}$$

(these are easier to understand after drawing pictures), and the monotone class lemma, applied to $\mathcal{O}$ with $\mathcal{A} = \mathcal{E}$, shows that

$$\mathcal{O} \supset \sigma(\mathcal{E}) = \mathcal{M} \otimes \mathcal{N},$$

as expected.

Second step: We now assume only that Y is a σ-finite measure space, and fix a sequence (Y_n) of sets in $\mathcal{N}$ such that

$$Y = \bigcup_{n\geqslant 1} Y_n \text{ and } \nu(Y_n) < +\infty.$$

We may assume that the Y_n are disjoint. Then, for any $C \in \mathcal{M}\otimes\mathcal{N}$ and any $x \in X$, we have a disjoint union

$$t_x(C) = \bigcup_{n\geqslant 1} t_x(C\cap Y_n) \text{ hence } \nu(t_x(C)) = \sum_{n\geqslant 1} \nu(t_x(C\cap Y_n)).$$

The first step, applied to the space $X \times Y_n$ (with the measure on Y_n induced by ν, which is finite) instead of $X \times Y$, shows that each function $x \mapsto \nu(t_x(C\cap Y_n))$ is measurable, and it follows therefore that the function $x \mapsto \nu(t_x(C))$ is measurable. $\square$

COROLLARY 4.1.4. *With the same assumptions as in the proposition, let $f : X\times Y \to$ **C** be measurable with respect to the product σ-algebra. Then, for any fixed $x \in X$, the map $t_x(f) : y \mapsto f(x,y)$ is measurable with respect to $\mathcal{N}$.*

PROOF. Indeed, we have $t_x(f) = f \circ i_x$, hence it is measurable as a composite of measurable functions. $\square$

We now state and prove the technical lemma used in the previous proof.

LEMMA 4.1.5 (Monotone class theorem). *Let X be a set, $\mathcal{O}$ a collection of subsets of X such that:*

(1) *If $A_n \in \mathcal{O}$ for all n and $A_n \subset A_{n+1}$ for all n, we have $\bigcup A_n \in \mathcal{O}$, i.e., $\mathcal{O}$ is stable under increasing countable union;*

(2) *If $A_n \in \mathcal{O}$ for all n and $A_n \supset A_{n+1}$ for all n, we have $\bigcap A_n \in \mathcal{O}$, i.e., $\mathcal{O}$ is stable under decreasing countable intersection.*

Such a collection of sets is called a monotone class *in X. If $\mathcal{O} \supset \mathcal{A}$, where $\mathcal{A}$ is an algebra of sets, i.e., $\mathcal{A}$ is stable under complement and finite intersections and unions, then we have*

$$\mathcal{O} \supset \sigma(\mathcal{A}).$$

PROOF. Let first $\mathcal{O}' \subset \mathcal{O}$ denote the intersection of all monotone classes that contain $\mathcal{A}$; it is immediate that this is also a monotone class. It suffices to show that $\mathcal{O}' \supset \sigma(\mathcal{A})$, and we observe that it is then enough to check that $\mathcal{O}'$ is itself a set algebra. Indeed, under this assumption, we can use the following trick

$$\bigcup_{n\geqslant 1} C_n = \bigcup_{N\geqslant 1}\Bigl(\bigcup_{1\leqslant n\leqslant N} C_n\Bigr) \in \mathcal{O}'$$

to reduce an arbitrary countable union of sets $C_n \in \mathcal{O}'$ into a countable *increasing union*, and deduce from the algebra property and monotonicity that $\mathcal{O}'$ is stable under countable unions. As it is also stable under complements, it will be a σ-algebra containing $\mathcal{A}$, hence containing $\sigma(\mathcal{A})$.

To check the three stability properties required of an algebra of sets is not very difficult, but somewhat tedious. First of all, let

$$\mathcal{G} = \{A \subset X \mid X - A \in \mathcal{O}'\}$$

be the collection of complements of sets in $\mathcal{O}'$. Because taking complements transforms increasing unions into decreasing intersections, it is very easy to see that $\mathcal{G}$ is a monotone class, and it contains $\mathcal{A}$ since the latter is a set algebra. Hence, by definition, we have $\mathcal{G} \supset \mathcal{O}'$, which means precisely that $\mathcal{O}'$ is stable under complement.

Using induction, it only remains to show that for A, $B \in \mathcal{O}'$, we have $A \cup B \in \mathcal{O}'$. The trick is to conclude in two steps. First, let

$$\mathcal{G}_1 = \{A \subset X \mid A \cup B \in \mathcal{O}' \text{ for all } B \in \mathcal{A}\},$$

the collection of sets which can be "added" to any set in $\mathcal{A}$ to obtain a union in $\mathcal{O}'$. We note first that

$$\mathcal{G}_1 \supset \mathcal{A}$$

(again because $\mathcal{A}$ itself is stable under finite unions). Using the formulas

$$\Big(\bigcup_{n \geqslant 1} A_n\Big) \cup B = \bigcup_{n \geqslant 1} (A_n \cup B)$$

$$\Big(\bigcap_{n \geqslant 1} A_n\Big) \cup B = \bigcap_{n \geqslant 1} (A_n \cup B),$$

we also see that $\mathcal{G}_1$ is a monotone class. Hence, once more, we have $\mathcal{G}_1 \supset \mathcal{O}'$, and this means that $\mathcal{O}'$ is stable under union *with a set in* $\mathcal{A}$.

Finally, let

$$\mathcal{G}_2 = \{A \subset X \mid A \cup B \in \mathcal{O}' \text{ for all } B \in \mathcal{O}'\}.$$

The preceeding step shows now that $\mathcal{G}_2 \supset \mathcal{A}$; again, $\mathcal{G}_2$ is a monotone class (the same formulas as above are ad-hoc); consequently, we get

$$\mathcal{G}_2 \supset \mathcal{O}',$$

which was the desired conclusion. $\square$

Now, assuming that *both* μ and ν are σ-finite measures, we can also apply Proposition 4.1.3 after exchanging the role of X and Y. It follows that $t^y(C) \in \mathcal{M}$ for all y, that the map $y \mapsto \mu(t^y(C))$ is measurable and that

$$C \mapsto \int_Y \mu(t^y(C)) d\nu(y) \tag{4.14}$$

is a measure on $X \times Y$. Not surprisingly, this is the same as the other one, since we can recall the earlier remark that

$$\int_X \nu(t_x(C)) d\mu(x) = \mu(A)\nu(B) = \int_Y \mu(t^y(C)) d\nu(y) \tag{4.15}$$

holds for all rectangles $C = A \times B$.

Proposition 4.1.6 (Existence of product measure). *Let* $(X, \mathcal{M})$ *and* $(Y, \mathcal{N})$ *be* σ*-finite measurables spaces. The measures on* $\mathcal{M} \otimes \mathcal{N}$ *defined by* (4.9) *and* (4.14) *are equal.*

This common measure is called the product measure *of* μ *and* ν*, denoted* $\mu \otimes \nu$.

We use a simple lemma to deduce this from the equality on rectangles.

Lemma 4.1.7. *Let* $(X, \mathcal{M})$ *and* $(Y, \mathcal{N})$ *be measurable spaces. If* μ_1 *and* μ_2 *are both measures on* $(X \times Y, \mathcal{M} \otimes \mathcal{N})$ *such that*

$$\mu_1(C) = \mu_2(C)$$

for any measurable rectangle $C = A \times B$*, and if there exists a sequence of disjoint measurable rectangles* $C_n \times D_n$ *with the property that*

$$X \times Y = \bigcup_{n \geqslant 1} (C_n \times D_n) \text{ and } \mu_i(C_n \times D_n) < +\infty,$$

then $\mu_1 = \mu_2$.

Proof. This is similar to the argument used to prove Part (2) of the previous proposition, and we also first consider the case where $\mu_1(X \times Y) = \mu_2(X \times Y) < +\infty$.

Let then

$$\mathcal{O} = \{C \in \mathcal{M} \otimes \mathcal{N} \mid \mu_1(C) = \mu_2(C)\},$$

a collection of subsets of C which contains the measurable rectangles by assumption. By (4.11), (4.12) and (4.13), we see also that $\mathcal{O}$ contains the set algebra of finite unions of measurable rectangles. We now check that it is a monotone class: if C_n is an increasing sequence of elements of $\mathcal{O}$, the continuity of measure leads to

$$\mu_1\Big(\bigcup_{n\geqslant 1} C_n\Big) = \lim_{n\to+\infty} \mu_1(C_n) = \lim_{n\to+\infty} \mu_2(C_n) = \mu_2\Big(\bigcup_{n\geqslant 1} C_n\Big),$$

and similarly for decreasing intersections (note that this is where we use $\mu_i(C_1) \leqslant \mu_i(X \times Y) < +\infty$.)

Thus $\mathcal{O}$ is a monotone class containing $\mathcal{E}$, so that by Lemma 4.1.5, we have $\mathcal{O} \supset \mathcal{M}\otimes\mathcal{N}$, which gives the equality of μ_1 and μ_2 in the case at hand.

More generally, the assumption states that we can write

$$X \times Y = \bigcup (C_n \times D_n),$$

where the sets $C_n \times D_n \in \mathcal{M} \otimes \mathcal{N}$ are disjoint and have finite measure (under μ_1 and μ_2). Let $C \in \mathcal{M} \otimes \mathcal{N}$; we have a disjoint union

$$C = \bigcup_{n\geqslant 1} (C \cap (C_n \times D_n))$$

and therefore, by additivity of measures, we derive

$$\mu_1(C) = \sum_n \mu_1(C \cap (C_n \times D_n)) = \sum_n \mu_2(C \cap (C_n \times D_n)) = \mu_2(C),$$

since the first case applies to μ_1 and μ_2 restricted to $C_n \times D_n$ for all n. ☐

Proof of Proposition 4.1.6. This is an immediate application of the lemma: let

$$\mu_1(C) = \int_X \nu(t_x(C))d\mu(x)$$

$$\mu_2(C) = \int_Y \nu(t^y(C))d\nu(y)$$

for $C \in \mathcal{M} \otimes \mathcal{N}$. According to Proposition 4.1.3, these are measures, and

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

for measurable rectangles. The σ-finiteness assumption implies that one can write $X \times Y$ as a disjoint union of rectangles with finite measure, and hence the lemma applies. ☐

Remark 4.1.8. It is useful to remark that the product measure $\mu \otimes \nu$ is also σ-finite, since we can write

$$X \times Y = \bigcup_{n,m} (X_n \times Y_m)$$

with

$$(\mu \otimes \nu)(X_n \times Y_m) = \mu(X_n)\nu(Y_m) < +\infty$$

if X_n (resp. Y_m) themselves form a decomposition of X (resp. Y) into a union of sets with finite measure.

It is therefore possible to construct, by induction, product measures of the type

$$\mu_1 \otimes \cdots \otimes \mu_n,$$

where μ_i is a σ-finite measure on X_i for all i. Lemma 4.1.7 shows that this operation is associative

$$\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$$

(because both measures coincide on rectangles).

For any function f defined on $X \times Y$, we write

$$\int_{X\times Y} fd(\mu \otimes \nu) = \int_{X\times Y} f(x,y)d\mu(x)d\nu(y) = \int\int fd\mu d\nu$$

for the integral with respect to the product measure.

4.2. Application to random variables

Let (Ω, Σ, P) be a probability space. Since the measure P is finite, the product measures $P^{\otimes n} = P \otimes \cdots \otimes P$ on Ω^n may be constructed, for all n, and these are all probability measures again.

Consider now random variables X and Y on Ω; the vector $Z = (X, Y)$ is a measurable map

$$(X,Y) \,:\, \Omega \to \mathbf{C}^2,$$

with respect to the σ-algebra $\mathcal{B} \otimes \mathcal{B} = \mathcal{B}_{\mathbf{C}^2}$. Now let $\mu = X(P)$ denote the law of X, $\nu = Y(P)$ that of Y. By definition, the measure $Z(P)$ on $\mathbf{C}^2$ (called the *joint law* of X and Y) satisfies

$$Z(P)(C) = P(Z^{-1}(C)) = P(X \in A \text{ and } Y \in B)$$

when $C = A \times B$ is a measurable rectangle in $\mathbf{C}^2$. The product measure $\mu \otimes \nu$, on the other hand, satisfies

$$(\mu \otimes \nu)(C) = \mu(A)\nu(B) = P(X \in A)P(Y \in B).$$

Comparing these, according to Proposition 4.1.6, we deduce immediately the following very useful characterization of independent random variables:

Lemma 4.2.1. *Let (Ω, Σ, P) be a probability space.*

(1) *Two random variables X and Y are independent if and only if*

$$(X,Y)(P) = X(P) \otimes Y(P),$$

i.e., if their joint law is the product of the laws of the variables.

(2) *More generally, a family $(X_i)_{i\in I}$ of random variables is independent if and only if, for any $n \geqslant 1$, and any distinct indices i_1, ..., i_n, we have*

$$(X_{i_1}, \ldots, X_{i_n})(P) = X_{i_1}(P) \otimes \cdots \otimes X_{i_n}(P).$$

This property is very handy. For instance, using it, one can quickly prove properties like that of Proposition 1.2.12 and Exercise 1.2.13. As an example of the last case, consider four random variables (X_i), $1 \leqslant i \leqslant 4$, which are independent, and two measurable maps

$$\varphi_1 \,:\, \mathbf{C}^2 \to \mathbf{C}, \qquad \varphi_2 \,:\, \mathbf{C}^2 \to \mathbf{C}.$$

Let

$$Y = \varphi_1(X_1, X_2), \qquad Z = \varphi_2(X_3, X_4).$$

We want to show that Y and Z are independent. For this, let $\psi = (\varphi_1, \varphi_2) : \mathbf{C}^4 \to \mathbf{C}$, and denote by $\mu_i = X_i(P)$ the laws of the X_i. By (1.11) and independence, we have

$$(Y, Z)(P) = \psi(X_1, X_2, X_3, X_4)(P) = \psi_*(\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \mu_4).$$

Now we use the following lemma:

LEMMA 4.2.2. (1) *Let* $(X_i, \mathcal{M}_i, \mu_i)$, *for* $i = 1$, $i = 2$, *be measured spaces with* $\mu_i(X_i) < +\infty$, *let* $(Y_i, \mathcal{N}_i)$, $i = 1, 2$, *be measurable spaces, and*

$$\psi = (\psi_1, \psi_2) : X_1 \times X_2 \to Y_1 \times Y_2$$

a measurable map with respect to the relevant product σ*-algebras. We then have*

$$\psi_*(\mu_1 \otimes \mu_2) = \psi_{1,*}(\mu_1) \otimes \psi_{2,*}(\mu_2)$$

on $Y_1 \times Y_2$.

(2) *Let* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *be probability spaces and* $p : X \times Y \to X$ *the first projection* $(x, y) \mapsto x$. *Then we have*

$$p_*(\mu \otimes \nu) = \mu.$$

PROOF. (1) For any $C = A \times B \in \mathcal{N}_1 \otimes \mathcal{N}_2$, we use the definitions to write

$$\begin{aligned}
\psi_*(\mu_1 \otimes \mu_2)(C) &= (\mu_1 \otimes \mu_2)(\psi^{-1}(C)) \\
&= (\mu_1 \otimes \mu_2)(\psi_1^{-1}(A) \times \psi_2^{-1}(B)) \\
&= \mu_1(\psi_1^{-1}(A))\mu_2(\psi_2^{-1}(B)) \\
&= \psi_{1,*}(\mu_1)(A)\psi_{2,*}(\mu_2)(B) \\
&= (\psi_{1,*}(\mu_1) \otimes \psi_{2,*}(\mu_2))(A \times B),
\end{aligned}$$

and hence the result follows from Proposition 4.1.6 since

$$\psi_*(\mu_1 \otimes \mu_2)(Y_1 \times Y_2) = \mu_1(X_1)\mu_2(X_2) < +\infty.$$

(2) Similarly, we have

$$p_*(\mu \otimes \nu)(C) = (\mu \otimes \nu)(p^{-1}(C)) = (\mu \otimes \nu)(C \times Y) = \mu(C)\nu(Y) = \mu(C)$$

for all $C \in \mathcal{M}$. $\square$

From the first part of this lemma, in the situation above, we obtain

$$\begin{aligned}
(Y, Z)(P) = \psi_*(\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \mu_4) &= \psi_*((\mu_1 \otimes \mu_2) \otimes (\mu_3 \otimes \mu_4)) \\
&= \varphi_{1,*}(\mu_1 \otimes \mu_2) \otimes \varphi_{2,*}(\mu_3 \otimes \mu_4) \\
&= \varphi_{1,*}(X_1, X_2)(P) \otimes \varphi_{2,*}(X_3, X_4)(P) \\
&= Y(P) \otimes Z(P),
\end{aligned}$$

and hence Y and Z are independent by Lemma 4.2.1.

Another important use of product measures, in probability, is that they can be used easily to produce independent random variables with laws arbitrarily specified.

PROPOSITION 4.2.3 (Existence of independent random variables). *Let* $n \geqslant 1$ *be an integer, and let* μ_i, $1 \leqslant i \leqslant n$, *be probability measures on* $(\mathbf{C}, \mathcal{B})$. *There exists a probability space* (Ω, Σ, P) *and random variables* $X_1, \ldots, X_n$ *on* Ω *such that the* (X_i) *are independent, and the law of* X_i *is* μ_i *for all* i.

Proof. Let $(\Omega_i, \Sigma_i, P_i)$ be any probability space together with a random variable Y_i such that $Y_i(P_i) = \mu_i$ (for instance, one can take $(\Omega_i, \Sigma_i, P_i) = (\mathbf{C}, \mathcal{B}, \mu_i)$ with $Y_i(z) = z$.)

We now take

$$\Omega = \Omega_1 \times \cdots \times \Omega_n,$$

with the product σ-algebra Σ, and with the product probability measure

$$P = P_1 \otimes \cdots \otimes P_n.$$

Define $X_i = Y_i(p_i)$, where $p_i : \Omega \to \Omega_i$ is the i-th projection map. Then the random variables (X_i) have the required properties.

Indeed, by Lemma 4.2.2, (2), we have $X_i(P) = Y_i(P_i) = \mu_i$, and by Lemma 4.2.2, (1), we obtain

$$(X_1, \ldots, X_n)(P) = (X_1, \ldots, X_n)(P_1 \otimes \cdots \otimes P_n) = X_1(P) \otimes \cdots \otimes X_n(P)$$

so that the (X_i) are independent. $\square$

This can be applied in particular with $\mu_i = \mu$ for all i, and hence can be used to produce arbitrarily long vectors $(X_1, \ldots, X_n)$ where the components are independent, and have the same law.

The law(s) of large numbers of the previous chapter suggest that it would be also interesting to prove a result of this type for an infinite family of measures. This is indeed possible, but we will only prove this in a later chapter.

However, even the simplest examples can have interesting consequences, as we illustrate:

Theorem 4.2.4 (Bernstein polynomials). *Let $f : [0,1] \to \mathbf{R}$ be a continuous function, and for $n \geqslant 1$, let B_n be the polynomial*

$$B_n = \sum_{k=0}^{n} \binom{n}{k} f\Big(\frac{k}{n}\Big) X^k (1-X)^{n-k} \in \mathbf{R}[X].$$

Then (B_n) converges to f uniformly *on $[0,1]$.*

Proof. Our proof will be based on a probabilistic interpretation of B_n. Let $x \in [0,1]$ be fixed, and let $(X_1, \ldots, X_n)$ be a vector of random variables taking values in $\{0,1\}$, which are independent and have the identical law $X_i(P) = \mu_x$ defined by

$$\mu_x(\{0\}) = 1 - x \text{ and } \mu_x(\{1\}) = x.$$

(again a Bernoulli law). The existence of this random vector follows from the discussion above, or one could be content with the finite probability space $\Omega = \{0,1\}^n$ with $X_i(\varepsilon) = \varepsilon_i$ for $(\varepsilon_i) \in \Omega$, and

$$P((\varepsilon_i)) = (1-x)^{|\{i \mid \varepsilon_i = 0\}|} x^{|\{i \mid \varepsilon_i = 1\}|}.$$

We denote that $E(X_i) = E(X_1) = x$ and $V(X_i) = V(X_1) = x(1-x)$ for all $i \leqslant n$. Now we claim that the following formula holds:

$$B_n(x) = E\Big(f\Big(\frac{X_1 + \cdots + X_n}{n}\Big)\Big). \tag{4.16}$$

Indeed, the random variable

$$S_n = X_1 + \cdots + X_n$$

takes values in $\{0, \dots, n\}$, so that $f(S_n/n)$ takes only the values $f(k/n)$, $0 \leqslant k \leqslant n$; precisely, by summing over the values of S_n, we find

$$E\Big(f\Big(\frac{X_1 + \cdots + X_n}{n}\Big)\Big) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) P(S_n = k). \tag{4.17}$$

Now the event $\{S_n = k\}$ corresponds to those ω such that exactly k among the values

$$X_i(\omega), \qquad 1 \leqslant i \leqslant n,$$

are equal to 1, the rest being 0. From the independence of the X_i's, and the description of their probability law, we have

$$\begin{aligned} P(S_n = k) &= \sum_{|I|=k} P(X_{i_1} = 1) \cdots P(X_{i_k} = 1) P(X_{j_1} = 0) \cdots P(X_{j_{n-k}} = 0) \\ &= \binom{n}{k} x^k (1-x)^{n-k} \end{aligned}$$

(where

$$I = \{i_1, \dots, i_k\}$$

ranges over all subsets of $\{1, \dots, n\}$ of order k, and

$$J = \{j_1, \dots, j_{n-k}\}$$

is the complement of I). Using this and (4.17), we obtain the formula (4.16).

Now note that we also have $E(S_n/n) = E(X_1) = x$, and therefore the law of large numbers suggests that, since S_n/n tends to be close to x, we should have

$$E(f(S_n/n)) \to f(x).$$

This is the case, and this can in fact be proved without appealing directly to the results of the previous chapter. For this, we write

$$|B_n(x) - f(x)| = \Big|E\Big(f\Big(\frac{S_n}{n}\Big) - f\Big(E\Big(\frac{S_n}{n}\Big)\Big)\Big)\Big|,$$

and use the following idea: if S_n/n is close to its expectation, the corresponding contribution will be small because of the (uniform) continuity of f; as for the remainder, where S_n/n is "far" from the mean, it will have small probability because, in effect, of the weak law of large numbers.

Now, for the details, fix any $\varepsilon > 0$. By uniform continuity, there exists $\delta > 0$ such that $|x - y| < \delta$ implies

$$|f(x) - f(y)| < \varepsilon.$$

Now denote A the event

$$A = \{|S_n/n - E(S_n/n)| < \delta\}.$$

We have

$$\Big|\int_A \Big(f\Big(\frac{S_n}{n}\Big) - f(x)\Big) dP\Big| \leqslant \int_A \Big|f\Big(\frac{S_n}{n}\Big) - f(x)\Big| dP \leqslant \varepsilon P(A) \leqslant \varepsilon, \tag{4.18}$$

and as for the complement B of A, we have

$$\Big|\int_B \Big(f\Big(\frac{S_n}{n}\Big) - f(x)\Big) dP\Big| \leqslant 2\|f\|_\infty P(B), \tag{4.19}$$

so that, using the Chebychev inequality (3.23), we derive

$$P(B) = P(|S_n/n - E(S_n/n)| \geqslant \delta) \leqslant \frac{V(S_n/n)}{\delta^2} = \frac{V(X_1)}{n\delta^2} \leqslant \frac{1}{n\delta^2}$$

(same computation as for (3.21), using also $V(X_1) = x(1-x) \leqslant 1$). To conclude, we have shown that

$$|B_n(x) - f(x)| \leqslant \varepsilon + 2\frac{\|f\|_\infty}{n\delta^2}.$$

For fixed ε, δ being therefore fixed, we find that

$$|B_n(x) - f(x)| < 2\varepsilon$$

for all

$$n > \frac{2\|f\|_\infty}{\varepsilon\delta^2},$$

and since this upper bound is independent of x, this gives the uniform convergence of (B_n) towards f. ☐

REMARK 4.2.5. This proof is quite enlightening. For instance, although the simple finite probability space $\Omega = \{0,1\}^n$ suffices for the construction of the random variables, it was not necessary to use this particular choice: only the structure (independence, and law) of the random variables was needed. In more complicated constructions, this feature of modern probability theory is particularly useful.

4.3. The Fubini–Tonelli theorems

The construction of the product measure in Section 4.1 means that the "change of order" formula (4.1) is valid when f is the characteristic function of a set $C \in \mathcal{M} \otimes \mathcal{N}$. We will deduce from this the general case, using linearity, positivity and limiting processes, as usual.

