2

The toolbox

In this and the following chapter, we introduce basic concepts necessary for understanding the flow of densities. These concepts may be studied in detail before continuing on to the core of our subject matter, which starts in Chapter 4, or, they may be skimmed on first reading to fix the location of important concepts for later reference.

We briefly outline here some essential concepts from measure theory, the theory of Lebesgue integration, and from the theory of the convergence of sequences of functions. This material is in no sense exhaustive; those desiring more detailed treatments should refer to Halmos [1950] and Royden [1968].

2.1 Measures and measure spaces

We start with the definition of a σ-algebra.

Definition 2.1.1. A collection $\mathcal{A}$ of subsets of a set X is a **σ-algebra** if:

(a) When $A \in \mathcal{A}$ then $X\backslash A \in \mathcal{A}$;
(b) Given a finite or infinite sequence $\{A_k\}$ of subsets of X, $A_k \in \mathcal{A}$, then the union $\bigcup_k A_k \in \mathcal{A}$; and
(c) $X \in \mathcal{A}$.

From this definition it follows immediately, by properties (a) and (c), that the empty set $\emptyset$ belongs to $\mathcal{A}$, since $\emptyset = X\backslash X$. Further, given a sequence $\{A_k\}$, $A_k \in \mathcal{A}$, then the intersection $\bigcap_k A_k \in \mathcal{A}$. To see this, note that

$$\bigcap_k A_k = X\backslash \bigcup_k (X\backslash A_k)$$

and then apply properties (a) and (b). Finally, the difference $A\backslash B$ of two sets A and B that belong to $\mathcal{A}$ also belongs to $\mathcal{A}$ because

$$A\backslash B = A \cap (X\backslash B).$$

Definition 2.1.2. A real-valued function μ defined on a σ-algebra $\mathcal{A}$ is a **measure** if:

(a) $\mu(\emptyset) = 0$;
(b) $\mu(A) \geqslant 0$ for all $A \in \mathcal{A}$; and
(c) $\mu(\cup_k A_k) = \Sigma_k \mu(A_k)$ if $\{A_k\}$ is a finite or infinite sequence of pairwise disjoint subsets of $\mathcal{A}$, that is, $A_i \cap A_j = \emptyset$ for $i \neq j$.

We do not exclude the possibility that $\mu(A) = \infty$ for some $A \in \mathcal{A}$.

Remark 2.1.1 This definition of a measure and the properties of a σ-algebra $\mathcal{A}$ as detailed in Definition 2.1.1 ensure that (1) if we know the measure of a set X and a subset A of X we can determine the measure of $X \backslash A$; and (2) if we know the measure of each disjoint subset A_k of $\mathcal{A}$ we can calculate the measure of their union. □

Definition 2.1.3. If $\mathcal{A}$ is a σ-algebra of subsets of X and if μ