THEOREM 4.3.1 (Tonelli theorem and Fubini theorem). *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measured spaces, and let $\mu \times \nu$ denote the product measure on $X \times Y$.*

(1) [Tonelli] *If $f : X \times Y \to [0, +\infty]$ is measurable, then the non-negative functions*

$$x \mapsto \int_Y f(x,y)d\nu(y)$$
$$y \mapsto \int_X f(x,y)d\mu(x)$$

are measurable with respect to $\mathcal{M}$ and $\mathcal{N}$, respectively, and we have

$$\int_{X\times Y} fd(\mu \otimes \nu) = \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x) = \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y). \tag{4.20}$$

(2) [Fubini] *If $f : X \times Y \to \mathbf{C}$ is in $L^1(\mu \otimes \nu)$, then, for ν-almost all $y \in Y$, the function*

$$t^y(f) : x \mapsto f(x,y)$$

is measurable and μ-integrable on X, for μ-almost all $x \in X$, the function

$$t_x(f) : y \mapsto f(x,y)$$

is measurable and ν-integrable, the functions

$$x \mapsto \int_Y f(x,y)d\nu(y)$$
$$y \mapsto \int_X f(x,y)d\mu(x)$$

are respectively ν-integrable and μ-integrable, and we have

$$\begin{aligned}\int_{X\times Y} fd(\mu\otimes\nu) &= \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x)\\ &= \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y).\end{aligned}\tag{4.21}$$

(3) *If $f : X\times Y \to \mathbf{C}$ is such that*

$$x \mapsto \int_Y |f(x,y)|d\nu(y)$$

is μ-integrable, then $f \in L^1(\mu\otimes\nu)$.

We first note that the operations $f \mapsto t_x(f)$ (which associate a function of the variable $y \in Y$ to a function of two variables) satisfy the following obvious formal properties, where $x \in X$ is arbitrary:

$$\begin{gathered}t_x(\alpha f+\beta g) = \alpha t_x(f) + \beta t_x(g)\\ t_x(f)^{\pm} = t_x(f^{\pm})\\ \text{If } f \leqslant g \text{ then } t_x(f) \leqslant t_x(g)\\ \text{If } f_n(x) \to f(x) \text{ for all } x, \text{ then } t_x(f_n) \to t_x(f) \text{ everywhere,}\end{gathered}$$

(and similarly for $f \mapsto t^y(f)$).

Proof. (1) We have already noticed that this is valid, by definition of the product measure, when f is the characteristic function of a measurable set $C \in \mathcal{M}\otimes\mathcal{N}$. By linearity, this remains true for any non-negative step function on $X\times Y$.

Now let $f \geqslant 0$ be any measurable non-negative function on $X\times Y$, and let (s_n) be any non-decreasing sequence of non-negative step functions on $X\times Y$ such that

$$s_n(x,y) \to f(x,y)$$

for all $(x,y) \in X\times Y$. The properties above show that, for any fixed x, the sequences $(t_x(s_n))_n$ converge pointwise to $t_x(f)$, and the convergence is non-decreasing. Hence, first of all, we find that $t_x(f)$ is measurable (this is also proved in Corollary 4.1.4), and then the monotone convergence theorem on Y gives

$$\int_Y t_x(f)d\nu = \lim_{n\to+\infty}\int_Y t_x(s_n)d\nu$$

for all x. Furthermore, this limit, seen as pointwise limit of non-negative functions on X, is also monotone, and by a second use of the monotone convergence theorem (on X), we find

$$\begin{aligned}\int_X \int_Y t_x(f) d\nu(y) d\mu(x) &= \lim_{n\to+\infty} \int_X \int_Y t_x(s_n) d\nu(y) d\mu(x)\\ &= \lim_{n\to+\infty} \int_{X\times Y} s_n d(\mu\otimes\nu) \text{ (case of step functions)}\\ &= \int_{X\times Y} f d(\mu\otimes\nu),\end{aligned}$$

where the last step was a third application of the monotone convergence theorem, this type to the original sequence $s_n \to f$ on $X \times Y$. Thus we have proved the first formula in (4.20), and the second is true simply by exchanging X and Y.

(2) Let $f \in L^1(\mu\otimes\nu)$. By Corollary 4.1.4, the functions $t_x(f)$ are measurable for all x. Moreover, by (4.20) (that we just proved), we have

$$\int_{X\times Y} |f| d(\mu\otimes\nu) = \int_X \Bigl(\int_Y |f(x,y)| d\nu(y)\Bigr) d\mu(x) < +\infty,$$

and this means that the function of x which is integrated must be finite almost everywhere, i.e., this means that $t_x(f) \in L^1(\nu)$ for μ-almost all $x \in X$.

Assume first that f is real-valued. Then, for any x for which $t_x(f) \in L^1(\nu)$, we have

$$\int_Y f(x,y) d\nu(y) = \int_Y t_x(f^+) d\nu(y) - \int_Y t_x(f^-) d\nu(y), \tag{4.22}$$

and the first part again shows that $x \mapsto \int f(x,y) d\nu(y)$ (which is defined almost everywhere, and extended for instance by zero elsewhere) is measurable. Since

$$\int_X \Bigl|\int_Y f(x,y) d\nu(y)\Bigr| d\mu(x) \leqslant \int_X \int_Y |f(x,y)| d\nu(y) d\mu(x) < +\infty,$$

this function is even μ-integrable. And so, integrating (4.22) on X, we obtain

$$\begin{aligned}\int_X \int_Y f(x,y) d\nu(y) d\mu(x) &= \int_X \int_Y t_x(f^+) d\nu(y) d\mu(x) - \int_X \int_Y t_x(f^-) d\nu(y) d\mu(x)\\ &= \int_{X\times Y} f^+ d(\mu\otimes\nu) - \int_{X\times Y} f^- d(\mu\otimes\nu)\\ &= \int_{X\times Y} f d(\mu\otimes\nu),\end{aligned}$$

(using Tonelli's theorem once more), which is (4.21). The case of f complex-valued is exactly similar, using

$$f = \mathrm{Re}(f) + i\,\mathrm{Im}(f), \qquad t_x(f) = t_x(\mathrm{Re}(f)) + i t_x(\mathrm{Im}(f)),$$

and finally the usual exchange of X and Y gives the second formula.

(3) By Tonelli's theorem, the assumption implies that

$$\int_{X\times Y} |f| d(\mu\otimes\nu) = \int_X \Bigl(\int_Y |f(x,y)| d\nu(y)\Bigr) d\mu(x) < +\infty$$

and hence that $f \in L^1(\mu\otimes\nu)$. ∎

Remark 4.3.2. One should see (3) as a convenient criterion to apply (2); indeed, it is not always clear that a given function f on $X \times Y$ is in $f \in L^1(\mu \otimes \nu)$. In practice, one often proves an estimate like

$$\int_Y |f(x,y)| d\nu(y) \leqslant g(x)$$

where $g \in L^1(\mu)$, to apply (3).

Example 4.3.3. One of the common practical applications of Fubini's theorem is in the evaluation of certain definite integrals involving functions which are themselves defined using integrals, as in Section 3.1. Thus, let $h : X \times Y \to \mathbf{C}$ be measurable and

$$f(x) = \int_Y h(x,y) d\nu(y),$$

assuming this is well-defined.

In many cases, one can evaluate

$$\int_X f(x) d\mu(x) = \int_Y \int_X h(x,y) d\mu(x) d\nu(y)$$

by a direct application of Fubini's theorem (which must, of course, be justified).

Here is one example among many. Consider the function[1]

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin\theta) d\theta$$

for $x \in \mathbf{R}$. Since the integrand is continuous over $\mathbf{R} \times [0,\pi]$, Proposition 3.1.1 shows that it is continuous, and in particular, measurable. Moreover, we have

$$|J_0(x)| \leqslant 1$$

for all x, and hence the function $x \mapsto e^{-x} J_0(x)$ on $[0,+\infty[$ is integrable with respect to Lebesgue measure.

This means, in particular, that Fubini's theorem is applicable, and we derive

$$\begin{aligned}\int_0^{+\infty} e^{-x} J_0(x) dx &= \frac{1}{\pi} \int_0^{+\infty} \int_0^{\pi} e^{-x} \cos(x \sin\theta) d\theta dx \\ &= \frac{1}{\pi} \int_0^{\pi} \int_0^{\infty} e^{-x} \cos(x \sin\theta) dx d\theta.\end{aligned}$$

Now the inner integral is elementary:

$$\begin{aligned}\int_0^{\infty} e^{-x} \cos(x \sin\theta) dx &= \mathrm{Re}\Bigl(\int_0^{\infty} e^{-x(1-i\sin\theta)} dx\Bigr) \\ &= \mathrm{Re}\Bigl(\Bigl[-\frac{1}{1-i\sin\theta} e^{-x(1-i\sin\theta)}\Bigr]_0^{+\infty}\Bigr) \\ &= \mathrm{Re}\Bigl(\frac{1}{1-i\sin\theta}\Bigr) = \frac{1}{1+\sin^2\theta},\end{aligned}$$

and therefore we have

$$\int_0^{+\infty} e^{-x} J_0(x) dx = \frac{1}{\pi} \int_0^{\pi} \frac{d\theta}{1+\sin^2\theta} = \frac{1}{\sqrt{2}}$$

(the last computation follows from the standard change of variable $u = \tan\theta$ on $[0,\pi/2[$, the other contribution being easily shown to be equal to it).

[1]. The simplest type among the Bessel functions.

EXAMPLE 4.3.4. One can also apply Fubini's theorem, together with Lemma 4.2.1, to recover easily some properties of independent random variables already mentioned in the previous chapter.

For instance, we recover the result of Exercise 2.3.5 as follows:

PROPOSITION 4.3.5. *Let (Ω, Σ, P) be a probability space, X and Y two independent integrable random variables. We then have $XY \in L^1(P)$ and*

$$E(XY) = E(X)E(Y).$$

PROOF. According to Lemma 4.2.2, $|X|$ and $|Y|$ are non-negative random variables which are still indepdendent. We first check the desired property in the case where $X \geqslant 0$, $Y \geqslant 0$; once this is done, it will follow that $XY \in L^1(P)$.

Let μ be the probability law of X, and ν that of Y. By Lemma 4.2.1, the joint law of X and Y is given by

$$(X, Y)(P) = \mu \otimes \nu.$$

Let $m : \mathbf{C}^2 \to \mathbf{C}$ denote the multiplication map. By Proposition 2.3.3, (3), we have

$$\begin{aligned} E(XY) &= \int_\Omega XY dP = \int_\Omega m(X, Y) dP \\ &= \int_{\mathbf{C}^2} md(X, Y)(P) \\ &= \int_{\mathbf{C}^2} xyd(\mu \otimes \nu)(x, y). \end{aligned} \tag{4.23}$$

But since m is measurable (for instance because it is continuous), Tonelli's theorem gives

$$\int_{\mathbf{C}^2} md(\mu \otimes \nu) = \int_{\mathbf{C}} \int_{\mathbf{C}} xyd\nu(y)d\mu(x) = \Big(\int_{\mathbf{C}} xd\mu(x)\Big)\Big(\int_{\mathbf{C}} yd\nu(y)\Big) = E(X)E(Y),$$

as desired.

Coming back to the general case, we have established that $XY \in L^1(P)$. Then the computation of $E(XY)$ proceeds like before, using (4.23), which are now valid for X and Y integrable because of Fubini's theorem. ☐

4.4. The Lebesgue integral on $\mathbf{R}^d$

We now come to the particularly important example of the Lebesgue measure and integral on $\mathbf{R}^d$.

For $d = 1$, we have assumed the existence of the Lebesgue measure λ_1. Now for $d \geqslant 2$, we denote by λ_d (or sometimes only λ, when the value of d is clear from context) the product measure

$$\lambda_d = \lambda_1 \times \cdots \times \lambda_1,$$

on the product σ-algebra, which is also the Borel σ-algebra of $\mathbf{R}^d$.

By definition, we have

$$\lambda(C) = (b_1 - a_1) \cdots (b_d - a_d) \text{ for } C = [a_1, b_1] \times \cdots \times [a_d, b_d]$$

(a generalized *cube*). This measure is also called the Lebesgue measure in dimension d; for $d = 2$, one may also speak of it as giving the *area* of a (measurable) subset of the plane, and for $d = 3$, one may speak of *volume*.

It is customary to write

$$\int_{\mathbf{R}^d} f(x_1,\ldots,x_d)dx_1\cdots dx_d = \int_{\mathbf{R}^d} f(x)dx$$

for the integral of a function f of d variables $x=(x_1,\ldots,x_d)$ with respect to the Lebesgue measure.

With the Lebesgue measure come the L^p spaces $L^p(d\lambda)=L^p(\mathbf{R}^d)$, $1\leqslant p\leqslant +\infty$. It is convenient to have simple criteria for functions to be L^p, by comparison with known functions. The following lemma generalizes Example 2.4.1.

Proposition 4.4.1. *Let $d\geqslant 1$ be fixed, and let*

$$\|t\| = \Bigl(\sum_{1\leqslant i\leqslant d} |t_i|^2\Bigr)^{1/2} \tag{4.24}$$

For $t\in\mathbf{R}^d$ be the Euclidian norm on $\mathbf{R}^d$, and

$$\|t\|_\infty = \max(|t_j|).$$

Let f be a non-negative measurable function $\mathbf{R}^d$ and $p\in[1,+\infty[$.

(1) If f is bounded, and if there exists a constant $C\geqslant 0$ and $\varepsilon>0$ such that

$$|f(x)|\leqslant C\|x\|^{-d-\varepsilon}, \qquad \textit{or} \qquad |f(x)|\leqslant C\|x\|_\infty^{-d-\varepsilon},$$

for $x\in\mathbf{R}^d$ with $\|x\|\geqslant 1$, or with $\|x\|_\infty\geqslant 1$, then $f\in L^1(\mathbf{R}^d)$.

(2) If f is bounded and there exist $C\geqslant 0$ and $\varepsilon>0$ such that

$$|f(x)|\leqslant C\|x\|^{-d/p-\varepsilon}, \qquad \textit{or} \qquad |f(x)|\leqslant C\|x\|_\infty^{-d/p-\varepsilon},$$

pour $x\in\mathbf{R}^d$ with $\|x\|\geqslant 1$, then $f\in L^p(\mathbf{R}^d)$.

(3) If f is integrable on $\{x\mid \|x\|>1\}$, or on $\{x\mid \|x\|_\infty>1\}$ and there exist $C\geqslant 0$ and $\varepsilon>0$ such that

$$|f(x)|\leqslant C\|x\|^{-d+\varepsilon}, \qquad \textit{or} \qquad |f(x)|\leqslant C\|x\|_\infty^{-d+\varepsilon},$$

for $\|x\|\leqslant 1$ or $\|x\|_\infty\leqslant 1$, then $f\in L^1(\mathbf{R}^d)$.

(4) If f is integrable on $\{x\mid \|x\|>1\}$, or on $\{x\mid \|x\|_\infty>1\}$ and there exist $C\geqslant 0$ and $\varepsilon>0$ such that

$$|f(x)|\leqslant C\|x\|^{-d/p+\varepsilon}, \qquad \textit{or} \qquad |f(x)|\leqslant C\|x\|_\infty^{-d/p+\varepsilon},$$

for $\|x\|\leqslant 1$ or $\|x\|_\infty\leqslant 1$, then $f\in L^p(\mathbf{R}^d)$.

Proof. Obviously, (2) and (4) follow from (1) and (3) applied to f^p instead of f. Moreover, we have

$$\frac{1}{\sqrt{d}}\|x\|_\infty\leqslant\|x\|\leqslant\|x\|_\infty,$$

for all $x\in\mathbf{R}^d$, the right-hand inequality coming e.g. from Cauchy's inequality, since

$$\max(|x_j|)\leqslant\sum_j|x_j|\leqslant\sqrt{d}\|x\|,$$

and this means we can work with either of $\|x\|$ or $\|x\|_\infty$. We choose the second possibility, and prove only (1), leaving (2) as an exercise.

It is enough to consider f defined by

$$f(x)=\begin{cases}0 & \text{if } \|x\|_\infty\leqslant 1\\ \|x\|_\infty^{-\alpha} & \end{cases}$$

for some $\alpha \geqslant 0$, and to determine for which values of α this belongs to $L^1(\mathbf{R}^d)$. We compute the integral of f by partitioning $\mathbf{R}^d$ according to the rough size of $\|x\|_\infty$:

$$\int_{\mathbf{R}^d} f(x) d\lambda_d(x) = \sum_{n \geqslant 1} \int_{A_n} \|x\|^{-\alpha} d\lambda_d(x),$$

where

$$A_n = \{x \in \mathbf{R}^d \mid n \leqslant \|x\|_\infty < n+1\}.$$

We get inequalities

$$\sum_{n \geqslant 1} (n+1)^{-\alpha} \lambda_d(A_n) \leqslant \int_{\mathbf{R}^d} f(x) d\lambda_d(x) \leqslant \sum_{n \geqslant 1} n^{-\alpha} \lambda_d(A_n)$$

Now, since

$$A_n =]-n-1, n+1[^d - [-n, n]^d,$$

we have

$$\lambda_d(A_n) = (2n+1)^d - (2n)^d \sim d(2n)^{d-1} \text{ as } n \to +\infty,$$

by definition of the product measure. Hence the inequalities above show that $f \in L^1(\mathbf{R}^d)$ if and only if $\alpha - (d-1) > 1$, i.e., if $\alpha > d$, which is exactly (1). $\square$

We now anticipate a bit some of the discussion of the next chapter to discuss one of the most important features of the Lebesgue integral in $\mathbf{R}^d$, namely the *change of variable formula*.

First, we recall some definitions of multi-variable calculus.

Definition 4.4.2 (Differentiable functions, diffeomorphisms). Let $d \geqslant 1$ be give, and let U, $V \subset \mathbf{R}^d$ be non-empty open sets, and

$$\varphi : U \to V$$

a map from U to V.

(1) For given $x \in U$, φ *is differentiable* at x if there exists a linear map

$$T : \mathbf{R}^d \to \mathbf{R}^d$$

such that

$$\varphi(x+h) = \varphi(x) + T(h) + o(\|h\|)$$

for all $h \in \mathbf{R}^d$ such that $x + h \in U$. The map T is then unique, and is called the *differential of* φ *at* x, denoted $T = D_x(\varphi)$.

(2) The map φ is differentiable on U if it is differentiable at all $x \in U$, and φ is *of* C^1 *class* if the map

$$\begin{cases} U \to L(\mathbf{R}^d) \simeq \mathbf{R}^{n^2} \\ x \mapsto D_x(\varphi) \end{cases}$$

is continuous on U, where $L(\mathbf{R}^d)$ is the vector space of all linear maps $T : \mathbf{R}^d \to \mathbf{R}^d$.

(3) The map φ is a diffeomorphism (resp. a C^1-diffomorphism) on U if φ is differentiable on U (resp., is of C^1-class on U), φ is bijective, and the inverse map

$$\psi = \varphi^{-1} : V \to U$$

is also differentiable (resp., of C^1-class) on V.

(4) If φ is differentiable on U, the *jacobian* of φ is the map

$$J_\varphi \begin{cases} U \to \mathbf{R} \\ x \mapsto \det(D_x(\varphi)) \end{cases},$$

which is continuous on U if φ is C^1.

In other words, the differential of φ at a point is the "best approximation" of f among the simplest functions, namely the linear maps. When $d = 1$, a linear map $T : \mathbf{R} \to \mathbf{R}$ is always of the type $T(x) = \alpha x$ for some $\alpha \in \mathbf{R}$; since α is canonically associated with T (because $\alpha = T(1)$), one can safely identify T with $\alpha \in \mathbf{R}$, and these numbers provide the usual derivative of φ.

Using the implicit function theorem, the following criterion can be proved:

Proposition 4.4.3. *Let U, $V \subset \mathbf{R}^d$ be non-empty open sets. A map $\varphi : U \to V$ is a C^1-diffeomorphism if and only if f is bijective, of C^1-class, and the differential $D_x(\varphi)$ is an* invertible *linear map for all $x \in U$. We then have*

$$D_y(\varphi^{-1}) = D_{\varphi^{-1}(y)}(\varphi)^{-1} = D_x(\varphi)^{-1}$$

for all $y \in V$, $y = \varphi(x)$ with $x \in U$. Moreover if $\psi = \varphi^{-1}$, we have

$$J_\psi(y) = J_\varphi(x)^{-1} \tag{4.25}$$

for $y = \varphi(x)$.

Remark 4.4.4. Concretely, if φ is given by "formulas", i.e., if

$$\varphi(x) = (\varphi_1(x_1, \ldots, x_d), \ldots, \varphi_d(x_1, \ldots, x_d))$$

where $\varphi_j : \mathbf{R}^d \to \mathbf{R}$, then φ is of C^1-class if and only if all the partial derivatives of the φ_j exist, are continuous functions on U. Then $D_x(\varphi)$ is the linear map given by the matrix of partial derivatives

$$D_x(\varphi) = \Big(\frac{\partial \varphi_i}{\partial x_j}\Big)_{i,j},$$

the determinant of which gives $J_\varphi(x)$.

It is customary to think of φ as giving a "change of variable"

$$y_i = \varphi_i(x_1, \ldots x_d),$$

where the inverse of φ gives the formula that can be used to go back to the original variables.

Example 4.4.5. (1) The simplest changes of variables are linear ones: if $T : \mathbf{R}^d \to \mathbf{R}^d$ is linear and invertible, then of course it is a C^1-diffeomorphism on $\mathbf{R}^d$ with $D_x(f) = T$ for all $x \in \mathbf{R}^d$, and $J_T(x) = |\det(T)|$ for all x.

(2) The translation maps defined by

$$\tau : x \mapsto x + a$$

for some fixed $a \in \mathbf{R}^d$ are also diffeomorphisms, where $D_x(\tau)$ equal to the identity matrix (and hence $J_\tau(x) = 1$) for all $x \in \mathbf{R}^d$.

(3) The "polar coordinates" in the plane given an important change of variable. Here we have $d = 2$, and one starts by noticing that every $(x, y) \in \mathbf{R}^2$ can be written

$$(x, y) = (r \cos\theta, r \sin\theta) = \varphi(r, \theta)$$

with $r \geqslant 0$ and $\theta \in [0, 2\pi[$. This is "almost" a diffemorphism, but minor adjustements are necessary to obtain one; first, since $(x, y) = 0$ is obtained from all $(0, \theta)$, to have an injective map one must restrict φ to $]-0, +\infty[\times[0, 2\pi[$. This is however not an open subset of $\mathbf{R}^2$, so we restrict further to obtain

$$\varphi : U \to V$$

where

$$U =]0, +\infty[\times]0, 2\pi[,$$

and to have a surjective map, the image is

$$V = \mathbf{R}^2 - \{(x, 0) \mid x \geqslant 0\}$$

(the plane minus the non-negative part of the real axis).

It is then easy to see that the polar coordinates, with this restriction, is a differentiable bijection from U to V. Its jacobian matrix is

$$D_{(r,\theta)}(\varphi) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

and the determinant is

$$J_\varphi(r, \theta) = r > 0,$$

which shows (using the proposition above) that φ is C^1-diffeomorphism.

As we will see, from the point of view of integration theory, having removed part of the plane will be irrelevant, since this is a set of Lebesgue measure zero.

Here is the change of variable formula, which will be proved in the next chapter.

Theorem 4.4.6 (Change of variable in $\mathbf{R}^d$). *Let $d \geqslant 1$ be fixed and let*

$$\varphi : U \to V$$

be a C^1-diffeomorphism between two non-empty open subsets of $\mathbf{R}^d$.

(1) *If f is a measurable function on V which is either non-negative or integrable with respect to the Lebesgue measure restricted to V, then we have*

$$\int_V f(y)d\lambda(y) = \int_U f(\varphi(x))|J_\varphi(x)|d\lambda(x),$$

which means, in the case where f is integrable, that if either of the two integrals exists, then so does the other, and they are equal.

(2) *If f is a measurable function on V which is either non-negative or such that $f \circ \varphi$ is integrable on U with respect to the Lebesgue measure, then we have*

$$\int_U f(\varphi(x))d\lambda(x) = \int_V f(y)|J_\varphi(\varphi^{-1}(y))|^{-1}d\lambda(y) = \int_V f(y)|J_{\varphi^{-1}}(y)|d\lambda(y).$$

(3) *The image measure of the Lebesque measure on U by φ is given by*

$$\varphi_*(d\lambda(x)) = |J_{\varphi^{-1}}(y)|d\lambda(y). \tag{4.26}$$

We only explain quickly why those three statements are equivalent (assuming they hold for all diffeomorphisms). First of all, assuming (1), we can apply it to

$$g(x) = f(x)|J_\varphi(\varphi^{-1}(x))|^{-1}$$

instead of f. By (4.25), we have

$$g(x) = f(x)|J_{\varphi^{-1}}(x)|,$$

and this gives the first equality in (2). One may also apply (1) directly to $\psi = \varphi^{-1}$, and this symmetry shows that (2) implies (1) also. Finally, (3) is equivalent with (2) because of the abstract formula

$$\int_V f(y)\varphi_*(d\lambda)(y) = \int_U f(\varphi(x))d\lambda(x),$$

given by (2.8) and (2.12) for any function on V which is either integrable or non-negative.

REMARK 4.4.7. (1) The presence of the absolute value of the jacobian in the formula is due to the fact that the Lebesgue integral is defined in a way which does not take into account "orientation". For instance, if $d = 1$ and $\varphi(x) = -x$, the differential is -1 and the jacobian satisfies $|J_\varphi(x)| = 1$; the change of variable formula becomes

$$\int_{[a,b]} f(x)d\lambda(x) = \int_{[-b,-a]} f(-y)d\lambda(y)$$

for $a < b$, which may be compared with the usual way of writing it

$$\int_a^b f(x)dx = -\int_{-a}^{-b} f(-y)dy = \int_{-b}^{-a} f(-y)dy$$

for a Riemann integral, where the correct sign is obtained from the orientation.

Since φ is a diffeomorphisms, the jacobian J_φ does not vanish on U, and hence if U is connected, the sign of $J_\varphi(x)$ will be the same for all $x \in U$, so that

$$|J_\varphi(x)| = \varepsilon J_{\varphi(x)},$$

in that case, with $\varepsilon = \pm 1$ independent of x.

(2) Intuitively, the jacobian factor has the following interpretation: for an invertible linear map T, it is well-known that the absolute value $|\det(T)|$ of the determinant is the "volume" (i.e., the d-dimensional Lebesgue measure) of the image of the unit cube under T; this is quite clear if T is diagonal or diagonalizable (each direction in the space is then stretched by a certain factor, and the product of these is the determinant), and otherwise is a consequence of the formula in any case: take $U = [0,1]^d$, $V = T(U)$ and $f = 1$ to derive

$$\lambda_d(T(U)) = \mathrm{Vol}(T(U)) = \int_{T(U)} f(x)d\lambda(x) = \int_U |\det(T)|d\lambda(y) = |\det(T)|.$$

Thus, one should think of $J_\varphi(x)$ as the "dilation coefficient" around x, for "infinitesimal" cubes centered at x. This may naturally suggest that the Lebesgue measure should obey the rule (4.26), after doing a few drawings if needed...

(3) In practice, for many changes of variables of interest, the "natural" definition does not lead immediately to a C^1-diffeomorphism between open sets, as we saw in the case of the polar coordinates. A more general statement, which follows immediately, is the following: let $\varphi : A \to B$, where A, $B \subset \mathbf{R}^d$ are such that

$$A = U \cup A_0, \qquad B = V \cup B_0,$$

where U and V are open and $\lambda(A_0) = \lambda(B_0) = 0$. If φ, restricted to U, is a C^1-diffeomorphism $U \to V$, then we have

$$\int_B f(y)d\lambda(y) = \int_A f(\varphi(x))|J_\varphi(x)|d\lambda(x),$$

for any measurable function f on B which is either integrable or non-negative, where $J_\varphi(x)$ can be extended arbitrarily (e.g., to take the value 0) on A_0.

EXAMPLE 4.4.8. (1) If T is an invertible linear map, we have

$$\int_{\mathbf{R}^d} f(T(x))d\lambda(x) = \frac{1}{\det(T)} \int_{\mathbf{R}^d} f(x)d\lambda(x)$$

for all f. In particular, if T is a *rotation*, or more generall $T \in O(d, \mathbf{R})$, a euclidean isometry, we have $\det(T) = \pm 1$ and

$$\int_{\mathbf{R}^d} f(T(x)) d\lambda(x) = \int_{\mathbf{R}^d} f(x) d\lambda(x),$$

or in other words $T_*(\lambda) = \lambda$: one says that *the Lebesgue measure on* $\mathbf{R}^d$ *is invariant under rotation.*

(2) Let $f : \mathbf{R}^2 \to \mathbf{C}$ be an integrable function. It can be integrated in polar coordinates (Example 4.4.5, (3)): since $Z = \{(a, 0) \mid a \geqslant 0\} \subset \mathbf{R}^2$ has measure zero, we can write

$$\int_{\mathbf{R}^2} f(x) d\lambda_2(x) = \int_{\mathbf{R}^2 - Z} f(x) d\lambda_2(x)$$

and therefore (with notation as in Example 4.4.5, (3)) we get

$$\begin{aligned}\int_{\mathbf{R}^2} f(x) d\lambda_2(x) &= \int_{V_1} f(r\cos\theta, r\sin\theta) r dr d\theta \\ &= \int_0^{2\pi} \int_0^{+\infty} f(r\cos\theta, r\sin\theta) r dr d\theta\end{aligned}$$

(Fubini's theorem justifies writing the integral in this manner, or indeed with the order of the r and θ variables interchanged).

Suppose now that f is *radial*, which means that there exists a function

$$g : [0, +\infty[\to \mathbf{C}$$

such that

$$f(x, y) = g(\sqrt{x^2 + y^2}) = g(r)$$

for all x and y. In that case, one can integrate over θ first, and get

$$\int_{\mathbf{R}^2} f(x) d\lambda_2(x) = \int_{-\pi}^{\pi} \int_0^{+\infty} g(r) r dr d\theta = 2\pi \int_0^{+\infty} g(r) r dr.$$

Here is an example of application of this:

Proposition 4.4.9 (Gaussians are probability measures). *Let μ be the measure on $\mathbf{R}$ given by*

$$\mu = e^{-\pi x^2} dx,$$

where dx denotes Lebesgue measure.

Then μ is a probability measure. Moreover, if X is any random variable with probability law $X(P) = \mu$, we have $X \in L^2$ and

$$E(X) = 0, \qquad V(X) = \frac{1}{2\pi}.$$

Proof. Let $f : \mathbf{R}^2 \to [0, +\infty[$ be the function defined by

$$f(x, y) = e^{-\pi(x^2+y^2)},$$

which is non-negative and radial. Thus the integral with respect to Lebesgue measure exists and by the formula above, we have

$$\int_{\mathbf{R}^2} e^{-\pi(x^2+y^2)} d\lambda_2(x) = 2\pi \int_0^{+\infty} e^{-\pi r^2} r dr = \left[-e^{-r^2}\right]_0^{+\infty} = 1.$$

However, by means of Tonelli's theorem, we also have

$$\int_{\mathbf{R}^2} e^{-\pi(x^2+y^2)} d\lambda_2(x) = \int_{\mathbf{R}} e^{-\pi x^2} \Big(\int_{\mathbf{R}} e^{-\pi y^2} d\lambda(y) \Big) d\lambda(x) = I^2$$

where

$$I = \int_{\mathbf{R}} e^{-\pi x^2} dx = \int_{\mathbf{R}} d\mu(x).$$

Comparing, we obtain

$$\int_{\mathbf{R}} d\mu(x) = 1,$$

(since the integral is clearly non-negative), showing that μ is a probability measure.

There remains to compute the expectation and variance of a random variable with law μ. We have first

$$E(X) = \int_{\mathbf{R}} x d\mu(x) = \int_{\mathbf{R}} x e^{-\pi x^2} dx = 0$$

since $x \mapsto x e^{-x^2}$ is clearly integrable and odd (substitute $\varphi(x) = -x$ to get the result). Then we get

$$V(x) = \int_{\mathbf{R}} x^2 d\mu(x) = \int_{\mathbf{R}} x^2 e^{-\pi x^2} dx = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-x^2} dx = \frac{1}{2\pi}$$

by a simple integration by parts. $\square$

More generally, the measure

$$\mu_{a,\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} d\lambda(x)$$

for $a \in \mathbf{R}$, $\sigma > 0$, is a probability measure with expectation a and variance σ^2 (cf. Remark 3.4.11). This follows from the case above by the change of variable

$$y = \frac{(x-a)}{\sigma},$$

and is left as an exercise.

CHAPTER 5

Integration and continuous functions

5.1. Introduction

When X is a topological space, we can consider the Borel σ-algebra generated by open sets, and also a distinguished class of functions on X, namely those which are continous. It is natural to consider the interaction between the two, in particular to consider special properties of integration of continuous functions with respect to measures defined on the Borel σ-algebra (those are called *Borel measures*). In this respect, we first note that by Corollary 1.1.10, any continuous function

$$f : X \to \mathbf{C}$$

is measurable with respect to the Borel σ-algebras.

However, to be able to say something more interesting, one must make some assumptions on the topological space as well as on the measures involved. There are two obvious obstacles: on the one hand, the vector space $C(X)$ of continuous (complex-valued) functions on X might be very small (it might contain only constant functions); on the other hand, even when $C(X)$ is "big", it might be that no interesting continuous function is μ-integrable, for a given Borel measure μ. (For instance, if $X = \mathbf{R}$ and ν is the counting measure, it is certainly a Borel measure, but

$$\int_{\mathbf{R}} f d\nu(x) = +\infty$$

for any continuous function $f \geqslant 0$ which is not identically zero, because, by continuity, such a function will be $\geqslant c > 0$, for some c, on some non-empty interval, containing infinitely many points).

The following easy proposition takes care of finding very general situations where *compactly supported* functions are integrable. We first recall for this purpose the definition of the support of a continuous function.

Definition 5.1.1 (Support). Let X be a topological space, and

$$f : X \to \mathbf{C}$$

a continuous function on X. The *support of f*, denoted $\mathrm{supp}(f)$, is the closed subset

$$\mathrm{supp}(f) = \bar{V} \text{ where } V = \{x \mid f(x) \neq 0\} \subset X.$$

If $\mathrm{supp}(X) \subset X$ is compact, the function f is said to be *compactly supported.* We denote by $C_c(X)$ the $\mathbf{C}$-vector space of compactly supported continuous functions on X.

One can therefore say that if $x \notin \mathrm{supp}(f)$, we have $f(x) = 0$. However, the converse does not hold; for instance, if $f(x) = x$ for $x \in \mathbf{R}$, we have

$$\mathrm{supp}(f) = \mathbf{R}, \qquad f(0) = 0.$$

When X is itself compact,[1] we have $C_c(X) = C(X)$, but otherwise the spaces are distinct (for instance, a non-zero constant function has compact support only if X is compact).

To check that $C_c(X)$ is a vector space, one uses the obvious relation

$$\operatorname{supp}(\alpha f + \beta g) \subset \operatorname{supp}(f) \cup \operatorname{supp}(g).$$

Note also an important immediate property of compactly-supported functions: they are bounded on X. Indeed, we have

$$|f(x)| \leqslant \sup_{x \in \operatorname{supp}(f)} |f(x)|$$

for all x, and of course f is bounded on the compact $\operatorname{supp}(f)$. We denote

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in X\},$$

for $f \in C_c(X)$, which is a norm on $C_c(X)$.

We have then the following simple-looking fact:

Proposition 5.1.2. *Let X be a topological space and μ a Borel measure which is* finite on compact sets*, i.e., such that*

$$\mu(K) < +\infty,$$

for any compact subset $K \subset X$. Then the map

$$\begin{cases} C_c(X) \to \mathbf{C} \\ f \mapsto \displaystyle\int_X f d\mu(x) \end{cases}$$

is well-defined, it is linear, and moreover it is positive*: for any $f \in C_c(X)$ which is non-negative on X, we have $\Lambda(f) \geqslant 0$.*

Proof. The only point that needs proof is that Λ is well-defined. However, if $f \in C_c(X)$, so is $|f|$, and since $|f|$ is bounded by the remark before the statement, and is zero outside the compact set $K = \operatorname{supp}(f)$, we have

$$\int_X |f(x)| d\mu(x) = \int_K |f(x)| d\mu(x) \leqslant \mu(K) \|f\|_\infty < +\infty$$

since we assumed that μ is finite on compact sets. ☐

Remark 5.1.3. Although it is tempting to state that

$$C_c(X) \subset L^1(\mu),$$

under the situation of the proposition, one must be aware that this is really an absure of notation, since $L^1(\mu)$ is the space of equivalence classes of functions, up to functions which are zero almost everywhere. Indeed, it is perfectly possible that the quotient map

$$C_c(X) \to L^1(\mu)$$

not be injective; a simple example is the Borel measure

$$\mu = (1 - \chi_{[-1,1]}) d\lambda$$

on $\mathbf{R}$, where λ is the Lebesgue measure. Of course, λ is finite on compact sets, and since

$$\mu(]-1/2, 1/2[) = 0,$$

[1] By convention, we consider that the compactness condition implies that X is separated.

we see that two continuous functions f_1 and f_2 which differ only on the interval $]-1/2, 1/2[$ actually define the same element in $L^1(\mu)$.

This proposition is extremely simple. It is therefore very striking that, for reasonable topological spaces (such as compact spaces, or $\mathbf{R}$) there is a converse: any linear map $C_c(X) \to \mathbf{C}$ which has the positivity property (that $\Lambda(f) \geqslant 0$ if $f \geqslant 0$) is obtained by integrating f against a fixed Borel measure μ. This powerful result is a very good way to construct measures; for instance, it may be applied to the Riemann integral, which is a well-defined map

$$\Lambda : C_c(\mathbf{R}) \to \mathbf{C},$$

and the resulting measure is the Lebesgue measure...

5.2. The Riesz representation theorem

Here is the converse to Proposition 5.1.2.

THEOREM 5.2.1. *Let X be* locally compact *topological space,*[2] *and let*

$$\Lambda : C_c(X) \to \mathbf{C}$$

be a linear map such that $\Lambda(f) \geqslant 0$ if $f \geqslant 0$.

(1) *There exists a σ-algebra $\mathcal{M} \supset \mathcal{B}_X$, and a complete measure μ on $\mathcal{M}$, such that μ is finite on compact sets, and*

$$\Lambda(f) = \int_X f d\mu(x) \text{ for all } f \in C_c(X).$$

(2) *In fact, there exists such a unique measure μ for which the following additional properties hold:*

(1) *For all $E \in \mathcal{M}$, we have*

$$\mu(E) = \inf\{\mu(U) \mid U \supset E \text{ is an open set containing } E\} \tag{5.1}$$

(2) *For all $E \in \mathcal{M}$, if E is either open or has finite measure, we have*

$$\mu(E) = \sup\{\mu(K) \mid K \subset E \text{ is compact}\}. \tag{5.2}$$

(3) *If X has the additional property that any open set in X is a countable union of compact sets, in which case X is called σ-compact, then the measure μ is unique as a measure on $(X, \mathcal{B}_X)$, i.e., without requiring* (5.1) *and (5.2).*

Before proving the theorem, here is its main application: the rigorous construction (and proof of existence) of Lebesgue measure. This is, indeed, highly enlightening.

EXAMPLE 5.2.2 (Construction of the Lebesgue measure). Let $X = \mathbf{R}$ and let Λ be the linear map

$$f \mapsto \int_{-\infty}^{\infty} f(x)dx = \int_a^b f(x)dx,$$

where the integral is a Riemann integral, and $[a, b] \subset \mathbf{R}$ is any interval such that $\mathrm{supp}(f) \subset [a, b]$. Applying Riesz's theorem, we obtain a Borel measure μ such that

$$\Lambda(f) = \int_{-\infty}^{+\infty} f(x)dx = \int_{\mathbf{R}} f(x)d\mu(x)$$

for any $f \in C_c(X)$. We claim that this measure *is the Lebesgue measure*, the existence of which was previously admitted (Theorem 1.3.1).

[2]. I.e., one for which every point x has one compact neighbourhood.

To check this, we remark first that μ is – by construction – a complete Borel measure, and it is enough to check that

$$\mu([a,b]) = b-a$$

for any real numbers $a \leqslant b$. The case $a = b$ is obvious, and so we assume that $a < b$. Then we construct the following sequences (f_n), (g_n) of continuous functions with compact support (well-defined in fact for $n > (2(b-a))^{-1}$):

$$f_n(x) = \begin{cases} 1 & \text{if } a \leqslant x \leqslant b \\ 0 & \text{if } x \leqslant a - 1/n \text{ or } x \geqslant b + 1/n \\ nx - (na-1) & \text{if } a - 1/n \leqslant x \leqslant a \\ -nx + (nb+1) & \text{if } b \leqslant x \leqslant b + 1/n \end{cases}$$

and

$$g_n(x) = \begin{cases} 1 & \text{if } a + 1/n \leqslant x \leqslant b - 1/n \\ 0 & \text{if } a \leqslant x \text{ or } x \geqslant b \\ nx - na & \text{if } a \leqslant x \leqslant a + 1/n \\ -nx + nb & \text{if } b - 1/n \leqslant x \leqslant b \end{cases}$$

(a graph of these functions will convey much more information than these dry formulas).

The definition implies immediately that

$$g_n \leqslant \chi_{[a,b]} \leqslant f_n$$

for $n > (2(b-a))^{-1}$, and after integrating with respect to μ, we derive the inequalities

$$\Lambda(g_n) = \int g_n d\mu \leqslant \mu([a,b]) \leqslant \int f_n d\mu = \Lambda(f_n)$$

for all n, using on the right and left the fact that integration of continuous functions is the same as applying Λ. In fact, the Riemann integrals of f_n and g_n can be computed very easily, and we derive

$$\Lambda(f_n) = (b-a) + \frac{1}{n} \text{ and } \Lambda(g_n) = (b-a) - \frac{1}{n},$$

so that $\mu([a,b]) = b-a$ follows after letting n go to infinity.

Since it is easy to check that $\mathbf{R}$ is σ-compact (see below where the case of $\mathbf{R}^d$ is explained), we derive in fact that the Lebesgue measure is the unique measure on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ which extends the length of intervals. We will see later another characterization of the Lebesgue measure which is related to this one.

The proof of this theorem is quite intricate; not only is it fairly technical, but there is quite a subtle point in the proof of the last part (unicity for σ-compact spaces). To understand the result, the first issue is to understand why an assumption like local compacity is needed. The point is that this is a way to ensure that $C_c(X)$ contains "many" functions. Intuitively, one requires $C_c(X)$ to contain (at least) sufficiently many functions to approximate arbitrarily closely the characteristic functions of nice sets in X, and this is not true for arbitrary topological spaces.

The precise existence results which are needed are the following:

Proposition 5.2.3 (Existence of continuous functions). *Let X be a locally compact topological space.*

(1) *For any compact set $K \subset X$, and any open neighbourhood V of K, i.e., with $K \subset V \subset X$, there exists $f \in C_c(X)$ such that*

$$\chi_K \leqslant f \preccurlyeq \chi_V \tag{5.3}$$

where the notation

$$f \preccurlyeq \chi_V$$

means that

$$\begin{cases} f \leqslant \chi_V \\ \operatorname{supp}(f) \subset V. \end{cases}$$

(2) *Let K_1 and K_2 be disjoint compact subsets of X. There exists $f \in C_c(X)$ such that $0 \leqslant f \leqslant 1$ and*

$$f(x) = \begin{cases} 0 & \text{if } x \in K_1 \\ 1 & \text{if } x \in K_2. \end{cases}$$

(3) *Let $K \subset X$ be a compact subset and let $V_1, \ldots, V_n$ be open sets such that*

$$K \subset V_1 \cup V_2 \cup \cdots \cup V_n.$$

For any function $g \in C_c(X)$, there exist functions $g_i \in C_c(X)$ such that $\operatorname{supp}(g_i) \subset V_i$ *for all i and*

$$\sum_{i=1}^{n} g_i(x) = g(x) \text{ for all } x \in K.$$

In addition, if $g \geqslant 0$, one can select g_i so that $g_i \geqslant 0$.

Remark 5.2.4. Note that it is possible that a function $f \in C_c(X)$ satisfies $f \leqslant \chi_V$, but without having $\operatorname{supp}(f) \subset V$ (for instance, take $X = [-2, 2]$, $V =]-1, 1[$ and

$$f(x) = \begin{cases} 0 & \text{if } |x| \geqslant 1, \\ 1 - x^2 & \text{if } |x| \leqslant 1 \end{cases}$$

for which $f \leqslant \chi_V$ but $\operatorname{supp}(f) = [-1, 1]$ is larger than V).

This is the reason for the introduction of the relation $f \preccurlyeq \chi_V$. Indeed, it will be quite important, at a technical point of the proof of the theorem, to ensure a condition of the type $\operatorname{supp}(f) \subset V$ (see the proof of Step 1 in the proof below).

Proof. These results are standard facts of topology. For (1), we recall only that one can give an easy construction of f when X is a metric space. Indeed, let $W \subset V$ be a relatively compact open neighbourhood of K, and let $F = X - W$ be the (closed) complement of W. We can then define

$$f(x) = \frac{d(x, F)}{d(x, F) + d(x, K)},$$

and check that it satisfies the required properties.

Part (2) can be deduced from (1) by applying the latter to $K = K_2$ and V any open neighbourhood of K_2 which is disjoint of K_1.

For (3), one shows first how to construct functions $f_i \in C_c(X)$ such that

$$0 \leqslant f_i \preccurlyeq \chi_{V_i},$$

and

$$1 = f_1(x) + \cdots + f_n(x)$$

for $x \in K$. The general statement follows by taking $g_i = g f_i$; we still have $g_i \preccurlyeq V_i$ of course, and also $g_i \geqslant 0$ if $g \geqslant 0$. ☐

Although the individual steps of the proof are all quite elementary and reasonable, the complete proof is somewhat lengthy and intricate. For simplicity, we will only consider the case where X is compact (so $C_c(X) = C(X)$), referring, e.g., to [**R**, Ch. 2] for the general case.

The general strategy is the following:

- Using Λ and the properties (5.1) and (5.2) as guide (which is a somewhat unmotivated way of proceeding), we will define a collection $\mathcal{M}$ of subsets of X, and a map
$$\mu : \mathcal{M} \to [0, +\infty]$$
though it will not be clear that either $\mathcal{M}$ is a σ-algebra, or that μ is a measure! However, similar ideas will show that a measure μ, if it exists, is unique among those which satisfy (5.1) and (5.2).
- We will show that $\mathcal{M}$ is a σ-algebra, that the open sets are in $\mathcal{M}$, and that μ is a measure on $(X, \mathcal{M})$. The regularity properties (5.1) and (5.2) of this measure μ will be valid essentially by construction.
- Next, we will show that
$$\Lambda(f) = \int_X f(x) d\mu(x) \tag{5.4}$$
for $f \in C(X)$. (Here we recall that we work with X compact).
- Finally, *using the results already proved*, we will show that *any* Borel measure μ on a σ-compact space which is finite on compact sets satisfies the regularity properties (5.1) and (5.2). Thus, the unicity under these additional conditions, which was already known, will imply the unicity without them...

As one can guess from this outline, it will be useful to introduce the following terminology:

Definition 5.2.5 (Radon measures). Let X be a locally compact topological space. A *Radon measure* on X is a measure μ, with respect to the Borel σ-algebra, which is finite on compact sets, and satisfies (5.1) for any Borel set $E \in \mathcal{B}$ (one says that μ is *outer-regular*), and (5.2) for any set E which is either open or has finite measure (one says that μ is *inner-regular*).

Thus, one can paraphrase the theorem of Riesz by saying that (1) it establishes a correspondance between non-negative linear forms on $C_c(X)$ and Radon measures on X; (2) if X is σ-compact, it says that any Borel measure finite on compact sets is a Radon measure.

We now come to the details; in a first reading, the next section might be only lightly skimmed...

5.3. Proof of the Riesz representation theorem

None of the steps in the outline above are obvious. But note that some hint on the way to proceed for the last points are clearly visible in the proof above that, applied to Λ coming from the Riemann integral, the measure that is obtained is precisely the Lebesgue measure.

We now start by the first point in the outline, before splitting the next two in a few steps the next two. The last (Part (3) of the theorem) will be dealt with at the very end.

First, to motivate the construction, we consider the question of uniqueness:

Uniqueness for μ among Radon measures. By (5.1), which holds for Radon measures, it is enough to show that the linear map Λ given by

$$\Lambda(f) = \int_X f(x) d\mu(x)$$

determines uniquely the measure $\mu(U)$ of any open subset $U \subset X$. The idea is that χ_U can be written as the pointwise limit of continuous functions (with support in U). More precisely, define

$$\mu^+(U) = \sup\{\Lambda(f) \mid 0 \leqslant f \preccurlyeq \chi_U\}.$$

We claim that $\mu^+(U) = \mu(U)$, which gives the required unicity. Indeed, we first have $\Lambda(f) \leqslant \mu(U)$ for any f appearing in the set defining $\mu^+(U)$, hence $\mu^+(U) \leqslant \mu(U)$. For the converse, we appeal to the assumed property (5.2) for U: we have

$$\mu(U) = \sup\{\mu(K) \mid K \subset U \text{ compact}\}.$$

Then consider any such compact set $K \subset U$; we must show that $\mu(K) \leqslant \mu^+(U)$, and Urysohn's Lemma is perfect for this purpose: it gives us a function $f \in C(X)$ with

$$0 \leqslant \chi_K \leqslant f \preccurlyeq \chi_U,$$

so that by integration we obtain

$$\mu(K) \leqslant \Lambda(f) \leqslant \mu(U),$$

as desired. □

This uniqueness result motivates (slightly) the general construction that we now attempt, to prove the existence of μ for a given linear form Λ. First we define a map μ^+ for *all* subsets of X (which is what is often called an *outer measure*), starting by writing the definition above for an open set U:

$$\mu^+(U) = \sup\{\Lambda(f) \mid f \in C(X) \text{ and } 0 \leqslant f \preccurlyeq \chi_U\}. \tag{5.5}$$

The use of the relation $f \preccurlyeq \chi_U$ instead of the more obvious $f \leqslant \chi_U$ is a subtle point (since, once μ has been constructed, it will follow by monotonicity that in fact

$$\mu(U) = \sup\{\Lambda(f) \mid f \leqslant \chi_U\}$$

for all open sets U); the usefulness of this is already suggested by the uniqueness argument.

We next define $\mu^+(E)$ using our goal to construct μ for which (5.1) holds: for any $E \subset X$, we let

$$\mu^+(E) = \inf\{\mu^+(U) \mid U \supset E \text{ is open}\}.$$

Note that it is clear that this definition coincides with the previous one when E is open (since we can take $U = E$ in the set defining the infimum).

This map is not a measure – it is defined for all sets, and usually there is no reasonable measure with this property. But we now define a certain subcollection of subsets of X, which are well-behaved in a way reminiscent of the definition of integrable functions in the Riemann sense. The definition is motivated, again, by the attemps to ensure that (5.2) holds: we first define

$$\mu^-(E) = \sup\{\mu^+(K) \mid K \subset E \text{ compact}\},$$

which corresponds to trying to approximate E by compact subsets, instead of using open super-sets. Then we define

$$\mathcal{M} = \{E \subset X \mid \mu^+(E) = \mu^-(E)\}.$$

Thus, we obtain, by definition, a map

$$\mu : \mathcal{M} \to [0, +\infty]$$

such that

$$\mu(E) = \mu^+(E) = \mu^-(E) \text{ for all } E \in \mathcal{M}.$$

Before proceeding, here are the few properties of this construction which are completely obvious;

- The map μ^+ is monotone: if $E \subset F$, we have
$$\mu^+(E) \leqslant \mu^+(F).$$
- Since X is compact, the constant function 1 is in $C(X)$ and is the only function appearing in the supremum defining $\mu^+(X)$; thus $\mu^+(X) = \Lambda(1)$, and by the above we get
$$0 \leqslant \mu^+(E) \leqslant \mu^+(X) = \Lambda(1)$$
for all $E \subset X$. In particular, μ is certainly finite on compact sets.
- Any compact set K is in $\mathcal{M}$ (because K is the largest compact subset inside K); in particular $\emptyset$, X are in $\mathcal{M}$, and obviously
$$\mu(\emptyset) = 0, \qquad \mu(X) = \Lambda(1).$$
- A set E is in $\mathcal{M}$ if, for any $\varepsilon > 0$, there exist a compact set K and an open set U, such that
$$K \subset E \subset U,$$
and

$$\mu^+(K) \geqslant \mu^+(E) - \varepsilon, \quad \mu^+(U) \leqslant \mu^+(E) + \varepsilon. \tag{5.6}$$

The next steps are the following five results:

(1) The "outer measure" μ^+ satisfies the countable subadditivity property
$$\mu^+\Bigl(\bigcup_{n\geqslant 1} E_n\Bigr) \leqslant \sum_{n\geqslant 1} \mu^+(E_n)$$
for all $E_n \subset X$.

(2) For K compact, we have the alternate formula

$$\mu(K) = \inf\{\Lambda(f) \mid \chi_K \leqslant f\}. \tag{5.7}$$

(3) Any open set is in $\mathcal{M}$.

(4) The collection $\mathcal{M}$ is stable under countable *disjoint* union, and
$$\mu\Bigl(\bigcup_{n\geqslant 1} E_n\Bigr) = \sum_{n\geqslant 1} \mu(E_n)$$
for a sequence of disjoint sets $E_n \in \mathcal{M}$.

(5) Finally, $\mathcal{M}$ is a σ-algebra (at which point we know that μ is a Borel measure on $(X, \mathcal{M})$ by the previous steps, and it is easy to check that it is a Radon measure by construction).

(6) The linear map Λ is the same as integration against the measure μ.

Once this is done, only the last uniqueness statement (Part (3)) of Theorem 5.2.1 will remain to be proved. We now complete these steps in turn. None is very hard, but the accumulation of small technical manipulations, which have to be done just right, is somewhat overwhelming at first.

Proof of Step 1. First of all, we show that if U_1 and U_2 are open subsets of X, and $U = U_1 \cup U_2$, we have

$$\mu^+(U) \leqslant \mu^+(U_1) + \mu^+(U_2). \tag{5.8}$$

For this, we use the definition. Let $g \in C(X)$ be such that $0 \leqslant g \preccurlyeq \chi_U$, and let $K = \mathrm{supp}(g)$, a compact subset of X. By Proposition 5.2.3, (3), there exist $g_1, g_2 \in C(X)$ such that

$$0 \leqslant g_i \preccurlyeq \chi_{U_i},$$

and $g = g_1 + g_2$. By linearity, we obtain

$$\Lambda(g) = \Lambda(g_1) + \Lambda(g_2) \leqslant \mu^+(U_1) + \mu^+(U_2),$$

and after taking the supremum over all g, we derive (5.8). Of course, an easy induction extends this subadditivity property to any finite union of open sets.

Now, let $E_n \subset X$ for $n \geqslant 1$. Fixing some $\varepsilon > 0$, we can, by definition, find open subsets

$$U_n \supset E_n$$

such that

$$\mu^+(U_n) \leqslant \mu^+(E_n) + \frac{\varepsilon}{2^n}$$

for all $n \geqslant 1$. Let now U be the union of the U_n, and f any function such that

$$0 \leqslant f \preccurlyeq \chi_U.$$

Since the support of f is contained in U,[3] it follows by compactness that there is some finite N such that

$$\mathrm{supp}(f) \subset U_1 \cup \cdots \cup U_N =: V.$$

We then have $f \preccurlyeq \chi_V$, and using the subadditivity for a finite union of open sets, we derive

$$\Lambda(f) \leqslant \mu^+(V) \leqslant \sum_{n=1}^{N} \mu^+(U_n) \leqslant \sum_{n\geqslant 1} \mu^+(U_n) \leqslant \sum_{n\geqslant 1} \mu^+(E_n) + \varepsilon.$$

This inequality holds for all f with $0 \leqslant f \preccurlyeq \chi_U$, and hence

$$\mu^+(U) \leqslant \sum_{n\geqslant 1} \mu^+(E_n) + \varepsilon,$$

and hence, using monotony of μ^+, we get

$$\mu^+\Bigl(\bigcup_{n\geqslant 1} E_n\Bigr) \leqslant \mu^+(U) \leqslant \sum_{n\geqslant 1} \mu^+(E_n) + \varepsilon,$$

from which the result follows by letting $\varepsilon \to 0$. ◻

Proof of Step 2. Let K be a compact set, and denote

$$\nu(K) = \inf\{\Lambda(f) \mid f \geqslant \chi_K\}.$$

We must show $\mu(K) = \nu(K)$, and we start by showing that $\mu(K) \leqslant \nu(K)$, i.e., that for any $f \geqslant \chi_K$, we have

$$\Lambda(f) \geqslant \mu(K) = \mu^+(K).$$

From the definition of μ^+, we need to construct open sets "close" to K. For this, we use the continuity of f. Indeed, for any α such that $0 < \alpha < 1$, we may consider

$$V_\alpha = \{x \mid f(x) > \alpha\} \subset X,$$

[3] At this point it is crucial to have imposed that $f \preccurlyeq \chi_U$ in the definition.

which is an open set containing K. Hence we have

$$\mu(K) \leqslant \mu^+(V_\alpha).$$

We now compare $\mu^+(V_\alpha)$ with $\Lambda(f)$. By definition, for any $\varepsilon > 0$, there exists $g \preccurlyeq \chi_{V_\alpha}$ such that

$$\mu^+(V_\alpha) \leqslant \Lambda(g) + \varepsilon,$$

and since $g \leqslant \chi_{V_\alpha}$ implies

$$\alpha g \leqslant \alpha \chi_{V_\alpha} \leqslant f,$$

we obtain by positivity that

$$\mu(K) \leqslant \mu^+(V_\alpha) \leqslant \Lambda(g) + \varepsilon \leqslant \Lambda(\alpha^{-1} f) + \varepsilon = \alpha^{-1}\Lambda(f) + \varepsilon.$$

We can now let $\alpha \to 1$, $\varepsilon \to 0$, and obtain

$$\mu(K) \leqslant \Lambda(f),$$

hence $\mu(K) \leqslant \nu(K)$.

Now, for the converse inequality, we have, for any $\varepsilon > 0$, some open set V such that $V \supset K$ and

$$\mu^+(V) \leqslant \mu(K) + \varepsilon.$$

By Proposition 5.2.3, (1), we can find a function $f \in C(X)$ such that

$$\chi_K \leqslant f \preccurlyeq \chi_V,$$

and it follows that

$$\nu(K) \leqslant \Lambda(f) \leqslant \mu^+(V) \leqslant \mu(K) + \varepsilon,$$

from which the equality (5.7) finally follows by letting $\varepsilon \to 0$. $\square$

Proof of Step 3. Let U be an open set. We must show that $\mu^+(U) \leqslant \mu^-(U)$. For this, let $\varepsilon > 0$ be given. By definition of $\mu^+(U)$, we can find f such that $f \preccurlyeq \chi_U$, in particular $f \leqslant \chi_U$, and

$$\Lambda(f) \geqslant \mu^+(U) - \varepsilon.$$

We must approach $\mu^+(U)$ from below by the measure of compact sets. The idea is that if $\Lambda(f)$ is very close to $\mu^+(U)$, the measure of the support of f should also be close to $\mu^+(U)$. So we consider the compact set $K = \mathrm{supp}(f)$. For any open set $W \supset K$, we have

$$f \preccurlyeq \chi_W,$$

and therefore $\Lambda(f) \leqslant \mu^+(W)$ by definition. Taking the infimum over W, we obtain

$$\Lambda(f) \leqslant \mu(K),$$

and then

$$\mu^+(U) \leqslant \Lambda(f) + \varepsilon \leqslant \mu(K) + \varepsilon,$$

and letting $\varepsilon \to 0$, it follows that (5.2) holds fo U, which means that $U \in \mathcal{M}$. $\square$

Proof of Step 4. We are now getting close. To prove the stability of $\mathcal{M}$ under disjoint countable unions and the additivity of μ for such unions, we first note that by Step 1, we have

$$\mu^+\Bigl(\bigcup_{n\geqslant 1} E_n\Bigr) \leqslant \sum_{n\geqslant 1} \mu^+(E_n) = \sum_{n\geqslant 1} \mu(E_n)$$

if $E_n \in \mathcal{M}$ for all n. So we need only show that

$$\mu^-\Bigl(\bigcup_{n\geqslant 1} E_n\Bigr) \geqslant \sum_{n\geqslant 1} \mu^-(E_n) = \sum_{n\geqslant 1} \mu(E_n)$$

to deduce both that the union E of the E_n is in $\mathcal{M}$, and that its measure is

$$\mu(E) = \sum_{n \geqslant 1} \mu(E_n).$$

The proof of the last inequality is very similar to that of Step 1, with compact sets replacing open sets, and lower bounds replacing upper-bounds. So we first consider the union $K = K_1 \cup K_2$ of two disjoint compact sets; this is again a compact set in X, and hence we know that $K \in \mathcal{M}$.

Fix $\varepsilon > 0$. By Step 2, we can find $g \in C(X)$ such that $\chi_K \preccurlyeq g$ and

$$\Lambda(g) \leqslant \mu(K) + \varepsilon.$$

Now we try to "share" the function g between K_1 and K_2. By Proposition 5.2.3, (2), there exists $f \in C(X)$ such that $0 \leqslant f \leqslant 1$, f is zero on K_1 and f is 1 on K_2. Then if we define

$$f_1 = (1 - f)g, \qquad f_2 = fg,$$

we have

$$\chi_{K_1} \leqslant f_1, \qquad \chi_{K_2} \leqslant f_2,$$

and hence by (5.7) and linearity, we get

$$\mu(K_1) + \mu(K_2) \leqslant \Lambda(f_1) + \Lambda(f_2) = \Lambda(f_1 + f_2) = \Lambda(g) \leqslant \mu(K) + \varepsilon,$$

and the desired inequality

$$\mu(K) \geqslant \mu(K_1) + \mu(K_2)$$

follows by letting $\varepsilon \to 0$.

Using induction, we also derive additivity of the measure for any finite union of disjoint compact sets. Now, for the general case, let $\varepsilon > 0$ be given. By (5.6), we can find compact sets $K_n \subset E_n$ (which are disjoint, since the E_n are) such that

$$\mu(K_n) \geqslant \mu(E_n) - \frac{\varepsilon}{2^n}$$

for $n \geqslant 1$. Then, for any finite $N \geqslant 1$, monotonicity and the case of compact sets gives

$$\mu^-(E) \geqslant \mu\Big(\bigcup_{n \leqslant N} K_n\Big) \geqslant \sum_{n=1}^{N} \mu(K_n) \geqslant \sum_{n=1}^{N} \mu(E_n) - \varepsilon,$$

and yet another limit, with $\varepsilon \to 0$ and $N \to +\infty$, leads to

$$\mu^-(E) \geqslant \sum \mu(E_n),$$

which was our goal. $\square$

Proof of Step 5. We already know that $\mathcal{M}$ contains the open sets (Step 3) and is stable under countable disjoint unions (Step 4). Now assume that we can show that, given E_1 and E_2 in $\mathcal{M}$, the difference

$$E = E_1 - E_2 = \{x \in E_1, \mid\ x \notin E_2\}$$

is also in $\mathcal{M}$. Then we will first deduce that $\mathcal{M}$ is stable under complement. Moreover, we will also get

$$E_1 \cup E_2 = (E_1 - E_2) \cup E_2 \in \mathcal{M},$$

which is a disjoint union, so that we get stability of $\mathcal{M}$ under finite unions (by induction from this case). Then, for any (E_n) in $\mathcal{M}$, we can write

$$\bigcup_{n \geqslant 1} E_n = \bigcup_{n \geqslant 1} F_n$$

where

$$F_n = E_n - \bigcup_{1 \leqslant j \leqslant n-1} E_j \in \mathcal{M},$$

and the F_n's are disjoint, leading to stability of $\mathcal{M}$ under any countable union. Complements give stability under intersection, and hence it will follow that $\mathcal{M}$ is a σ-algebra containing the Borel sets.

So we proceed to show that $E_1 - E_2$ is in $\mathcal{M}$ if E_1 and E_2 are. This is not difficult, because both the measure of E_1 and E_2 can be approximated both by compact or open sets, and their complements just exchange compact and open.

First, however, we note that if $E \in \mathcal{M}$ is arbitrary and $\varepsilon > 0$, we know that we can find U open and K compact, with

$$K \subset E \subset V,$$

and

$$\mu(V) - \varepsilon/2 < \mu(E) < \mu(K) + \varepsilon/2.$$

The set $V - K = V \cap (X - K)$ is then open, hence belongs to $\mathcal{M}$ by Step 3, and because of these inequalities it satisfies

$$\mu(V - K) < \varepsilon$$

(in view of the disjoint union $V = (V - K) \cup K$ which gives

$$\mu(V) = \mu(V - K) + \mu(K)$$

by the additivity of Step 4.)

Now, applying this to E_1, $E_2 \in \mathcal{M}$, with difference E, we get (for any $\varepsilon > 0$) open sets V_1 and V_2, and compact sets K_1 and K_2, such that

$$K_i \subset E_i \subset V_i \text{ and } \mu(V_i - K_i) < \varepsilon.$$

Then

$$E \subset (V_1 - K_2) \subset (K_1 - V_2) \cup (V_1 - K_1) \cup (V_2 - K_2),$$

so that by Step 1 again, we have

$$\mu^+(E) \leqslant \mu(K_1 - V_2) + 2\varepsilon.$$

Since $K_1 - V_2 \subset E$ is compact, we have

$$\mu^-(E) \geqslant \mu(K_1 - V_2) \geqslant \mu^+(F) - 2\varepsilon,$$

and letting $\varepsilon \to 0$ gives $E \in \mathcal{M}$. □

At this point, we know that $\mathcal{M}$ is a σ-algebra, and that it contains the Borel σ-algebra since it contains the open sets (Step 5 with Step 3), so that Step 4 shows that μ is a Borel measure. The "obvious" property $\mu(X) = \Lambda(1)$ measure shows that μ is finite on compact sets; the regularity property (5.1) is just the definition of μ^+ (which coincides with μ on $\mathcal{M}$), and (5.2) is the definition of μ^-. Hence μ is indeed a Radon measure. Now the inner regularity also ensures that μ is complete: if $E \in \mathcal{M}$ satisfies $\mu(E) = 0$ and F is any subset of E, then any compact subset K of F is also a compact subset of E, hence satisfies

$$\mu(K) \leqslant \mu(E) = 0,$$

leading to $\mu^-(F) = 0 = \mu^+(F)$.

Therefore, the proof of Parts (1) and (2) of the Riesz Representation Theorem will be concluded after the last step.

Proof of Step 6. Since μ is a finite Borel measure, we already know that any $f \in C(X)$ is integrable with respect to μ. Now, to prove that (5.4) holds for $f \in C(X)$, we can obviously assume that f is real valued, and in fact it suffices to prove the one-sided inequality

$$\Lambda(f) \leqslant \int_X f(x) d\mu(x), \tag{5.9}$$

since linearity of Λ we lead to the converse inequality by applying this to $-f$:

$$\Lambda(f) = -\Lambda(-f) \geqslant -\int_X -f(x) d\mu(x) = \int_X f(x) d\mu(x).$$

Moreover, since

$$\Lambda(1) = \mu(X) = \int d\mu,$$

we can assume $f \geqslant 0$ (replacing f, if needed, by $f + \|f\|_\infty \geqslant 0$.) Writing then $M = \|f\|_\infty$, we have

$$f(x) \in [0, M]$$

for all $x \in X$.

Trying to adapt roughly the argument used in the case of Riemann integration, we want to bound f by (continuous versions of) step functions converging to f. For this, denote $K = \operatorname{supp}(f)$ and consider $n \geqslant 1$; we define the sets

$$E_i = f^{-1}\Big(\Big] \frac{i}{n}, \frac{i+1}{n} \Big] \Big) \cap K$$

for $-1 \leqslant i \leqslant Mn$.

These sets are in $\mathcal{B}_X$ since f is measurable, hence in $\mathcal{M}$. They are also disjoint, and of course they cover K. Now, for any $\varepsilon > 0$, we can find open sets

$$V_i \supset E_i$$

such that

$$\mu(V_i) \leqslant \mu(E_i) + \varepsilon,$$

by (5.1), and we may assume that

$$V_i \subset f^{-1}\Big(\Big] \frac{i}{n}, \frac{i+1}{n} + \varepsilon \Big[\Big),$$

by replacing V_i, if needed, by the open set

$$V_i \cap f^{-1}\Big(\Big] \frac{i}{n}, \frac{i+1}{n} + \varepsilon \Big[\Big).$$

Now we use Proposition 5.2.3, (3) to construct functions $g_i \preccurlyeq \chi_{V_i}$ such that

$$\sum g_i(x) = 1$$

for all $x \in K$. Consequently, since $\operatorname{supp}(f) = K$, we get

$$\Lambda(f) = \Lambda\Big(\sum_i f g_i \Big) = \sum_i \Lambda(f g_i) \leqslant \sum_i \Big(\frac{i+1}{n} + \varepsilon \Big) \Lambda(g_i)$$

since

$$0 \leqslant f g_i \leqslant ((i+1)/n + \varepsilon) g_i,$$

and thus

$$\Lambda(f) \leqslant \sum_i \Big(\frac{i+1}{n} + \varepsilon \Big) (\mu(E_i) + \varepsilon)$$

since $\Lambda(g_i) \leqslant \mu(V_i) \leqslant \mu(E_i) + \varepsilon$. Letting, $\varepsilon \to 0$, we find

$$\Lambda(f) \leqslant \sum_i \frac{i+1}{n}\mu(E_i) = \int_X s_n(x)d\mu(x) + \frac{1}{n}\sum_i \mu(E_i)$$
$$\leqslant \int_X s_n(x)d\mu(x) + \frac{\mu(K)}{n},$$

where s_n is the step function

$$s_n = \sum_i \frac{i}{n}\chi_{E_i}.$$

Obviously, we have $s_n \leqslant f$, and finally we derive

$$\Lambda(f) \leqslant \int_X f(x)d\mu(x) + \frac{\mu(K)}{n},$$

for all $n \geqslant 1$, which leads to (5.9) once n tends to infinity... □

Remark 5.3.1. In the case of constructing the Lebesgue measure on $[0,1]$, this last Step 6 may be omitted, since we have already checked in that special case that the measure obtained from the Riemann integral satisfies

$$\mu([a,b]) = b - a$$

for $a < b$.

We now come to the proof of Part (3). The subtle point is that it will depend on an application of what has already been proved...

Proposition 5.3.2 (Radon measures on σ-compact spaces). *Let X be a σ-compact topological space, and μ any Borel measure on X finite on compact sets. Then μ is a Radon measure. Additionally, the inner regularity* (5.2) *holds for any $E \in \mathcal{B}$, not only for those E which are open or have finite measure.*

Proof. Since μ is finite on compact sets, we can consider the linear map

$$\Lambda : f \mapsto \int_X f(x)d\mu(x)$$

on $C_c(X)$.

As we observed in great generality in Proposition 5.1.2, this is a well-defined non-negative linear map. Hence, according to Part (2) of the Riesz Representation Theorem – which we already proved –, there exists a *Radon* measure ν on X such that

$$\int_X f(x)d\mu(x) = \Lambda(f) = \int_X f(x)d\nu(x)$$

for any $f \in C_c(X)$. Now we will simply show that $\mu = \nu$, which will give the result.

For this, we start by the proof that inner regularity for ν holds for all Borel subsets (because X is σ-compact; note that if X is compact, this step is not needed as there are no Borel set with infinite measure). Indeed, we can then write

$$X = \bigcup_{n \geqslant 1} K_n \text{ with } K_n \text{ compact},$$

and we can assume $K_n \subset K_{n+1}$ (replacing K_n, if needed, with the compact set $K_n \cup K_{n-1} \cup \cdots \cup K_1$). Then if $E \in \mathcal{B}$ has infinite measure, we have

$$\lim_{n \to +\infty} \nu(E \cap K_n) = \nu(E) = +\infty.$$

Given any $N > 0$, we can find n such that $\nu(E \cap K_n) \geqslant N$. Then (5.2), applied to $E \cap K_n$, shows that there exists a compact set $K \subset E \cap K_n \subset E$ with

$$\nu(K) \geqslant \nu(E \cap K_n) - 1 \geqslant N - 1,$$

and hence

$$\sup\{\nu(K) \mid K \subset E\} = +\infty,$$

which is (5.2) for E.

Once this is out of the way, we have the property

$$\int_X f d\mu(x) = \int_X f d\nu(x) \text{ for all } f \in C_c(X). \tag{5.10}$$

We will first show that

$$\mu(V) = \nu(V) \tag{5.11}$$

if $V \subset X$ is open.

For this, we use the σ-compactness of X to write

$$V = \bigcup_{n \geqslant 1} K_n$$

where K_n is compact for all n and $K_n \subset K_{n+1}$. Using Proposition 5.2.3, (1), we can find functions $f_n \in C_c(X)$ such that

$$\chi_{K_n} \leqslant f_n \leqslant \chi_{K_{n+1}}$$

for all n, and

$$0 \leqslant f_n \leqslant f_{n+1}$$

for all n, and moreover

$$\lim_{n \to +\infty} f_n(x) = \chi_V(x)$$

for all $x \in X$: indeed, all these quantities are 0 for $x \notin V$, and otherwise we have $x \in K_n$ for all n large enough, so that $f_n(x) = 1$ for all n large enough. We now apply the monotone convergence theorem for both μ and for ν, and we obtain

$$\mu(V) = \lim_{n \to +\infty} \int_X f_n(x) d\mu(x) = \lim_{n \to +\infty} \int_X f_n(x) d\nu(x) = \nu(V),$$

which gives (5.11).

Now we extend this to all Borel subsets. First, if $E \subset X$ is a Borel set such that $\nu(E) < +\infty$, and $\varepsilon > 0$, we can apply the regularity of ν to find K and V with

$$K \subset E \subset V,$$

and K is compact, V is open, and $\nu(V - K) < \varepsilon$.

However, since $V - K \subset X$ is also open, it follows that we also have

$$\mu(V - K) = \nu(V - K) < \varepsilon,$$

and hence – using monotonicity and (5.11), – we get the inequalities

$$\mu(E) \leqslant \mu(V) = \nu(V) \leqslant \nu(E) + \varepsilon$$

and[4]

$$\nu(E) \leqslant \nu(V) = \mu(V) \leqslant \mu(E) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this gives $\mu(E) = \nu(E)$.

[4] Write the disjoint union $V = (V - E) \cup E$ to get $\mu(V) = \mu(V - E) + \mu(E)$ and $\mu(V - E) \leqslant \mu(V - K) < \varepsilon$ since $K \subset E$.

Finally, when $E \in \mathcal{M}$ has infinite measure, we just observe that regularity for ν and E shows that, for any $N \geqslant 1$, we can find a compact subset $K \subset E$ with $\nu(K) > N$; since $\mu(K) = \nu(K)$ by the previous case, we get $\mu(E) = +\infty = \nu(E)$ by letting $N \to +\infty$. $\square$

Note that this proof has shown the following useful fact:

Corollary 5.3.3 (Integration determines measures). *Let X be a σ-compact topological space, and let μ_1 and μ_2 be Borel measures on X finite on compact sets, i.e., Radon measures on X. If we have*

$$\int_X f(x)d\mu_1(x) = \int_X f(x)d\mu_2(x) \text{ for all } f \in C_c(X),$$

then $\mu_1 = \mu_2$.

Proof. This is contained in the previous proof (with μ and ν instead of μ_1 and μ_2, starting from (5.10)). $\square$

The following is also useful:

Corollary 5.3.4. *Let X be a σ-compact topological space, μ_1 and μ_2 two Borel measures on X finite on compact sets. Let*

$$\mathcal{K} = \{E \in \mathcal{B} \mid \mu_1(E) = \mu_2(E)\}.$$

(1) *If $\mathcal{K}$ contains all compact sets in X, then $\mu_1 = \mu_2$.*
(2) *If $\mathcal{K}$ contains all open subsets in X, then $\mu_1 = \mu_2$.*

Proof. Since μ_1 and μ_2 are Radon measures and (5.2) holds for all $E \in \mathcal{B}$ by the lemma above, we can use (5.2) in case (1), or (5.1) in case (2). $\square$

In Section 5.6, we will derive more consequences of the uniqueness statement in the Riesz Representation Theorem, but before this we derive some important consequences of the first two parts.

5.4. Approximation theorems

Consider a Borel measure μ on X, finite on compact sets. Then the compactly-supported functions are not only integrable and bounded, their p-th power is also integrable for all $p \geqslant 1$. Thus, continuous functions provide a subset of the L^p-spaces $L^p(X, \mu)$, for any $p \geqslant 1$ (including $p = +\infty$). As one can imagine, the continuity of these functions makes them fairly well-behaved compared with arbitrary measurable functions, and it is natural to ask "how large" is the space they generate in the L^p spaces. It is very convenient that this space is *dense* for the L^p-norm, if $p < +\infty$.

Theorem 5.4.1 (Density of continuous functions in L^p). *Let X be a locally compact space, μ a Radon measure on X.*

(1) *For any $p \in [1, +\infty[$, the image of $C_c(X)$ in $L^p(\mu)$ is dense in $L^p(\mu)$ for the L^p-norm, i.e., for any $f \in L^p(X, \mu)$ and any $\varepsilon > 0$, there exists $g \in C_c(X)$ such that*

$$\Big(\int_X |f(x) - g(x)|^p d\mu(x)\Big)^{1/p} < \varepsilon.$$

(2) *For $p = +\infty$, the closure of $C_c(X)$ in $L^\infty(X)$, with respect to the L^∞-norm, is contained in the image of the space $C_b(X)$ of bounded continuous functions.*

Remark 5.4.2. Most often, the closure of $C_c(X)$ in L^∞ is not equal $C_b(X)$. For instance, when $X = \mathbf{R}^d$, the closure of $C_c(\mathbf{R}^d)$ in $L^\infty(\mathbf{R}^d)$ is the space $C_0(\mathbf{R}^d)$ of continuous functions such that

$$\lim_{\|x\|\to+\infty} f(x) = 0$$

(functions going to 0 at infinity). Indeed, consider a sequence (f_n) of functions in $C_c(\mathbf{R}^d)$, converging uniformly to f. We know that f is continous. Moreover, we can write

$$|f(x)| \leqslant |f(x) - f_n(x)| + |f_n(x)|$$

for any x and any $n \geqslant 1$, and given an arbitrariy $\varepsilon > 0$, we find first n_0 such that

$$\|f - f_{n_0}\|_\infty < \varepsilon$$

from which it follows that

$$|f(x)| \leqslant \varepsilon + |f_{n_0}(x)|$$

for all x, and then, since $f_{n_0}(x) = 0$ if $x \notin \mathrm{supp}(f_n)$, we find that $|f(x)| < \varepsilon$ for all x such that $\|x\|$ is large enough; hence the limit f is in $C_0(\mathbf{R}^d)$.

Conversely, given $f \in C_0(\mathbf{R}^d)$, we can approach it arbitrarily closely by a sequence of functions in $C_c(\mathbf{R}^d)$ of the type

$$f_n = f g_n$$

where $g_n \in C_c(\mathbf{R}^d)$ satisfies

$$\chi_{[-n,n]^d} \leqslant g_n \leqslant \chi_{[-2n,2n]^d},$$

Indeed, we have

$$\|f - f_n\|_\infty \leqslant \sup\{|f(x)| \mid \|x\| \leqslant n\} \to 0.$$

Proof of Theorem 5.4.1. Part (2) is simply the expression of the fact that a uniformly convergent sequence of continuous functions has a limit which is continuous (and that any convergent sequence in a normed vector space is bounded).

To prove (1), the idea is to reduce the problem to the approximation of characteristic functions of measurable sets, and for the latter to appeal to the regularity conditions (5.1) and (5.2).

More precisely, let us denote by V the $\mathbf{C}$-vector space which is the closure of the image of $C_c(X)$ in $L^p(\mu)$. By Lemma 5.4.3 below, it follows that V contains all functions of the type $f = \chi_E$, where $E \subset X$ is a measurable set of finite measure (these functions are obviously in $L^p(\mu)$ for $p < +\infty$).

Now by linearity, we deduce that V contains all non-negative step functions which are in $L^p(\mu)$. Consider now $f \geqslant 0$ such that $f \in L^p(\mu)$, and let (s_n) be, as usual, a non-decreasing sequence of non-negative step functions such that

$$s_n(x) \to f(x)$$

for all $x \in X$ (as given by Proposition 2.2.4). Since $s_n \leqslant f$, we have also $s_n \in L^p(\mu)$, and we now prove that the convergence $s_n \to f$ holds also in $L^p(\mu)$. From this, using linearity again, it follows finally that $V = L^p(\mu)$, which is the conclusion we want.

Let $g_n = |f - s_n|^p$. We have $g_n(x) \to 0$ for all x, and we want to apply the dominated convergence theorem. Using the inequality

$$(a + b)^p \leqslant 2^{p/q}(|a|^p + |b|^p)$$

(where $p^{-1} + q^{-1} = 1$, this being a "trivial" version of the Hölder inequality), we get

$$|g_n| \leqslant 2^{p/q}(|f|^p + |s_n|^p) \leqslant 2^{1+p/q}|f|^p \in L^1(\mu),$$

so that we can indeed apply the dominated convergence theorem, which gives

$$\|f - s_n\|_p^p = \int_X |g_n| d\mu \to 0$$

as desired. $\square$

Here is the technical lemma we used:

LEMMA 5.4.3. *Let X be locally compact, μ a Radon measure on X and $p \in [1, +\infty[$. For any $E \subset X$ with finite measure, and any $\varepsilon > 0$, there exists $f \in C_c(X)$ such that*

$$\|f - \chi_E\|_p < \varepsilon.$$

PROOF. Given $\varepsilon > 0$, the regularity of Radon measures (5.2) and (5.1) shows that there exist $K \subset E$ compact, $U \supset E$ open, such that

$$\mu(U - K) < \varepsilon.$$

Furthermore, by Urysohn's lemma, there exists f continuous on X such that

$$\chi_K \leqslant f \leqslant \chi_U$$

and $\mathrm{supp}(f) \subset U$ (what we denoted $f \preccurlyeq \chi_U$ in the proof of Riesz's Theorem). In particular, f is compactly supported. We then have

$$\int_X |f - \chi_E|^p d\mu \leqslant \int_{U-K} |f - \chi_E|^p d\mu \leqslant \mu(U - K) < \varepsilon$$

since $|f - \chi_E| \leqslant 1$ on X and f coincides with χ_E both U and inside K. Thus we get

$$\|f - \chi_E\|_p < \varepsilon^{1/p},$$

and the result follows (changing ε into ε^p). $\square$

REMARK 5.4.4. If $X = \mathbf{R}$, or an open subset of $\mathbf{R}^d$ more generally, one can also ask about more regularity than continuity for the L^p space. For instance, the following is easy to show:

PROPOSITION 5.4.5. *Let μ be a Radon measure on $\mathbf{R}$, and let $p \in [1, +\infty[$. For any $k \geqslant 1$, the image of the space $C_c^k(\mathbf{R}^d)$ of C^k functions with compact support is dense in $L^p(\mu)$.*

PROOF. Using the previous theorem, it is enough to approximate continuous functions with compact support using C^k functions. Let $f \in C_c(\mathbf{R})$ be given, and let $K = \mathrm{supp}(f)$. Using the Weierstrass approximate theorem, the function f is uniform limit, on K, of polynomials f_n (see Theorem 4.2.4). In $L^p(K, \mu)$, we then have also $f_n \to f$. However, this does not extend to $L^p(\mathbf{R}, \mu)$ because the polynomials typically are not integrable "at infinity", nor compactly supported. To work around this, let $N \geqslant 0$ be such that $K \subset [-N, N]$, and consider

$$g_n = f_n \varphi$$

where φ is a function in $C_c^k(\mathbf{R})$ which satisfies

$$\chi_{[-N,N]} \leqslant \varphi \leqslant \chi_{[-2N,2N]},$$

We still have $g_n \to f$ uniformly, but also $g_n \to f$ in $L^p(\mathbf{R}, \mu)$.

The explicit construction of φ is not particularly difficult (one must connect the "top" on $[-N,N]$ where it should be equal to 1 to the "bottom" outside $[-2N,2N]$ where it vanishes, with C^k connections); here is one possible solution:

$$\varphi(x)=\begin{cases}1 & \text{if } -N\leqslant x\leqslant N\\ \Big(1-\Big(\dfrac{x}{N}-1\Big)^k\Big)^k & \text{if } N\leqslant x\leqslant 2N\\ \Big(1-(-1)^k\Big(\dfrac{x}{N}+1\Big)^k\Big)^k & \text{if } -2N\leqslant x\leqslant -N\\ 0 & \text{if } |x|\geqslant 2N.\end{cases}$$

□

In the next chapter (in particular Corollary 6.4.7), even stronger versions will be proved, at least when μ is the Lebesgue measure.

5.5. Simple applications

We now present two simple but important applications of the approximation theorem (Theorem 5.4.1. The idea is the same in both cases: given a property to check for all $f\in L^p$, it suffices to do so when f is continuous and compactly supported, provided the property in question is "continuous" with respect to the L^p-norm.

Proposition 5.5.1 (Continuity of translation operators). *Let $d\geqslant 1$ be an integer and let $p\in[1,+\infty[$. For any $f\in L^p(\mathbf{R}^d)$, we have*

$$\lim_{h\to 0}\int_{\mathbf{R}^d}|f(x+h)-f(x)|^p d\lambda_n(x)=0. \tag{5.12}$$

Here $L^p(\mathbf{R}^d)$ refers, of course, to the Lebesgue measure on $\mathbf{R}^d$.

Remark 5.5.2. (1) This statement is not obvious at all, because the function which is being integrated, namely

$$x\mapsto f(x+h)-f(x)$$

if $p=1$, has no reason to be small when $h\neq 0$, since f is not continuous but merely integrable. Indeed, the fact that the statement fails for $p=+\infty$ shows that this result is only an average statement for the differences $f(x+h)-f(x)$. For instance, consider $f=\chi_{[0,1]}\in L^\infty(\mathbf{R})$. Then, for any $h\neq 0$, we have

$$\sup\{|f(x+h)-f(x)|\}=1,$$

(e.g., because $f(0)-f(-h)=1$ if $h>0$).

(2) Here is an abstract formulation (for readers familiar with the basic facts of functional analysis) which gives some insight into the meaning of the result. For any $h\in\mathbf{R}$, one can consider the translation operator defined by

$$T_h\ \begin{cases}L^p(\mathbf{R}^d)\to L^p(\mathbf{R}^d)\\ f\mapsto (x\mapsto f(x+h)).\end{cases}$$

It follows from the invariance of Lebesgue measure under translation (that we will prove below in Theorem 5.6.1) that

$$\|T_h(f)\|_p=\|f\|_p$$

for $f\in L^p(\mu)$, so that T_h is an isometry (and in particular, it is continuous). Moreover $T_{h+j}=T_h\circ T_j$.

Now the proposition may be translated to the assertion that the map

$$\rho : h \mapsto T_h$$

is itself continous at 0, with respect to the operator norm on the space of linear maps. Indeed, we have $T_0 = \mathrm{Id}$, and one may write (5.12) in the form

$$\lim_{h\to 0} \|f - T_h(f)\|_p = 0,$$

for any $f \in L^p(\mathbf{R}^d)$. The implication $\lim T_h = T_0$ is then a consequence of the Banach-Steinhaus theorem of functional analysis.

Proof. Notice first that, given $f \in L^p(\mathbf{R}^d)$ and a fixed real number h, the function

$$g(x) = f(x+h)$$

is well-defined in L^p (changing f on a set of measure zero also changes g on a set of measure zero). Also, we have $g \in L^p(\mathbf{R}^d)$ (and indeed $\|g\|_p = \|f\|_p$, e.g., by the change of variable formula). Thus the integrals in (5.12) exist for all $h \in \mathbf{R}$.

Now we start by proving the result when $f \in C_c(\mathbf{R}^d)$. Let

$$\psi(h) = \int_{\mathbf{R}^d} |f(x+h) - f(x)|^p d\lambda_n(x) = \int_{\mathbf{R}^d} \varphi(x,h) d\lambda_n(x)$$

where

$$\varphi(x,h) = |f(x+h) - f(x)|^p$$

for $(x,h) \in \mathbf{R}^d \times]-1,1[^d$. When f is continuous, the function φ is continuous at $h = 0$ for any fixed x. Moreover, we have

$$0 \leqslant \varphi(x,h) \leqslant 2^{p/q}(|f(x+h)|^p + |f(x)|^p) \leqslant 2^{p/q}\|f\|_\infty^p \chi_Q(x)$$

where χ_Q is the characteristic function of a compact set Q large enough so that

$$-h + \mathrm{supp}(f) \subset Q$$

for $h \in]-1,1[^d$ (if $\mathrm{supp}(h) \subset [-A,A]^d$, one may take for instance $Q = [-A-1, A+1]^d$).

Since Q is compact, the function χ_Q is integrable. We can therefore apply Proposition 3.1.1 and deduce that ψ is continuous at $h = 0$. Since $\psi(0) = 0$, this leads exactly to the result.

Now we apply the approximation theorem. Let $f \in L^p(\mu)$ be any function, and let $\varepsilon > 0$ be arbitrary. By approximation, there exists $g \in C_c(X)$ such that

$$\|f - g\|_p < \varepsilon,$$

and we then have

$$|f(x+h) - f(x)|^p \leqslant 3^{p/q}(|f(x+h) - g(x+h)|^p + |g(x+h) - g(x)|^p + |g(x) - f(x)|^p),$$

so that

$$\int_{\mathbf{R}^d} |f(x+h) - f(x)|^p d\lambda_n(x) \leqslant 3^{p/q}\Big(2\|f-g\|_p^p + \int_{\mathbf{R}^d} |g(x+h) - g(x)|^p d\lambda_n(x)\Big).$$

Since g is continuous with compact support, we get from the previous case that

$$0 \leqslant \limsup_{h\to 0} \int_{\mathbf{R}^d} |f(x+h) - f(x)|^p d\lambda_n(x) \leqslant 3^{1+p/q}\varepsilon^p,$$

and since $\varepsilon > 0$ is arbitrary, the result follows. ☐

The next application concerns the Fourier transform. Recall (see Section 3.2) that, for $f \in L^1(\mathbf{R})$, we have defined the Fourier transform $\hat{f} : \mathbf{R} \to \mathbf{C}$ by the formula

$$\hat{f}(t) = \int_{\mathbf{R}} f(y)e(-yt)dy$$

where $e(z) = e^{2i\pi z}$ for $z \in \mathbf{C}$. We showed that $\hat{f}$ is a bounded continuous function (Proposition 3.2.3). In fact, more is true:

THEOREM 5.5.3 (Riemann-Lebesgue lemma). *Let $f \in L^1(\mathbf{R})$. Then the Fourier transform of f goes to zero at infinity, i.e., we have*

$$\lim_{t \to \pm\infty} \hat{f}(t) = 0.$$

We give three different (but related...) proofs of this fact, to illustrate various techniques. The common feature is that the approximation theorem is used (implicitly or explicitly).

FIRST PROOF OF THE RIEMANN-LEBESGUE LEMMA. We write

$$\hat{f}(t) = \int_{\mathbf{R}} f(x)e(-xt)dx = -\int_{\mathbf{R}} f(x)e\Bigl(-t\Bigl(x + \frac{1}{2t}\Bigr)\Bigr)dx = -\int_{\mathbf{R}} f\Bigl(y - \frac{1}{2t}\Bigr)e(-yt)dy$$

(because $e(1/2) = e^{i\pi} = -1$, by the change of variable $x = y - (2t)^{-1}$). Hence we have also

$$\hat{f}(t) = \frac{1}{2}\int_{\mathbf{R}} \Bigl(f(x) - f\Bigl(x - \frac{1}{2t}\Bigr)\Bigr)e(-xt)dx,$$

from which we get

$$|\hat{f}(t)| \leqslant \int_{\mathbf{R}} \Bigl|f(x) - f\Bigl(x + \frac{1}{2t}\Bigr)\Bigr|dx.$$

As $t \to \pm\infty$, we have $1/(2t) \to 0$, and hence, applying (5.12) with $p = 1$, we get

$$\lim_{t \to \pm\infty} \hat{f}(t) = 0.$$

$\square$

SECOND PROOF. We reduce more directly to regular functions. Precisely, assume first that $f \in L^1(\mathbf{R})$ is compactly support and of C^1 class. By Proposition 3.2.3, (3), we get

$$-2i\pi t\hat{f}(t) = \hat{f'}(t)$$

for $t \in \mathbf{R}$, and this shows that the continuous function

$$g(t) = |t|\hat{f}(t)$$

is bounded on $\mathbf{R}$, which of course implies the conclusion

$$\lim_{t \to \pm\infty} \hat{f}(t) = 0$$

in that case – indeed, in a rather stronger quantitative form.

Now, we know that $C_c^1(\mathbf{R})$ is dense in $L^1(\mathbf{R})$ for the L^1-norm (by Proposition 5.4.5), and we proceed as before: given any function $f \in L^1(\mathbf{R})$ and any $\varepsilon > 0$, there exists $g \in C_c^1(\mathbf{R})$ such that

$$\|f - g\|_1 < \varepsilon,$$

and it follows that

$$|\hat{f}(t)| \leqslant |\hat{g}(t)| + \int_{\mathbf{R}} |f(y) - g(y)|dy \leqslant |\hat{g}(t)| + \varepsilon$$

for any $t \in \mathbf{R}$. Letting $t \to \pm\infty$, this gives

$$\limsup_{t\to\pm\infty} |\hat{f}(t)| \leqslant \varepsilon,$$

and again the result is obtained by letting $\varepsilon \to 0$. Or one can conclude by noting that the Fourier transform map is continuous from L^1 to the space of bounded continuous functions, and since the image of the dense subspace of C^1 functions with compact support lies in the closed (for the uniform convergence norm) subspace $C_0(\mathbf{R})$ of $C_b(\mathbf{R})$, the image of the whole of L^1 must be in this subspace. ☐

Third proof. This time, we start by looking at $f = \chi_{[a,b]}$, the characteristic function of a compact interval. In that case, the Fourier transform is easy to compute: for $t \neq 0$, we have

$$\hat{f}(t) = \int_a^b e(-xt)dx = -\frac{1}{2i\pi t}(e(-bt) - e(-at)),$$

and therefore

$$|\hat{f}(t)| \leqslant (\pi t)^{-1} \to 0$$

as $|t| \to +\infty$. Using linearity, we deduce the same property for finite linear combinations of such characteristic functions.

Now let $f \in C_c(\mathbf{R})$ be given, and let $K = \mathrm{supp}(f) \subset [-A, A]$ for some $A > 0$. Since f is uniformly continuous on $[-A, A]$, one can write it as a uniform limit (on $[-A, A]$ first, then on $\mathbf{R}$) of such finite combinations of characteristic functions of compact intervals. The same argument as the one used in the second proof then proves that the Riemann-Lebesgue lemma holds for f, and the general case is obtained as before using the approximation theorem. ☐

5.6. Application of uniqueness properties of Borel measures

In this section, we will use the unicity properties of Radon measures in the case of $X = \mathbf{R}^n$ to give a very useful characterization of Lebesgue measure in terms of invariant properties. In turn, this will lead to a fairly direct proof of the change of variable formula in $\mathbf{R}^n$ that was stated in the previous chapter.

Theorem 5.6.1 (Invariance of Lebesgue measure under translation). *Let $n \geqslant 1$ be an integer.*

(1) *The Lebesgue measure λ_n on $\mathbf{R}^n$ is invariant under translations, i.e., for any Borel set $E \subset \mathbf{R}^n$ and any $t \in \mathbf{R}^n$, we have*

$$\lambda_n(\{y \in \mathbf{R}^n \mid y = x + t \text{ for some } x \in E\}) = \lambda_n(E).$$

Equivalently, if

$$\tau_t : x \mapsto x + t$$

denotes the translation by t map on $\mathbf{R}^n$, we have

$$\tau_{t,*}(\lambda_n) = \lambda_n$$

for all $t \in \mathbf{R}^n$.

(2) *Conversely, let μ be a Borel measure on $\mathbf{R}^n$ such that μ is finite on compact sets and μ is invariant under translation. Then we have*

$$\mu(E) = c\lambda_n(E)$$

for all $E \subset \mathbf{R}^n$ measurable, where $c \geqslant 0$ is a constant, namely $c = \mu([0, 1]^n)$.

As a preliminary, we recall the following important fact:

LEMMA 5.6.2 (σ-compacity in $\mathbf{R}^n$). *For $n \geqslant 1$, the space $\mathbf{R}^n$ with the usual topology is σ-compact.*

PROOF. Indeed, given an open set $U \subset \mathbf{R}^n$, let $\mathcal{K}$ be the collection of all closed balls contained in U which have finite rational radius $r > 0$ and rational center $x_0 \in \mathbf{Q}^n$; this is a countable collection of compact sets, and we have the countable union

$$U = \bigcup_{K \in \mathcal{K}} K$$

since U is open: for any $x \in U$, we can find an open ball $V = B(x, r)$ with center x and radius $r > 0$ such that $x \in V$; then given a positive rational radius $r_0 < r/2$ and a point $x_0 \in \mathbf{Q}^n \cap B(x, r_0)$, the closed ball $\bar{B}(x_0, r_0)$ is included in $B(x, r)$, hence in U, by the triangle inequality, and therefore it is an element of $\mathcal{K}$ such that $x \in \bar{B}(x_0, r_0)$. □

PROOF. (1) Let $t \in \mathbf{R}^n$ be given, and let $\mu = \tau_{t,*}(\lambda)$ be the image measure. This is a Borel measure on $\mathbf{R}^n$, and since τ_t is a homeomorphism, it is finite on compact sets (the inverse image of a compact being still compact). Thus (by Proposition 5.3.4), we need only check that $\mu(U) = \lambda(U)$ when $U \subset \mathbf{R}^n$ is open.

For this purpose, we first consider the case $n = 1$. Then, we know that U is the union of (at most) countably many open intervals, which are its connected components:

$$U = \bigcup_{n \geqslant 1}]a_n, b_n[$$

with $a_n < b_n$ for all $n \geqslant 1$. Since translating an interval gives another interval of the same length, the property $\lambda(\tau_{t,*}(U))$ is clear in that case by additivity of the measures.

Now, for $n \geqslant 2$, we proceed by induction on n using Fubini's Theorem. We first observe that the translation τ_t can be written as a composition of n translations in each of the n independent coordinate directions successively; since composition of measures behaves as expected (i.e., $(f \circ g)_* = f_* \circ g_*$), we can assume that t has a single non-zero component, and because of the symmetry, we may as well suppose that

$$t = (t_1, 0, \ldots, 0)$$

for some $t_1 \in \mathbf{R}$. We then have

$$\mu(U) = \int_{\mathbf{R}^{n-1}} \Big(\int_{\mathbf{R}} f(u + t_1, x) d\lambda_1(u) \Big) d\lambda_{n-1}(x)$$

for any $U \subset \mathbf{R}^n$, where $f(x, u)$ is the characteristic function of U. When $x \in \mathbf{R}^{n-1}$ is fixed, this is the characteristic function of a "slice" $t_x(U)$, which is an open subset of $\mathbf{R}$. Applying the case $n = 1$, we find that

$$\int_{\mathbf{R}} f(u + t_1, x) d\lambda_1(u) = \int_{\mathbf{R}} f(u, x) d\lambda_1(u),$$

and therefore

$$\mu(U) = \int_{\mathbf{R}^n} f(u, x) d\lambda_1 \otimes d\lambda_{n-1} = \lambda_n(U),$$

as claimed.

(2) For the converse (which is the main point of this theorem), we first denote $Q = [0,1]^n$ (the compact unit cube in $\mathbf{R}^n$), and we write $c = \mu(Q) < +\infty$; we want to show that $\mu = c\lambda_n$. Let then

$$\mathcal{K} = \{E \in \mathcal{B} \mid \mu(E) = c\lambda_n(E)\},$$

be the collection of sets where μ and $c\lambda$ coincide; this is not empty since it contains Q. By Proposition 5.3.4, it is enough to prove that $\mathcal{K}$ contains all open sets in $\mathbf{R}^n$. Before going to the proof of this, we notice that $\mathcal{K}$ is obviously stable under disjoint countable unions.

The main idea is that $\mathcal{K}$ has also the following additional "divisibility" property: if a set $E \in \mathcal{K}$ can be expressed as a finite disjoint union of translates of a *fixed* set $F \subset \mathbf{R}^n$ (measurable of course), then this set must also be in $\mathcal{K}$. Indeed, the expression

$$E = \bigcup_{1 \leqslant i \leqslant n} (t_i + F)$$

(for some $n \geqslant 1$) implies

$$nc\lambda_n(F) = c\lambda_n(E) = \mu(E) = \sum \mu(t_i + F) = n\mu(F)$$

and therefore $\mu(F) = c\lambda_n(F)$.

We want to apply this, e.g., to claim that $R = [0, 1/2]^n \in \mathcal{K}$ because Q is a union of 2^n translates of R; however, we must first deal with the technical issue of showing that the boundary parts (which prevent us from writing a really disjoint union) have measure zero.

But more generally, if $H \subset \mathbf{R}^n$ is an affine hyperplane parallel to one of the coordinate axes, i.e., a set of the type

$$H = \{x \in \mathbf{R}^n \mid x_i = a\}$$

(for some fixed $a \in \mathbf{R}$ and $i \leqslant n$), we have $\mu(H) = 0$. Indeed, by invariance under translation we can assume $a = 0$, and then we can write

$$H = \bigcup_{m \in \mathbf{Z}^n} (m + Q')$$

where

$$Q' = \{x \in [0, 1]^n \mid x_i = 0\}.$$

But since we have a disjoint union

$$\bigcup_{j \geqslant 1} (e_i j^{-1} + Q') \subset Q,$$

the invariance under translation gives

$$\sum_{i \geqslant 1} \mu(Q') \leqslant c < +\infty,$$

which is only possible if $\mu(Q') = 0$. Here e_i denotes the i-th canonical basis vector. It follows therefore that $\mu(H) = 0$ as claimed. In particular, we see that $[0, 1[^n$ is in $\mathcal{K}$.

Now, applying the previous idea to cubes of the type

$$K_{x,k} = \prod_{i=1}^{n} \Big[x_i, x_i + \frac{1}{2^k}\Big]$$

where $k \geqslant 0$, $x \in \mathbf{R}^n$, we see that all of these lie in $\mathcal{K}$: indeed – up to various sets which are subsets of finitely many hyperplanes – a disjoint union of translates of $K_{x,k}$ is a translate of

$$Q'' = \Big[0, \frac{1}{2^k}\Big]^n,$$

while $[0, 1]^n \in \mathcal{K}$ is itself the disjoint union of 2^k translates of Q''.

Now, any open set U can be written as a countable union of closed cubes of the type above (this is the same proof as the one used in Lemma 5.6.2), and hence we the result that $\mathcal{K}$ contains all open sets, as desired. $\square$

We now use this property to derive an easy proof of the change of variable. The reader is encouraged to read to proof of the change of variable formula first when $n = 1$; in that case, certain obvious simplifications will arise, and the argument will probably be more transparent. In particular, the next lemma, which is the case of a linear substitution, it not necessary for $n = 1$.

LEMMA 5.6.3 (Linear change of variable). *Let $T \in GL(\mathbf{R}^n)$ be an invertible linear map. We then have*

$$T_*(\lambda_n) = |\det(T)|^{-1}\lambda_n.$$

PROOF. Since the right-hand side is a (positive) multiple of the Lebesgue measure, the previous theorem shows that it is natural to try to prove that the measure on the left-hand side is invariant under translation (which will imply the result, up to the identification of the proportionality constant).

But the invariance of $\mu = T_*(\lambda_n)$ is easy. First, it is certainly a Borel measure finite on compact sets (since T is a homeomorphism, again). Now, for any $t \in \mathbf{R}$, let τ_t denote the corresponding translation map

$$\tau_t(x) = x + t.$$

We then have

$$\tau_{t,*}(\mu) = (\tau_t \circ T)_*(\lambda_n),$$

(since composition and direct images work well together), and then we note that, by linearity, we have

$$(\tau_t \circ T)(x) = T(x) + t = T(x + T^{-1}(t)) \qquad \text{for all } x \in \mathbf{R}^n,$$

i.e., we have $\tau_t \circ T = T \circ \tau_{T^{-1}(t)}$. But since λ_n is invariant under *all* translations, it follows that

$$\tau_{t,*}(\mu) = (\tau_t \circ T)_*(\lambda_n) = T_*(\tau_{T^{-1}(t),*}(\lambda_n)) = T_*(\lambda_n) = \mu.$$

By Theorem 5.6.1, (2), we deduce that there exists a constant $c \geqslant 0$ such that

$$T_*(\lambda_n) = c\lambda_n,$$

and in fact that

$$c = T_*(\lambda_n)(Q) = \lambda_n(T^{-1}(Q)) = |\det(T)|^{-1}$$

where $Q = [0,1]^n$, where the last step results from the geometric interpretation of the determinant (Remark 4.4.7). $\square$

We will now proceed to the proof of the general change of variable formula. Let U and $V \subset \mathbf{R}^n$ be open sets, and let

$$\varphi : U \to V$$

be a C^1-diffeomorphism. In order to prove Theorem 4.4.6, we use the interpretation in terms of image measures: we must show that

$$\varphi_*(d\lambda(x)) = |J_{\varphi^{-1}}(y)|d\lambda(y), \tag{5.13}$$

where both sides are Borel measures on the set V.

In order to simplify the notation, we write $J(y) = |J_{\varphi^{-1}}(y)|$, and we denote

$$\mu_1 = \varphi_*(d\lambda(x)), \qquad \mu_2 = J(y)d\lambda(y).$$

We must show that these measures are identical. First of all, they are certainly finite on compact sets (in the case of μ_1, this is because, φ being a homeomorphism, the inverse image of a compact set is compact; in the case of μ_2, this is simply because the function J is continuous, hence bounded on compact sets). Thus each is a Radon measure, and it is sufficient to prove that they coincide on open sets (Proposition 5.3.4). However, the argument will be simplified by first observing that it is enough to show that $\mu_1 \leqslant \mu_2$. (Compare with the proof of (5.4), which was Step 6 in our proof of the Riesz Representation Theorem).

Lemma 5.6.4 (Comparison of measures). *With notation as above, we have*

$$\mu_1(E) \leqslant \mu_2(E)$$

for any Borel set $E \subset \mathbf{R}^n$.

Proof. Since μ_1 and μ_2 are Radon measures (because $\mathbf{R}^n$ is σ-compact), we have

$$\mu_1(E) = \inf\{\mu_1(U) \,:\, U \supset E \text{ open }\},$$
$$\mu_2(E) = \inf\{\mu_2(U) \,:\, U \supset E \text{ open }\},$$

and therefore this will follow for all E if we can show that the inequality $\mu_1(W) \leqslant \mu_2(W)$ is valid when $W \subset V$ is open. In turn, if W is open, we can write it as a countable union of cubes

$$Q_j = [a_{j,1}, b_{j,1}] \times \cdots \times [a_{j,n}, b_{j,n}], \qquad j \geqslant 1$$

where the intersections of two of these cubes are either empty or contained in affine hyperplanes parallel to some coordinate axes. Thus if $\mu_1(Q_j) \leqslant \mu_2(Q_j)$ for all j, we get

$$\mu_1(W) \leqslant \sum_j \mu_1(Q_j) \leqslant \sum_j \mu_2(Q_j) = \mu_2(W),$$

where it is important to note that *in the last equality* we use the fact that if $Z \subset \mathbf{R}^n$ has Lebesgue-measure zero, it also satisfies $\mu_2(Z) = 0$, by definition of measures of the type $f d\mu$. *This property is not obvious at all concerning* μ_1, and hence we have to be quite careful at this point.

Now we proceed to show that $\mu_1(Q) \leqslant \mu_2(Q)$ when Q is any cube. We bootstrap this from a weaker, but simpler inequality: we claim that, for any invertible linear map $T \in GL(n, \mathbf{R})$, we have

$$\mu_1(Q) \leqslant |\det(T)|^{-1} M^n \lambda(Q), \tag{5.14}$$

where

$$M = \sup\{\|T \circ D_y \varphi^{-1}\| \mid y \in Q\}.$$

For this, we first deal with the case $T = 1$; then, according to the mean-valued theorem in multi-variable calculus, we know that

$$\varphi^{-1}(Q) = \{\varphi^{-1}(x) \mid x \in Q\}$$

is contained in another cube which has diameter at most M times the diameter of Q (the diameter of a cube is defined as $\max_i |b_i - a_i|$ if the sides are $[a_i, b_i]$). In that case, the inequality is clear from the formula for the measure of a cube.

Now, given any $T \in GL(n, \mathbf{R})$, we will reduce to this simple case by replacing φ with $\psi = \varphi \circ T^{-1}$ and applying Lemma 5.6.3: we have

$$\begin{aligned}\mu_1(Q) = \lambda(\varphi^{-1}(Q)) &= \lambda(T^{-1}(T \circ \varphi^{-1}(Q)) \\ &= T_*((\varphi \circ T^{-1})^{-1}(Q)) \\ &= |\det(T)|^{-1}\lambda(\psi^{-1}(Q)) \leqslant |\det(T)|^{-1}M^n\lambda(Q),\end{aligned}$$

since (in the computation of M for ψ) we have $D_y(\psi^{-1})) = T \circ D_y(\varphi^{-1})$.

This being done, we finally proceed to the main idea: the inequality (5.14) is not sufficient because M can be too large; but since φ^{-1} is C^1, we can decompose Q into smaller cubes where the continuous function $y \mapsto D_y(\varphi^{-1})$ is *almost* constant; on each of these, applying the inequality is very close to what we want.

Thus let $\varepsilon > 0$ be any real number; using the uniform continuity of $y \mapsto D_y(\varphi^{-1})$ on Q, we can find finitely many closed cubes $Q_1, \ldots, Q_m$, with disjoint interior, and points $y_j \in Q_j$, such that, first of all

$$J(y_j) = \min_{y \in Q_j} |J(y)|,$$

and also, for $T_j = (D_{y_j}\varphi^{-1})^{-1}$, we have

$$M_j = \sup_{y \in Q_j}\{\|T_j \circ D_y\varphi^{-1}\|\} \leqslant 1 + \varepsilon.$$

Then, using monotony and additivity of measures (and, again, being careful that μ_2 vanishes for Lebesgue-negligible sets, but not necessarily μ_1), we apply (5.14) on each Q_j separely, and obtain

$$\mu_1(Q) \leqslant \sum_{j=1}^{m} \mu_1(Q_j) \leqslant \sum_{i=1}^{m} (1+\varepsilon)^n |\det(T_j)|^{-1}\lambda(Q_j).$$

Now we have

$$|\det(T_j)|^{-1} = J(y_j)$$

by definition, and we can express

$$\sum_{i=1}^{m} J(y_j)\lambda(Q_j) = \int_Q s(y)d\lambda_n(y)$$

where the function s is a step function taking value $J(y_j)$ on Q_j. By construction of the y_j, we have $s \leqslant J$, and therefore we derive

$$\mu_1(Q) \leqslant (1+\varepsilon)^n \int_Q J(y)d\lambda_n(y) = (1+\varepsilon)^n\mu_2(Q).$$

Now letting $\varepsilon \to 0$, we obtain the inequality $\mu_1(Q) \leqslant \mu_2(Q)$ that we were aiming at. □

Finally, to conclude the proof of (5.13), we first easily obtain from the lemma the inequality

$$\int_U f(\varphi(x))d\lambda_n(x) \leqslant \int_V f(y)J(y)d\lambda_n(y),$$

valid for any measurable $f : V \to [0, +\infty]$. We apply this to φ^{-1} now: for any $g : U \to [0, +\infty]$, the computation of the revelant Jacobian using the chain rule leads to

$$\int_V g(\varphi^{-1}(y))d\lambda_n(y) \leqslant \int_U g(x)J(\varphi(x))^{-1}d\lambda_n(x),$$

and if we select g so that

$$g(\varphi^{-1}(y)) = f(y)J(y),$$

for a given $f : V \to [0, +\infty]$, i.e.,

$$g(x) = f(\varphi(x))J(\varphi(x)),$$

we obtain the converse inequality

$$\int_U f(\varphi(x))d\lambda_n(x) \geqslant \int_V f(y)J(y)d\lambda_n(y).$$

5.7. Probabilistic applications of Riesz's Theorem

In Section 3.4, we stated the Central Limit Theorem of probability (Theorem 3.4.9). The statement is easier to understand in terms of the important notion of "convergence in law" and its characterization in terms of distribution functions. We now define these terms...

Definition 5.7.1 (Convergence in law and distribution function). Let (Ω, Σ, P) be a probability space.

(1) Let (X_n) be a sequence of random variables, with laws μ_n, and let X (resp. μ) be a random variable (resp. a Borel probability measure on $\mathbf{C}$). Then (X_n) *converges in law* to X (resp. to μ) if and only if

$$\lim_{n\to+\infty} E(f(X_n)) = \lim_{n\to+\infty} \int_{\mathbf{C}} f(x)d\mu_n(x) = \int_X f(x)d\mu(x) = E(f(X))$$

for any continuous compactly-supported function $f \in C_c(\mathbf{C})$. We denote this

$$X_n \overset{\text{law}}{\Longrightarrow} X, \qquad \text{or} \qquad X_n \overset{\text{law}}{\Longrightarrow} \mu.$$

(2) Let X be a real-valued random variable (resp. μ a Borel probability measure on $\mathbf{R}$). The *distribution function* of X (resp. of μ) is the function, denoted F_X or F_μ, defined on $\mathbf{R}$ by

$$F(x) = P(X \leqslant x) = \mu(]-\infty, x]).$$

Remark 5.7.2. (1) Note that, in defining convergence in law, nothing requires in fact that the random variables be defined on the same probability space, since the condition to check only depend on the laws of the (X_n), which are measures on $\mathbf{C}$. This means that this notion of convergence is quite different than what is usually encountered; in particular, it does not make sense to ask if convergence in law has any relation with almost sure convergence.

(2) It is often useful to visualize a probability measure on $\mathbf{R}$ by looking at the graph of its distribution function. This gives often very useful intuitive understanding of the behavior of random variables which have this measure as probability law.

(3) From Corollary 5.3.3, we see – since $\mathbf{C}$ is σ-compact – that the measure μ such that $X_n \overset{\text{law}}{\Longrightarrow} \mu$, if it exists (of course, it may not...), is *unique* (as a Borel probability measure on $\mathbf{C}$).

In the next proposition, which seems merely technical but turns out to be rather crucial, the space of compactly supported continuous functions, which is used to "test" for convergence in law, is replaced by the larger one of bounded continuous functions.

Proposition 5.7.3. *Let (X_n) be a sequence of random variables on a fixed probability space, with laws μ_n, and let μ be a Borel probability measure on $\mathbf{C}$. Then we have $X_n \overset{law}{\Longrightarrow} \mu$ if and only if, for any* bounded *continuous function $f \in C(\mathbf{C}) \cap L^\infty(\mathbf{C})$, we have*

$$\lim_{n\to+\infty} E(X_n) = \int_{\mathbf{C}} f(x) d\mu(x). \tag{5.15}$$

Proof. The condition indicated is stronger than the definition, since compactly-supported functions are bounded, and hence we must only show that convergence in law implies (5.15). Considering separately the real and imaginary parts of a complex-valued bounded function f, we see that we may assume that f is real-valued, and after writing $f = f^+ - f^-$, we may also assume that $f \geqslant 0$ (this is all because the condition to be checked is linear).

The main point of the proof is the following deceptively simple point: in addition to the supply of compactly-supported test functions that we have by definition, our assumption that each μ_n and μ are all probability measures implies that (5.15) holds for the constant function $f = 1$, which is bounded and continuous, *but not compactly supported*: indeed, we have

$$\mu_n(\mathbf{R}) = \int_{\mathbf{R}} d\mu_n(x) \to 1 = \mu(\mathbf{R}).$$

Now fix $f \geqslant 0$ continuous and bounded. We need to truncate it appropriately to enter the realm of compactly-supported functions. For instance, for any integer $N > 0$, let h_N denote the compactly supported continuous function such that

$$h_N(x) = \begin{cases} 1 & \text{if } |x| \leqslant N, \\ 0 & \text{if } |x| > 2N, \end{cases}$$

and h_N is affine-linear on $[-2N, -N]$ and $[N, 2N]$. Thus we have in particular $0 \leqslant h_N \leqslant 1$.

We now write the inequalities

$$\int_{\mathbf{R}} f h_N d\mu_n \leqslant \int_{\mathbf{R}} f d\mu_n = \int_{\mathbf{R}} f h_N d\mu_n + \int_{\mathbf{R}} f(1-h_N) d\mu_n, \tag{5.16}$$

valid for all $n \geqslant 1$. Since $f h_N \in C_c(\mathbf{R})$, we know that

$$\lim_{n\to+\infty} \int_{\mathbf{R}} f h_N d\mu_n = \int_{\mathbf{R}} f h_N d\mu$$

because $X_n \overset{law}{\Longrightarrow} \mu$. Moreover, since $1 - h_N \geqslant 0$, we get

$$\int_{\mathbf{R}} f(1-h_N) d\mu_n \leqslant \|f\|_\infty \int_{\mathbf{R}} (1-h_N) d\mu_n$$

and

$$\int_{\mathbf{R}} (1-h_N) d\mu_n = 1 - \int_{\mathbf{R}} h_N d\mu_n \to 1 - \int_{\mathbf{R}} h_N d\mu = \int_{\mathbf{R}} (1-h_N) d\mu$$

where we have used the crucial observation above and $h_N \in C_c(\mathbf{R})$.

Now let $n \to +\infty$ in (5.16) for fixed N; we obtain

$$\int_{\mathbf{R}} f h_N d\mu \leqslant \liminf_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \limsup_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \int_{\mathbf{R}} f h_N d\mu + \|f\|_\infty \int_{\mathbf{R}} (1-h_N) d\mu.$$

We now finally let $N \to +\infty$ to make the truncation "vanish away": since $h_N(x) \to 1$ for all $x \in \mathbf{R}$, we have

$$f(x)h_N(x) \to f(x) \text{ and } 0 \leqslant f(x)h_N(x) \leqslant f(x),$$
$$1 - h_N(x) \to 0 \text{ and } 0 \leqslant 1 - h_N(x) \leqslant 1,$$

so that we may apply twice the dominated convergence theorem and get

$$\int_{\mathbf{R}} f h_N d\mu \to \int_{\mathbf{R}} f d\mu \text{ and } \int_{\mathbf{R}} (1 - h_N) d\mu \to 0,$$

which then give

$$\int_{\mathbf{R}} f d\mu \leqslant \liminf_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \limsup_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \int_{\mathbf{R}} f d\mu.$$

This means precisely that we have

$$\lim_{n\to+\infty} \int_{\mathbf{R}} f(x) d\mu_n(x) = \int_{\mathbf{R}} f(x) d\mu(x),$$

as claimed. □

Here are the basic properties of the distribution function, which explain how to use it to visualize a probability measure.

Proposition 5.7.4. (1) *Let μ be a Borel probability measure on $\mathbf{R}$. The distribution function F of μ is non-negative, non-decreasing, and satisfies*

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1, \qquad \textit{and} \qquad \mu(]a, b]) = F(b) - F(a)$$

for any real numbers $a < b$.

(2) *In addition, for any $x \in \mathbf{R}$, the limits of F at x exist on the left and on the right, and F is continuous at all $x \in \mathbf{R}$, except possibly on a set D, at most countable, characterized by*

$$D = \{x \in \mathbf{R} \mid \mu(\{x\}) > 0.$$

(3) *The distribution function characterizes μ uniquely, i.e., if μ and ν are Borel probability measures with $F_\mu = F_\nu$, then we have $\mu = \nu$.*

Proof. (1) is elementary using basic properties of measures (in particular the fact that $\mu(\mathbf{R}) = 1 < +\infty$).

(2) The existence of limits on the left and on the right is a general property of any monotonic function. Then the continuity outside of at most countably many points is a consequence (note that if $F(x^-) < F(x^+)$, there exists a rational $y(x)$ such that $F(x^-) < y(x) < F(x^+)$, and distinct x must least to distinct $y(x)$ by monotony).

Finally, for any $x \in \mathbf{R}$, we find that

$$\begin{aligned}\mu(\{x\}) &= \lim_{n\to+\infty} \mu(]x - n^{-1}, x + n^{-1}]) \\ &= \lim_{n\to+\infty} F(x + n^{-1}) - F(x - n^{-1}) = F(x^+) - F(x^-)\end{aligned}$$

which gives the characterization of the points where F is not continuous.

In a similar way, we see that for any open interval $I =]a, b[$, we have

$$\begin{aligned}\mu(I) &= \lim_{n\to+\infty} \mu(]a, b - n^{-1}]) = \lim_{n\to+\infty} (F(b - n^{-1}) - F(a)) \\ &= F(b^-) - F(a)\end{aligned}$$

if $-\infty < a < b < +\infty$, and we get in the same manner

$$\mu(]-\infty, a[) = F(a^-),$$
$$\mu(]a, +\infty[) = 1 - F(a),$$
$$\mu(\mathbf{R}) = 1,$$

which shows that $\mu(I)$ only depends on F for open intervals. Countable additivity shows that the same is true for any open set (written as usual as the union of its countably many connected components). Now, since a Borel probability measure on $\mathbf{R}$ is a Radon measure, we derive from Proposition 5.3.4, (2) that F determines μ. $\square$

This proposition leads to two natural questions: first, given a function F on $\mathbf{R}$ which satisfies the conditions in (1), does there exist a Borel probability measure μ on $\mathbf{R}$ for which the distribution function is F? And, secondly, given the distribution function F_μ, can we use it directly to express an integral like

$$\int_{\mathbf{R}} f(x) d\mu(x)$$

for suitable functions f?

The next formula gives a partial answer:

PROPOSITION 5.7.5 (Integration by parts using distribution function). *Let μ be a Borel probability measure on $\mathbf{R}$ with distribution function F. For any $f \in C_c^1(\mathbf{R})$, i.e., a C^1 function compactly supported, we have*

$$\int_{\mathbf{R}} f(x) d\mu(x) = -\int_{\mathbf{R}} f'(x) F(x) dx.$$

In particular, if X is a random variable with law μ, we have

$$E(f(X)) = -\int_{\mathbf{R}} f'(x) F(x) dx.$$

PROOF. Formally, this is an "integration by parts" formula, where F plays the role of "indefinite integral" of μ. To give a proof, we start from the right-hand side

$$\int_{\mathbf{R}} f'(x) F(x) dx,$$

which is certainly well-defined since the function $x \mapsto f'(x)F(x)$ is bounded and compactly supported. Now we write F itself as an integral:

$$F(x) = \mu(]-\infty, x]) = \int_{\mathbf{R}} \chi(x, y) d\mu(y),$$

where $\chi : \mathbf{R}^2 \to \mathbf{R}$ is the characteristic function of the closed subset

$$D = \{(x, y) \mid y \leqslant x\} \subset \mathbf{R}^2$$

of the plane. Having expressed the integral as a double integral, namely

$$\int_{\mathbf{R}} f'(x) F(x) dx = \int_{\mathbf{R}} \int_{\mathbf{R}} f'(x) \chi(x, y) d\mu(y) dx,$$

we proceed to apply Fubini's theorem, which is certainly permitted here since

$$|f'(x)\chi(x, y)| \leqslant \|f'\|_\infty \chi_K(x)$$

for all x and y, where $K = \mathrm{supp}(f)$, and we have

$$\int_{\mathbf{R}} \int_{\mathbf{R}} \chi_K(x) d\mu(y) dx \leqslant \lambda(K) < +\infty.$$

Fubini's Theorem then leads to

$$\begin{aligned}\int_{\mathbf{R}} f'(x)F(x)dx &= \int_{\mathbf{R}} \int_{\mathbf{R}} f'(x)\chi(x,y)dxd\mu(y)\\ &= \int_{\mathbf{R}} \int_{[y,+\infty]} f'(x)dxd\mu(y)\\ &= -\int_{\mathbf{R}} f(y)d\mu(y).\end{aligned}$$

□

Because of this expression, we may expect that one can give a definition of convergence in law which uses only the distribution function. This is indeed possible, and indeed the resulting equivalent definition is often taken as starting point for the theory.

PROPOSITION 5.7.6. *Let (X_n), $n \geqslant 1$, be a sequence of real-valued random variables, with law μ_n and distribution functions F_n, respectively. Let μ be a Borel probability measure on $\mathbf{R}$ with distribution function F. Then we have $X_n \overset{law}{\Longrightarrow} X$ if and only if*

$$\lim_{n\to+\infty} F_n(x_0) = F(x_0)$$

for all x_0 such that F is continuous at x_0.

PROOF. First, assume that $F_n(x) \to F(x)$ at all x where F is continuous. If we start with $f \in C_c^1(\mathbf{R})$, then Proposition 5.7.5, gives

$$E(f(X_n)) = \int_{\mathbf{R}} f(x)d\mu_n(x) = -\int_{\mathbf{R}} f'(x)F_n(x)d\lambda(x),$$

and since the set of discontinuities of F is at most countable, hence has measure zero, we have

$$f'(x)F_n(x) \to f'(x)F(x)$$

for almost all $x \in \mathbf{R}$. Moreover, we have

$$|f'(x)F_n(x)| \leqslant \|f'\|_\infty \chi_K(x) \in L^1(\mathbf{R}),$$

(where $K = \mathrm{supp}(f')$ which is compact) and the Dominated Convergence Theorem is applicable and leads to

$$-\int_{\mathbf{R}} f'(x)F_n(x)d\lambda(x) \to -\int_{\mathbf{R}} f'(x)F(x)d\lambda(x) = \int_{\mathbf{R}} f(x)d\mu(x),$$

by Propostion 5.7.5 again.

We now extend this to $f \in C_c(\mathbf{R})$, not necessarily differentiable. For any $\varepsilon > 0$, we know that there exists $g \in C_c^1(\mathbf{R})$ such that

$$\|f - g\|_\infty < \varepsilon$$

(see the proof of Proposition 5.4.5), and we then write

$$\begin{aligned}\Big|\int_{\mathbf{R}} fd\mu_n - \int_{\mathbf{R}} fd\mu\Big| &\leqslant \Big|\int_{\mathbf{R}} (f-g)d\mu_n\Big| + \Big|\int_{\mathbf{R}} gd\mu_n - \int_{\mathbf{R}} gd\mu\Big| + \Big|\int_{\mathbf{R}} (g-f)d\mu\Big|\\ &\leqslant 2\varepsilon + \Big|\int_{\mathbf{R}} gd\mu_n - \int_{\mathbf{R}} gd\mu\Big|.\end{aligned}$$

Letting $n \to +\infty$, we find first that

$$\limsup_{n\to+\infty} \Big(\int_{\mathbf{R}} f d\mu_n - \int_{\mathbf{R}} f d\mu\Big) \leqslant 2\varepsilon,$$

and then that the limit exists and is zero since ε was arbitrarily small. This proves the first half of the result.

Now for the converse; since, formally, we must prove (5.15) with

$$f(x) = \chi_{]-\infty,x_0]}(x),$$

where x_0 is arbitrary, we will use Proposition 5.7.3 (note that f is not compactly supported; it is also not continuous, of course, and this explains the coming restriction on x_0). Now, for any $\varepsilon > 0$, we consider the function f given by

$$f(x) = \begin{cases} 1 & \text{if } x \leqslant x_0 \\ 0 & \text{if } x > x_0 + \varepsilon \end{cases}$$

and extended by linearity on $[x_0, x_0 + \varepsilon]$. We therefore have

$$\chi_{]-\infty,x_0]} \leqslant f \leqslant \chi_{]-\infty,x_0+\varepsilon]},$$

and by integration we obtain

$$F_n(x_0) = \mu_n(]-\infty, x_0]) \leqslant \int_{\mathbf{R}} f d\mu_n \to \int_{\mathbf{R}} f d\mu$$

as $n \to +\infty$, by Proposition 5.7.3, as well as

$$\int_{\mathbf{R}} f d\mu = \lim_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \mu(]-\infty, x_0 + \varepsilon]) = F(x_0 + \varepsilon).$$

This, and an analogue argument with $x_0+\varepsilon$ replaced by $x_0-\varepsilon$, leads to the inequalities

$$F(x_0 - \varepsilon) \leqslant \liminf_{n\to+\infty} F_n(x_0) \leqslant \limsup_{n\to+\infty} F_n(x_0) \leqslant F(x_0 + \varepsilon).$$

Now $\varepsilon > 0$ was arbitarily small; if we let $\varepsilon \to 0$, we see that we can conclude *provided* F is continuous at x_0, so that the extreme lower and upper bounds both converge to $F(x_0)$, which proves that for any such x_0, we have

$$F_n(x_0) \to F(x_0)$$

as $n \to +\infty$. ☐

Remark 5.7.7. The restriction on the set of x which is allowed is necessary to obtain a "good" notion. For instance, consider a sequence (x_n) of positive real numbers converging to (say) 0 (e.g., $x_n = 1/n$), and let μ_n be the Dirac measure at x_n, μ the Dirac measure at 0. For any continuous (compactly-supported) function f on $\mathbf{R}$, we have

$$\int_{\mathbf{R}} f(x) d\mu_n(x) = f(x_n),$$

and continuity implies that this converges to

$$f(0) = \int_{\mathbf{R}} f(x) d\mu(x).$$

Thus, μ_n converges in law to μ, and this certainly seems very reasonable. However, the distribution functions are given by

$$F_n(x) = \begin{cases} 0 & \text{if } x < x_n, \\ 1 & \text{if } x \geqslant x_n, \end{cases}$$

and

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geqslant 0, \end{cases}$$

Now for $x = 0$, we have $F_n(0) = 0$ for all n, since $x_n > 0$ by assumption, and this of course does not converge to $F(0) = 1$.

We see now that we may formulate the Central Limite Theorem as follows:

THEOREM 5.7.8 (Central Limit Theorem). *Let (X_n) be a sequence of real-valued random variables on some probability space (Ω, Σ, P). If the (X_n) are independent and identically distributed, and if $X_n \in L^2(\Omega)$, and $E(X_n) = 0$, then*

$$\frac{S_n}{\sqrt{n}} \stackrel{law}{\Longrightarrow} \mu_{0,\sigma^2},$$

where $\sigma^2 = E(X_n^2)$ is the common variance of the X_n's and μ_{0,σ^2} is the Gaussian measure with expectation 0 and variance σ^2.

CHAPTER 6

The convolution product

6.1. Definition

The goal of this chapter is to study a very important construction, called the *convolution product* (or simply convolution) of two functions on Euclidean space $\mathbf{R}^d$. Formally, this operation is defined as follows: given complex-valued functions f and g definde on $\mathbf{R}^d$, we define (or wish to define) the convolution $f \star g$ as a function on $\mathbf{R}^d$ given by

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)dt \tag{6.1}$$

for $x \in \mathbf{R}^d$, where dt denotes integration with respect to the Lebesgue measure on $\mathbf{R}^d$. Obviously, some conditions will be required for this integral to make sense, and there are various conditions that ensure that it is well defined, as we will see.

To present things rigorously and conveniently, we make the following (temporary, and not standard) definition:

Definition 6.1.1 (Convolution domain). Let f and g be measurable complex-valued functions on $\mathbf{R}^d$. The *convolution domain* of f and g, denoted $\mathcal{C}(f,g)$, is the set of all $x \in \mathbf{R}^d$ such that the integral above makes sense, i.e., such that

$$t \mapsto f(x-t)g(t)$$

is in $L^1(\mathbf{R}^d)$. For all $x \in \mathcal{C}(f,g)$, we let

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)dt.$$

Lemma 6.1.2. (1) *Let $x \in \mathcal{C}(f,g)$. Then $x \in \mathcal{C}(g,f)$ and we have $f \star g(x) = g \star f(x)$.*

(2) *Let $x \in \mathcal{C}(f,g) \cap \mathcal{C}(f,h)$, and let α, $\beta \in \mathbf{C}$ be constants. Then $x \in \mathcal{C}(f, \alpha g + \beta h)$ and*

$$f \star (\alpha g + \beta h)(x) = \alpha f \star g(x) + \beta f \star h(x).$$

(3) *If $x \notin \mathrm{supp}(f) + \mathrm{supp}(g)$, then we have $x \in \mathcal{C}(f,g)$ and $f \star g(x) = 0$, where the "sumset" is defined by*

$$A + B = \{z \in \mathbf{R}^d \mid z = x + y \text{ for some } x \in A \text{ and } y \in B\}$$

for any subsets A, $B \subset \mathbf{R}^d$. In other words, we have

$$\mathrm{supp}(f \star g) \subset \mathrm{supp}(f) + \mathrm{supp}(g) \tag{6.2}$$

if $\mathcal{C}(f,g) = \mathbf{R}^d$.

(4) *We have*

$$\mathcal{C}(f,g) = \mathcal{C}(|f|,|g|),$$

and moreover

$$|f \star g(x)| \leqslant |f| \star |g|(x) \tag{6.3}$$

for $x \in \mathcal{C}(f,g)$.

Proof. All these properties are very elementary. First, (1) is a direct consequence of the invariance of Lebesgue measure under translations, since

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)dt = \int_{\mathbf{R}^d} f(y)g(x-y)dy = g \star f(x).$$

Then (2) is quite obvious by linearity, and for (3), it is enough to notice that if $x \notin \mathrm{supp}(f) + \mathrm{supp}(g)$, it must be the case that for any $t \in\in \mathrm{supp}(g)$, the element $x-t$ is *not* in $\mathrm{supp}(f)$; this implies that $f(x-t)g(t) = 0$, and since this holds for all t, we have trivially $f \star g = 0$. Finally, (4) is also immediate since

$$|f(x-t)g(t)| = |f(x-t)||g(t)|$$

and

$$\Big|\int_{\mathbf{R}^d} f(x-t)g(t)dt\Big| \leqslant \int_{\mathbf{R}^d} |f(x-t)||g(t)|dt.$$

□

6.2. Existence of the convolution product

We start by considering the properties of the convolution for non-negative functions, in which case it is of course defined everywhere (but may take the value $+\infty$).

Proposition 6.2.1. *Let f and g be non-negative measurable functions on $\mathbf{R}^d$, and let $f \star g$ denote the convolution of f and g, taking values in $[0, +\infty]$. We then have*

$$\int_{\mathbf{R}^d} f \star g(x)dx = \Big(\int_{\mathbf{R}^d} f(x)dx\Big)\Big(\int_{\mathbf{R}^d} g(x)dx\Big). \tag{6.4}$$

Proof. This is a simple application of Tonelli's theorem and of the invariance of the Lebesgue measure under translations; indeed, these results give

$$\begin{aligned}\int_{\mathbf{R}^d} f \star g(x)dx &= \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} f(x-t)g(t)dtdx \\ &= \int_{\mathbf{R}^d} g(t)\Big(\int_{\mathbf{R}^d} f(x-t)dx\Big)dt \\ &= \Big(\int_{\mathbf{R}^d} f(x)dx\Big)\Big(\int_{\mathbf{R}^d} g(t)dt\Big).\end{aligned}$$

□

This result quickly leads to the first important case of existence of the convolution product, namely when one function is integrable, and the other in some L^p space.

Theorem 6.2.2 (Convolution $L^1 \star L^p$). *Let $f \in L^p(\mathbf{R}^d)$ and $g \in L^1(\mathbf{R}^d)$ where $1 \leqslant p \leqslant +\infty$. Then $\mathcal{C}(f,g)$ contains almost all $x \in \mathbf{R}^d$, and the function $f \star g$ which is thus defined almost everywhere is in $L^p(\mathbf{R}^d)$ and satisfies*

$$\|f \star g\|_p \leqslant \|f\|_p\|g\|_1, \tag{6.5}$$

Moreover, if we also have $p = 1$, then

$$\int_{\mathbf{R}^d} f \star g(x)dx = \Big(\int_{\mathbf{R}^d} f(x)dx\Big)\Big(\int_{\mathbf{R}^d} g(x)dx\Big), \tag{6.6}$$

and if we have three functions f, g, $h \in L^1(\mathbf{R}^d)$, we have

$$(f \star g) \star h = f \star (g \star h). \tag{6.7}$$

Proof. Assume first that $p = 1$. By applying the previous proposition to the non-negative measurable functions $|f|$ and $|g|$, we see that

$$\int_{\mathbf{R}^d} |f| \star |g| d\lambda = \|f\|_1 \|g\|_1 < +\infty.$$

By Proposition 2.2.2, (1), this implies that $|f| \star |g|$ is finite almost everywhere, and therefore, according to Lemma 6.1.2, (4), the set $\mathcal{C}(f, g) = \mathcal{C}(|f|, |g|)$ does contain almost all $x \in \mathbf{R}^d$.

This being established, we integrate (6.3) over $x \in \mathbf{R}^d$, and obtain

$$\|f \star g\|_1 \leqslant \int_{\mathbf{R}^d} |f| \star |g| d\lambda = \|f\|_1 \|g\|_1 < +\infty$$

by (6.4).

From this, we see that $f \star g \in L^1(\mathbf{R}^d)$, and now we can repeat the computation leading to (6.4) and see that it is now possible to apply Fubini's Theorem instead of Tonelli's Theorem, and obtain from this the identity (6.6).

Similarly, another easy application of Fubini's Theorem gives the associativity of the convolution product for three functions in L^1: we have

$$\begin{aligned}
(f \star g) \star h(x) &= (g \star f) \star h(x) \\
&= \int_{\mathbf{R}^d} (g \star f)(x - t) h(t) dt \\
&= \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} g(x - t - v) f(v) dv h(t) dt \\
&= \int_{\mathbf{R}^d} f(v)(g \star h)(x - v) dv \\
&= f \star (g \star h)(x).
\end{aligned}$$

Now we must still prove the existence of $f \star g$ when $f \in L^p(\mathbf{R}^d)$ and $g \in L^1(\mathbf{R}^d)$. First of all, the case $p = \infty$ is immediate. Next, if $p < +\infty$, we may assume that $g \neq 0$ (i.e., is not zero almost everywhere). Then we consider the probability measure on $\mathbf{R}^d$ given by

$$\mu = \frac{|g|}{\|g\|_1} d\lambda.$$

Applying Hölder's inequality, we obtain

$$\int_{\mathbf{R}^d} |f(x - t)||g(t)| dt = \|g\|_1 \int_{\mathbf{R}^d} |f(x - t)| d\mu(t) \leqslant \|g\|_1 \Big(\int_{\mathbf{R}^d} |f(x - t)|^p d\mu(t) \Big)^{1/p},$$

and hence

$$\begin{aligned}
\int_{\mathbf{R}^d} \Big(\int_{\mathbf{R}^d} |f(x - t)||g(t)| dt \Big)^p dx &\leqslant \|g\|_1^p \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} |f(x - t)|^p dx d\mu(t) \\
&= \|g\|_1^p \Big(\int_{\mathbf{R}^d} d\mu(t) \Big) \Big(\int_{\mathbf{R}^d} |f(x)|^p dx \Big) = \|f\|_p^p \|g\|_1^p < +\infty,
\end{aligned}$$

using once more the invariance under translation of the Lebesuge measure, and the definition of μ. Hence, as before, we see that $(|f| \star |g|)^p$ is finite almost everywhere, and thus $\mathcal{C}(f, g)$ contains almost all $x \in \mathbf{R}^d$, and finally we also get the inequality

$$\|f \star g\|_p \leqslant \|f\|_p \|g\|_1.$$

$\square$

Remark 6.2.3. (1) From this result, we see that we have defined an operation

$$\begin{cases} L^1(\mathbf{R}^d) \times L^1(\mathbf{R}^d) \to L^1(\mathbf{R}^d) \\ (f, g) \mapsto f \star g \end{cases}$$

which is commutative, associative, and associative with respect to addition of functions. This is why this is called the "convolution product". It should already be mentioned that, in sharp constrast with the "usual" product, the convolution product does not have a *unit*: there is no function $\delta \in L^1(\mathbf{R}^d)$ such that $f \star \delta = f = \delta \star f$ for all f. However, a crucial property of convolution will turn out to be the existence of "approximate units", which have very important applications (see Section 6.4).

(2) The inequality (6.5) is quite important, as it shows that the convolution product, seen as a bilinear map, is continuous on $L^1 \times L^1$. In particular, it follows that whenever we have convergent sequences (f_n) and (g_n) in L^1, with

$$\lim_{n\to+\infty} f_n = f \in L^1(\mathbf{R}^d), \qquad \lim_{n\to+\infty} g_n = g \in L^1(\mathbf{R}^d),$$

the sequence $(f_n \star g_n)$ also converges (in $L^1(\mathbf{R}^d)$) to $f \star g$. We recall the easy argument: we have

$$\begin{aligned} \|f_n \star g_n - f \star g\|_1 &\leqslant \|f_n \star (g_n - g)\|_1 + \|(f_n - f) \star g\|_1 \\ &\leqslant \|f_n\|_1 \|g_n - g\|_p + \|g\|_p \|f_n - f\|_1 \to 0 \end{aligned}$$

as $n \to +\infty$ (since the sequence $(\|f_n\|_1)$ is bounded).

Similarly, if $g \in L^1(\mathbf{R}^d)$, and if $1 \leqslant p \leqslant +\infty$, the operation of convolution with the fixed function g defines a continuous linear map

$$\begin{cases} L^p(\mathbf{R}^d) \to L^p(\mathbf{R}^d) \\ f \mapsto f \star g \end{cases}$$

(the continuity comes from (6.5)).

Before we go on to another example of existence of convolution products, here is a first concrete probabilistic interpretation of this operation. (Another important motivation is given by the link with the Fourier transform, which is formally the identity $\widehat{f \star g} = \hat{f}\hat{g}$).

Proposition 6.2.4 (Density of sums of independent variables). *Let (Ω, Σ, P) be a probability space, and let X and Y be real-valued* independent *random variables. Assume that the laws of X and Y are given by*

$$X(P) = f(x)dx, \ \text{and}\ Y(P) = g(x)dx$$

where f and g are non-negative functions on $\mathbf{R}$. Then the law of $X + Y$ is given by the probability measure

$$(X + Y)(P) = (f \star g)(x)dx.$$

Proof. We must show that, for any Borel subset $B \subset \mathbf{R}$, we have

$$P(X + Y \in B) = \int_B f \star g(x)dx.$$

Now, by definition, we have

$$\begin{aligned} \int_B f \star g(x)dx &= \int_B \int_{\mathbf{R}} f(x - t)g(t)dtdx \\ &= \int_E f(u)g(v)dudv \end{aligned}$$

where

$$E = \{(u,v) \in \mathbf{R}^2 \mid u+v \in B\}.$$

But since the joint law $(X,Y)(P)$ of X and Y is given by

$$(X,Y)(P) = X(P) \otimes Y(P)$$

(because they are independent, see Lemma 4.2.1), we have

$$\int_E d(X,Y)P = P((X,Y) \in E) = P(X+Y \in B),$$

as claimed. □

Remark 6.2.5. More generally, if μ and ν are arbitrary probability measures on $\mathbf{R}$, we can define a convolution measure by

$$\mu \star \nu = s_*(\mu \otimes \nu)$$

where $s : \mathbf{C}^2 \to \mathbf{C}$ is the addition map

$$s : (x,y) \mapsto x+y.$$

(so that $(\mu \star \nu)(B) = (\mu \otimes \nu)(s^{-1}(B))$).

From the above, we see that

$$(f(x)dx) \star (g(x)dx) = (f \star g)(x)dx,$$

for f and g non-negative with integral 1. We also see by the same method that

$$(X+Y)(P) = X(P) \star Y(P)$$

for any pair of independent random variables (X,Y).

It is interesting to see that, in this more general framework, there *is* a "unit" for the convolution of measures: denoting by δ the Diract measure at the point 0 (see Example 1.2.7, (2)), we have

$$\mu \star \delta = \delta \star \mu = \mu$$

for any probability measure μ.

Now we come to the second importance case of existence of convolutions.

Proposition 6.2.6 (Convolution $L^p \star L^q$ for $1/p + 1/q = 1$). *Let $p \in [1,+\infty]$ and let q be the complementary exponent, so that $1/p+1/q = 1$. For any $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$, we have $\mathcal{C}(f,g) = \mathbf{R}^d$ and $f \star g \in L^\infty$. Indeed, we have*

$$\|f \star g\|_\infty \leqslant \|f\|_p \|g\|_q. \tag{6.8}$$

Proof. We apply Hölder's inequality (3.12): if f and g are non-negative first, we have

$$\begin{aligned} f \star g(x) &= \int_{\mathbf{R}^d} f(x-t)g(t)dt \\ &\leqslant \|g\|_q \Big(\int_{\mathbf{R}^d} f(x-t)^p dt\Big)^{1/p} \\ &= \|g\|_q \|f\|_p < +\infty \end{aligned}$$

for all x, having used (once more) the invariance property of the Lebesgue measure. Now come back to any $f \in L^p$ and $g \in L^q$; by Lemma 6.1.2, (4) we see that $\mathcal{C}(f,g) \in L^\infty(\mathbf{R}^d)$ and then that

$$|f \star g(x)| \leqslant \|f\|_p \|g\|_q$$

for all $x \in \mathbf{R}^d$, leading to (6.8). □

Remark 6.2.7. As in the previous case, the inequality (6.8) implies that this product is continuous: if $f_n \to f$ in L^p and $g_n \to g$ in L^q, then $f_n \star g_n \to f \star g$ in L^∞ (which means uniformly over $\mathbf{R}^d$).

Our last case is a bit more technical, but it is also very useful in some situations where the previous ones do not apply.

Definition 6.2.8 (Local L^p spaces, L^p functions with compact support). Let X be any topological space and μ a Borel measure on X, finite on compact sets. For any $1 \leqslant p \leqslant +\infty$, we denote

$$L^p_c(X) = \{f \in L^p(X) \mid \operatorname{supp}(f) = K \cup Z \text{ where } K \text{ is compact and } \mu(Z) = 0\}$$
$$L^p_{loc}(X) = \{f \mid f \in L^p(K) \text{ for any compact subset } K \subset X\}.$$

The vector space L^p_c is the space of *L^p functions with compact support*, and L^p_{loc} is the space of *functions which are locally in L^p*.

Proposition 6.2.9 (Convolution $L^1_{loc} \star L^\infty_c$). *Let $f \in L^1_{loc}(\mathbf{R}^d)$ and $g \in L^\infty_c(\mathbf{R}^d)$. Then $\mathcal{C}(f, g) = \mathbf{R}^d$, hence the convolution $f \star g$ is defined for all $x \in \mathbf{R}^d$.*

Proof. We reduce, as usual, to the case where $f \geqslant 0$ and $g \geqslant 0$. We can write

$$\int_{\mathbf{R}^d} f(x-t)g(t)dt = \int_K f(x-t)g(t)dt$$

where K is a compact subset such that $g(x) = 0$ outside $K \cup Z$, where $\lambda(Z) = 0$. Then we find

$$f \star g(x) \leqslant \|g\|_\infty \int_K f(x-t)dt = \|g\|_\infty \int_{x-K} f(u)du < +\infty$$

since the translated set $x - K = \{x - k \mid k \in K\}$ is still compact and $f \in L^1_{loc}$. $\square$

Example 6.2.10. Any continuous function is automatically in dans $L^1_{loc}(\mathbf{R}^d)$, but not in L^1 in general (for instance $f = 1$). Correspondingly, it is easy to see that $f \star g$ is not bounded in general.

One should note that the three cases of existence of the convolution $f \star g$ that we have seen ($L^1 \star L^p$, $L^p \star L^q$, $L^1_{loc} \star L^\infty_c$) are not exclusive of each other: it may well be the case that two (or three) of them apply for the same pair of functions (f, g).

6.3. Regularization properties of the convolution operation

If we view $f \star g$ as a function defined by an integral depending on a parameter x, the function that we integrate is

$$h(x, t) = f(x-t)g(t),$$

which is extremely simple from the point of view of the results of Section 3.1: for a gixed $t \in \mathbf{R}^d$, it is a "translate" of f, multiplied by a constant $g(t)$. One may expect from this that $f \star g$ will inherit the regularity properties of f. And since $f \star g = g \star f$, it will also get the corresponding properties coming from g.

Here are some examples of this principle. We start by showing that, under suitable conditions, the convolution is differentiable. First of all, we recall a convenient notation for partial derivatives of a function. Given a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, where $\alpha_i \geqslant 0$ are integers, we denote

$$|\alpha| = \alpha_1 + \cdots + \alpha_d \tag{6.9}$$

and

$$\partial_\alpha = \frac{\partial^{|\alpha|}}{\partial x^\alpha} = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}, \tag{6.10}$$

which we view as an operator acting on functions defined on open subsets of $\mathbf{R}^d$ which are of class $C^{|\alpha|}$, so that the partial derivative in question is well-defined.

PROPOSITION 6.3.1 (Derivatives of convolution). *Let $k \geqslant 1$, $f \in L^1(\mathbf{R}^d)$ and let $\varphi \in C^k(\mathbf{R}^d)$ be such that*

$$\partial_\alpha \varphi \in L^\infty$$

for all multi-indices α such that $|\alpha| \leqslant k$, or in other words, such that all partial derivatives of order at most k are bounded. Then we have

$$f \star \varphi \in C^k(\mathbf{R}^d),$$

and, for all $|\alpha| \leqslant k$, the partial derivative corresponding to α is given by

$$\partial_\alpha(f \star \varphi) = f \star (\partial_\alpha \varphi).$$

PROOF. First of all, note that all the convolutions which are written down in this statement are of the type

$$f \star \partial_\alpha \varphi$$

with $|\alpha| \leqslant k$. Hence, they are well-defined because of the assumption on the derivatives of φ (this is the case $L^1 \star L^\infty$ of Proposition 6.2.6), and they are functions in $L^\infty(\mathbf{R}^d)$.

Using an easy induction and the definition of partial derivatives, we can assume that $d = 1$ and $k = 1$ (note a small subtlety: if $f \in L^1(\mathbf{R}^d)$, it is not necessarily the case that all partial functions of the type

$$g(x) = f(x, x_2, \ldots, x_d)$$

are in $L^1(\mathbf{R})$ for fixed x_i, $i \geqslant 2$; however, an easy argument involving Fubini's theorem shows that this is true for almost all x_2, $\ldots$, $x_n d$, which is enough to conclude).

Now, in that special case, we have

$$f \star \varphi(x) = \varphi \star f(x) = \int_{\mathbf{R}} f(t)\varphi(x-t)dt,$$

and we can apply the criterion of differentiability under the integral sign (Proposition 3.1.4) with derivative given for each fixed t by

$$\frac{\partial}{\partial x} f(t)\varphi(x-t) = f(t)\varphi'(x-t).$$

The domination estimate

$$|f(t)\varphi'(x-t)| \leqslant \|\varphi'\|_\infty |f(t)|$$

with $f \in L^1(\mathbf{R})$ shows that Proposition 3.1.4 is indeed applicable, and gives the desired result. □

EXAMPLE 6.3.2. Consider for instance $f \in C_c^k(\mathbf{R})$ and $g \in C_c^m(\mathbf{R})$. Then we may apply the proposition first with f and g, and deduce that $f \star g$ is of class C^m on $\mathbf{R}$. But then we may exchange the two arguments, and conclude that $f \star g \in C_c^{k+m}(\mathbf{R})$ (see (6.2) to check the support condition) with

$$(f \star g)^{(k+m)} = f^{(k)} \star g^{(m)},$$

for instance. For the first derivative, one may write either of the two formulas:

$$(f \star g)' = f' \star g = f \star g'.$$

Here is a further regularity property, which is a bit deeper, since the functions involved in the convolution do not have any obvious regularity by themselves.

Proposition 6.3.3 (Average regularization effect). *Let $1 \leqslant p \leqslant +\infty$ and let q be the complementary exponent of p. For any $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$, the convolution $f \star g \in L^\infty$ is uniformly continuous on $\mathbf{R}^d$. Moreover, if $1 < p, q < +\infty$, we have*

$$\lim_{\|x\| \to +\infty} f \star g(x) = 0.$$

Proof. For the first part, we can assume by symmetry that $p < +\infty$ (exchanging f and g in the case $p = +\infty$). For $x \in \mathbf{R}^d$ and $h \in \mathbf{R}^d$, we obtain the obvious upper bound

$$\begin{aligned} |f \star g(x+h) - f \star g(x)| &\leqslant \int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)||g(t)|dt \\ &\leqslant \|g\|_q \Big(\int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)|^p dt \Big)^{1/p} \end{aligned}$$

by Hölder's inequality. Using the linear change of variable $x - t = u$, the integral with respect to t becomes

$$\int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)|^p dt = \int_{\mathbf{R}^d} |f(u+h) - f(u)|^p du.$$

Now the last integral, say $\varepsilon(h)$, is independent of x, and by Proposition 5.5.1, we have $\varepsilon(h) \to 0$ as $h \to 0$. Thus the upper bound

$$|f \star g(x+h) - f \star g(x)| \leqslant \varepsilon(h)^{1/p} \|g\|_q$$

shows that $f \star g$ is uniformly continuous, as claimed.

Now assume that both p and q are $< \infty$. Then, according to the approximation theorem (Theorem 5.4.1), there are sequences (f_n) and (g_n) of continuous functions with compact support such that $f_n \to f$ in L^p and $g_n \to g$ in L^q. By continuity of the convolution (Remark 6.2.7), we see that

$$f_n \star g_n \to f \star g$$

in L^∞, i.e., uniformly on $\mathbf{R}^d$. Since $f_n \star g_n \in C_c(\mathbf{R}^d)$ by (6.2), we get

$$\lim_{\|x\| \to +\infty} f \star g(x) = 0$$

by Remark 5.4.2. □

Example 6.3.4. Here is a very nice classical application of the first part of this last result. Consider two measurable sets A, $B \subset \mathbf{R}^d$, such that

$$\lambda(A) > 0, \qquad \lambda(B) > 0.$$

Then we claim that the set

$$C = A + B = \{c \in \mathbf{R}^d \mid c = a + b \text{ for some } a \in A,\ b \in B\}$$

has non-empty interior, or in other words, there exists $c = a + b \in C$ and $r > 0$ such that the open ball centered at c with radius r is entirely contained in C. In particular, when $d = 1$, this means that the "sum" of any two sets with positive measure contains a non-empty interval! (Of course, it may well be that neither A nor B has this property, e.g., if $A = B = \mathbf{R} - \mathbf{Q}$ is the set of irrational numbers).

The proof of this striking property is very simple when using the convolution. Indeed, by first replacing A and B (if needed) by $A \cap D_R$ and $B \cap D_R$, where D_R is the ball centered at 0 with radius R large enough, we can clearly assume that

$$0 < \lambda(A),\ \lambda(B) < +\infty.$$

In that case, the functions $f = \chi_A$, $g = \chi_B$ are both in $L^2(\mathbf{R}^d)$ (this is why the truncation was needed). By Proposition 6.3.3, the convolution $h = f \star g \geqslant 0$ is a uniformly continuous function. Moreover, since f and g are non-negative, we have also $h \geqslant 0$ and

$$\int h dx = \int f dx \int g dx = \lambda(A)\lambda(B) > 0,$$

which implies that $h \geqslant 0$ is not identically zero. Consider now any c such that $h(c) > 0$. By continuity of h, there exists an open ball centered at c, with radius $r > 0$, such that $h(x) > 0$ for all $x \in U$. Then, since (as in the proof of Lemma 6.1.2, (3)) we have $h(x) = 0$ when $x \notin A + B$, it must be the case that $U \subset A + B$.

6.4. Approximation by convolution

Part of the significance of the regularity properties of convolutions lies in the fact that it is possible to approach a given function f (in some suitable sense) using convolution products $f \star \varphi$, where φ is an "approximate unit" for convolution. These approximate units can be constructed with almost arbitrary regularity properties, and the approximations $f \star \varphi$ inherit them.

Before starting, we explain quickly why there is no exact unit for the convolution product on $L^1(\mathbf{R})$ (the case of $\mathbf{R}^d$ is of course similar).

Lemma 6.4.1 (Properties of a convolution unit). *Let $\delta \in L^1(\mathbf{R})$ be a unit for convolution, i.e., such that $f \star \delta = f$ for any $f \in L^1(\mathbf{R})$. Then*

(1) *The support of δ is $\{0\}$.*
(2) *We have $\int_{\mathbf{R}} \delta(x)dx = 1$.*

These two properties are of course incompatible for functions. However, if one weakens the first condition by asking that the support of δ be contained in a small interval $[-\varepsilon, \varepsilon]$ with $\varepsilon > 0$, then there is no difficulty in constructing functions in $L^1(\mathbf{R})$ with the corresponding properties (e.g., $1/(2\varepsilon)\chi_{[-\varepsilon,\varepsilon]}$). It is then natural to expect that $f \star \varphi$ will be, if not equal to, at least close to f, because of the continuity properties of convolution.

Remark 6.4.2. (1) The quickest argument to see that δ does not exist uses the Fourier transform which implies by an easy computation that

$$\widehat{f \star \delta} = \hat{f}\hat{\delta}$$

for any $f \in L^1(\mathbf{R})$. For any $t \in \mathbf{R}$, we can find a function f with $\hat{f}(t) > 0$; then since both sides are continuous functions, comparison gives $\hat{\delta}(t) = 1$, or in other words, the Fourier transform of δ needs to be the unit 1 for the usual multiplication of functions. However, from the Riemann-Lebesgue Lemma (Theorem 5.5.3), we see that this is not possible.

(2) In the land of measures, there *is* an object which has the properties of a unit for convolution, namely the Dirac measure at 0, which is supported on $\{0\}$ and for which the function 1 has integral 1. One can indeed construct, in some generality, a theory of convolution of measures, for which this Dirac measure δ is a unit.

PROOF. Property (2) is immediate by taking the integral of $f \star \delta$ for any function f with non-zero integral.

For Property (1), take f to be the characteristic function of any set N not containing 0, and $x = 0$; then we find that

$$0 = \int_N \delta(-t)dt$$

(since this result is only for illustration, we assumed that the point $x = 0$ is in the set where the functions $f \star \delta$ and f coincide). It is easy to see that this implies that δ is zero almost everywhere on any such set, which is the desired conclusion. □

These properties suggest the following definition.

DEFINITION 6.4.3 (Dirac sequence). A *Dirac sequence* in $\mathbf{R}^d$ is a sequence (φ_n) of non-negative measurable functions such that

$$\int_{\mathbf{R}^d} \varphi_n(t)dt = 1 \text{ for all } n \geqslant 1 \tag{6.11}$$

$$\lim_{n \to +\infty} \int_{\|t\| > \eta} \varphi_n(t)dt = 0 \text{ for all } \eta > 0, \tag{6.12}$$

where $\|t\|$ is the euclidean norm (4.24) on $\mathbf{R}^d$.

REMARK 6.4.4. On may replace the euclidean norm by any equivalent one, for instance, by

$$\|t\|_\infty = \max_{1 \leqslant i \leqslant d} |t_i|, \qquad \text{or} \qquad \|t\|_1 = \sum_{1 \leqslant i \leqslant d} |t_i|.$$

We start with some examples of such sequences.

EXAMPLE 6.4.5. (1) Let $\varphi_n = (2n)^d \chi_{[-n^{-1}, n^{-1}]^d}$. This is a Dirac sequence since (6.11) is clear and we have

$$\int_{\|t\| > \eta} \varphi_n(t)dt = 0$$

for all $n > \eta^{-1}$, which obviously gives (6.12).

(2) More generally, let (φ_n) be any sequence of non-negative functions in L^1 for which (6.11) holds, and for which

$$\operatorname{supp}(\varphi_n) \subset \{\|x\| < \varepsilon_n\}$$

for some $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$ as $n \to +\infty$. Then the same argument shows that (φ_n) is a Dirac sequence.

(3) Condition (6.12) means that, for any $\eta > 0$, the sequence (φ_n) restricted to

$$U_\eta = \{t \mid \|t\| > \eta\}$$

tends to zero in $L^1(U_\eta)$. The intuitive meaning is that most of that “mass” of the function (which is always 1) is concentrated very close to 0.

For instance, let $\psi \geqslant 0$ be any function in L^1 such that

$$I = \int_{\mathbf{R}^d} \psi(t)dt > 0.$$

Define then $\varphi = \psi / I$ and

$$\varphi_n(t) = n^d varphi(nt)$$

for $n \geqslant 1$. Then (φ_n) is also a Dirac sequence.

Indeed, the linear substitution $u = nt$ (with Jacobian given by $\det(n\mathrm{Id}) = n^d$) gives first

$$\int_{\mathbf{R}^d} \varphi_n(t)dt = n^d \int_{\mathbf{R}^d} \varphi(nt)dt = \int_{\mathbf{R}^d} \varphi(u)du = 1$$

for all n, and then, for any $\eta > 0$, we get

$$\int_{\|t\|>\eta} \varphi_n(t)dt = \int_{\|t\|>n\eta} \varphi(t)dt,$$

and since φ is integrable and the sets $\{\|t\| > b\eta\}$ are decreasing with empty intersection, we obtain

$$\lim_{n\to+\infty} \int_{\|t\|>n\eta} \varphi(t)dt = 0$$

(Lemma 2.3.11).

An typical example is given by

$$\varphi(t) = e^{-\pi\|t\|^2},$$

since Fubini's Theorem and Proposition 4.4.9 imply that

$$\int_{\mathbf{R}^d} e^{-\pi\|t\|^2} dt = \int_{\mathbf{R}^d} \exp\Bigl(-\pi \sum_{1\leqslant i\leqslant d} t_i^2\Bigr)dt_1 \cdots dt_d = \Bigl(\int_{\mathbf{R}} e^{-\pi t^2} dt\Bigr)^d = 1.$$

Now comes the important approximation theorem.

Theorem 6.4.6 (Approximation by convolution). *Let $d \geqslant 1$ be an integer and let (φ_n) be a fixed Dirac sequence on $\mathbf{R}^d$.*

(1) *For any p such that $1 \leqslant p < \infty$, and any $f \in L^p(\mathbf{R}^d)$, we have*

$$f \star \varphi_n \to f \text{ in } L^p(\mathbf{R}^d),$$

i.e.,

$$\lim_{n\to+\infty} \int_{\mathbf{R}^d} |f \star \varphi_n(x) - f(x)|^p dx = 0.$$

(2) *If $f \in L^\infty$ is continuous at a point $x \in \mathbf{R}^d$, then*

$$(f \star \varphi_n)(x) \to f(x),$$

and if $f \in L^\infty$ is uniformly continuous on $\mathbf{R}^d$, then

$$f \star \varphi_n \to f \text{ uniformly on } \mathbf{R}^d.$$

Proof. We first check quickly that all convolutions in this statement are among which exist according to the earlier results; for (1) since $\varphi_n \in L^1$ by definition, this is a case of $L^p \star L^1$ (Theorem 6.2.2), so that $f \star \varphi_n \in L^p(\mathbf{R}^d)$ – which also shows that the convergence in L^p makes sense. For (2), this is the case $L^1 \star L^\infty$ with convolution in L^∞ (Proposition 6.2.6 with $p = 1$, $q = \infty$); moreover, by Proposition 6.3.3, the function $f \star \varphi_n$ is itself uniformly continuous on $\mathbf{R}^d$, and it makes sense to speak of its value at the given point x.

To begin the proofs, we use (6.11) cleverly to write

$$f \star \varphi_n(x) - f(x) = \int_{\mathbf{R}^d} (f(x-t) - f(x))\varphi_n(t)dt \tag{6.13}$$

for $n \geqslant 1$ (and only almost x in the first case). We must now show that (in suitable sense) this quantity is small. The idea is that this will come from two reasons: first, $f(x-t) - f(x)$ will be small (maybe on average only) if t is close to 0; then we expect that the contribution of those t which are *not* small is negligible because of the second property of Dirac sequences (6.12).

It is therefore natural to split the integration in two parts, where $\|t\| \leqslant \eta$ and where $\|t\| > \eta$, for some parameter $\eta > 0$ that will be chosen at the end.

We first implement this in the first case with $p < +\infty$. By Hölder's inequality applied to the constant function 1 and to

$$t \mapsto |f(x-t) - f(x)|,$$

integrated with respect to the probability measure $\varphi_n(t)dt$ (here we use the fact that $\varphi_n \geqslant 0$), we obtain

$$|f \star \varphi_n(x) - f(x)|^p \leqslant \int_{\mathbf{R}^d} |f(x-t) - f(x)|^p \varphi_n(t)dt,$$

for almost all x, hence after integrating over $x \in \mathbf{R}^d$, we get

$$\|f \star \varphi_n - f\|_p^p \leqslant S_\eta + T_\eta$$

where $\eta > 0$ is a parameter and

$$S_\eta = \int_{\mathbf{R}^d} \int_{\|t\| \leqslant \eta} |f(x-t) - f(x)|^p \varphi_n(t)dtdx,$$

$$T_\eta = \int_{\mathbf{R}^d} \int_{\|t\| > \eta} |f(x-t) - f(x)|^p \varphi_n(t)dtdx.$$

We deal first with the small values of t. Fix some $\varepsilon > 0$. By Tonelli's Theorem, we have

$$S_\eta = \int_{\|t\| \leqslant \eta} \Bigl(\int_{\mathbf{R}^d} |f(x-t) - f(x)|^p dx \Bigr) \varphi_n(t)dt,$$

and by Proposition 5.5.1, we know that we can select $\eta > 0$ so that

$$\int_{\mathbf{R}^d} |f(x-t) - f(x)|^p dx < \varepsilon,$$

whenever $\|t\| < \eta$. For such a value of η (which we fix), we have therefore

$$S_\eta \leqslant \varepsilon \int_{\|t\| \leqslant \eta} \varphi_n(t)dt \leqslant \varepsilon$$

for all n (using (6.11) and monotony).

Now we come to T_η; here we write the easy upper bound

$$|f(x-t) - f(x)|^p \leqslant 2^{p/q}\{|f(x-t)|^p + |f(x)|^p\},$$

and by Tonelli's Theorem again and the invariance under translation of the Lebesgue measure, we get

$$T_\eta \leqslant 2^p \|f\|_p^p \int_{\|t\| > \eta} \varphi_n(t)dt.$$

Now we can use (6.12): this last quantity goes to 0 as $n \to +\infty$, and therefore, for all n large enough, we have

$$T_\eta \leqslant \varepsilon.$$

Finally, we have found that

$$\|f \star \varphi_n - f\|_p^p \leqslant 2\varepsilon,$$

for all n large enough, and this is the desired conclusion.

We come to case (2) for a fixed value of x first. Starting from (6.13), we have now

$$|f \star \varphi_n(x) - f(x)| \leqslant S_\eta + T_\eta$$

for any fixed $\eta > 0$, with

$$S_\eta = \int_{\|t\| \leqslant \eta} |f(x-t) - f(x)| \varphi_n(t) dt,$$

$$T_\eta = \int_{\|t\| > \eta} (|f(x-t)| + |f(x)|) \varphi_n(t) dt.$$

Since f is continuous at x by assimption, for any $\varepsilon > 0$ we can find some $\eta > 0$ such that

$$|f(x-t) - f(x)| < \varepsilon$$

whenever $\|t\| < \eta$. Fixing η in this manner, we get

$$S_\eta \leqslant \varepsilon$$

for all n and

$$T_\eta \leqslant 2\|f\|_\infty \int_{\|t\| > \eta} \varphi_n(t) dt,$$

from which we conclude as before. Finally, if f is uniformly continuous on $\mathbf{R}^d$, we can select η in such a way that the above upper bound is valid for $x \in \mathbf{R}^d$ uniformly. □

This theorem will allow us to deduce an important corollary.

Corollary 6.4.7. *For any $d \geqslant 1$ and p such that $1 \leqslant p < +\infty$, the vector space $C_c^\infty(\mathbf{R}^d)$ of compactly supported C^∞ functions is dense in $L^p(\mathbf{R}^d)$.*

The idea of the proof is to construct a Dirac sequence in $C_c^\infty(\mathbf{R}^d)$. Because of the examples, above, it is enough to find a single suitable function $\psi \in C_c^\infty(\mathbf{R}^d)$.

Lemma 6.4.8. (1) *For $d \geqslant 1$, there exists $\psi \in C_c^\infty(\mathbf{R}^d)$ such that $\psi \neq 0$ and $\psi \geqslant 0$.*
(2) *For $d \geqslant 1$, there exists a Dirac sequence (φ_n) with $\varphi_n \in C_c^\infty(\mathbf{R}^d)$ for all $n \geqslant 1$.*

Proof. (1) This is a well-known fact from multi-variable calculus; we recall on construction. First, let

$$f : \mathbf{R} \to [0,1]$$

be defined by

$$f(x) = \begin{cases} e^{-1/x} & \text{for } x \geqslant 0 \\ 0 & \text{for } x < 0. \end{cases}$$

We see immediately that f is C^∞ on $\mathbf{R} - \{0\}$. Now, an easy induction argument shows that for $k \geqslant 1$ and $x > 0$, we have

$$f^{(k)}(x) = P_k(x^{-1}) f(x)$$

for some polynomial P_k of degree k. From this, we deduce that

$$\lim_{x \to 0} f^{(k)}(x) = 0$$

for all $k \geqslant 1$. By induction again, this shows that f est C^∞ on $\mathbf{R}$ with $f^{(k)}(0) = 0$. Of course, we have $f \geqslant 0$ and $f \neq 0$.

Now we can simply define

$$\psi(t) = f(1 - \|t\|^2),$$

for $t \in \mathbf{R}^d$. This is a smooth function, non-negative and non-zero, and its support is the unit ball in $\mathbf{R}^d$.

(2) Let ψ be as in (1). Since $\|\psi\|_1 > 0$ (because $\psi \neq 0$ and $\psi \geqslant 0$ and ψ is continuous), we can construct φ_n as in Example 6.4.5, (3), namely

$$\varphi_n(t) = \frac{n^d}{\|\psi\|_1}\psi(nt) \in C_c^\infty(\mathbf{R}^d).$$

$\square$

Proof of the Theorem. Let p with $1 \leqslant p < +\infty$ be given and $f \in L^p(\mathbf{R}^d)$. Fix $\varepsilon > 0$; we first note that by (Théorème 5.4.1), we can find a function

$$g \in C_c(\mathbf{R}^d) \subset L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$$

such that

$$\|f - g\|_p < \varepsilon.$$

Now fix a Dirac sequence (φ_n) in $C_c^\infty(\mathbf{R}^d)$. By Theorem 6.4.6, (1), we have

$$g \star \varphi_n \to g$$

in $L^p(\mathbf{R}^d)$. But by Proposition 6.3.1, we see immediately that $g \star \varphi_n \in C^\infty(\mathbf{R}^d)$ for all n. Moreover (this is why we replaced f by g) we have

$$\operatorname{supp}(g \star \varphi_n) \subset \operatorname{supp}(g) + \operatorname{supp}(\varphi_n)$$

which is compact. Thus the sequence $(g \star \varphi_n)$ is in $C_c^\infty(\mathbf{R}^d)$ and converges to $g \star \varphi_n \to g$ in L^p. In particular, for all n large enough, we find that

$$\|f - g \star \varphi_n\|_p \leqslant 2\varepsilon,$$

and since f and $\varepsilon > 0$ were arbitrary, this implies the density of $C_c(\mathbf{R}^d)$ in $L^p(\mathbf{R}^d)$. $\square$

CHAPTER 7

Questions for the oral examination

Every oral exams consists of answering as good as possible two randomly chosen questions from the following list of questions within 15 minutes. Wrong statements are counted, but it is *not* counted whether all the answers can be given during 15 minutes, i.e. understanding and correct formulation are more important than speed. After posing the two questions three minutes of preparation time are granted. A Definition/Theorem/Remark/Example being stated means that the question is on (some of) the assertions (including proofs). Please try to work out answers in groups or alone as preparation of the exam.

Independent of the chosen question there might be one or two additional small questions on historical personalities, on which you should be prepared via, e.g., wikipedia: Cantor, Fubini, Kolmogorov, Lebesgue, Lévy, Riesz, Tonelli, Vitali.

(1) Definition 1.1.1. and Lemma 1.1.3. with proof.
(2) Example 1.1.4.
(3) Definition 1.1.6 with Remark 1.1.7. (1) and Remark 1.1.7. (2).
(4) Corollary 1.1.10 with proof and Remark 1.1.7. (3).
(5) Lemma 1.1.9. with proof.
(6) Lemma 1.1.12. with proof.
(7) Remark 1.1.13.
(8) Definition 1.2.1 and Proposition 1.2.3 with proof.
(9) Definition 1.2.4 and Remark 1.2.5.
(10) Proposition 1.2.6. with proof.
(11) Example 1.2.7.
(12) Proposition 1.2.8. with proof.
(13) Example 1.3.3 (2).
(14) Proposition 2.1.4. with proof.
(15) Proposition 2.2.2. with proof.
(16) Theorem 2.2.3. with proof.
(17) Proposition 2.2.4. with proof.
(18) Corollary 2.2.5. (1) and (2) with proof.
(19) Corollary 2.2.5. (3) and (4) with proof.
(20) Proposition 2.3.3. with proof.
(21) Fatou's Lemma (Lemma 2.3.8.) with proof.
(22) Dominated convergence (Theorem 2.3.6) with proof.
(23) Lemma 2.3.11. with proof.
(24) Example 2.4.2.
(25) Example 2.4.3.
(26) Proposition 3.1.1. with proof.
(27) Proposition 3.1.4. with proof.
(28) Proposition 3.2.3. with proof.
(29) Proposition 3.3.3. with proof.

(30) Proposition 3.3.4. with proof.
(31) Exercise 3.3.6. with proof.
(32) Proposition 3.3.8. (1) with proof.
(33) Proposition 3.3.8. (2) with proof.
(34) Proposition 3.3.8. (3) with proof.
(35) Proposition 3.3.14. with proof.
(36) Proposition 3.3.17. with proof.
(37) Proposition 3.3.18. with proof.
(38) Proposition 3.4.5. with proof.
(39) Proposition 3.4.4. with proof.
(40) Lemma 4.1.5. with proof.
(41) Proposition 4.1.6. with proof.
(42) Lemma 4.2.1. with proof.
(43) Theorem 4.2.4. with proof.
(44) Theorem 4.3.1. (1) with proof.
(45) Theorem 4.3.1. (2) with proof.
(46) Proposition 4.3.5. with proof.
(47) State and explain Theorem 4.4.6. and Remark 4.4.7. (2).
(48) Proposition 5.1.2. with proof.
(49) State and explain Theorem 5.2.1.
(50) Example 5.2.2.
(51) Uniqueness proof of Theorem 5.2.1.
(52) Proposition 5.3.2. with proof.
(53) Theorem 5.4.1. with proof.
(54) Lemma 5.4.3. with proof.
(55) Proposition 5.5.1. with proof.
(56) A proof of the Riemann Lebesgue Lemma (Theorem 5.5.3.)
(57) Theorem 5.6.1. (1) with proof.
(58) Theorem 5.6.1. (2) with proof.
(59) Lemma 5.6.3. with proof.
(60) Lemma 5.6.4. with proof.
(61) Proof of (5.13) (page 124–125).
(62) Lemma 6.1.2. and proof.
(63) Proposition 6.2.1 and proof.
(64) Theorem 6.2.2. and proof.
(65) Proposition 6.2.4. and proof.
(66) Serie 4, exercise 2.
(67) Serie 6, exercise 1.
(68) Serie 9, exercise 2.
(69) Serie 11, exercise 4.
(70) Serie 11, exercise 5.
(71) Serie 12, exercise 1 – explain the main idea of this counterexample to uniqueness of the Riesz representation theorem.
(72) Serie 13, exercise 3.

Bibliography

[A] W. Appel : *Mathematics for physics... and for physicists*, Princeton Univ. Press 2006.
[Bi] P. Billingsley : *Probability and measure*, 3rd edition, Wiley 1995.
[F] G. Folland : *Real Analysis*, Wiley 1984.
[M] P. Malliavin : *Intégration et probabilités, analyse de Fourier et analyse spectrale*, Masson 1982.
[R] W. Rudin : *Real and complex analysis*, 3rd edition.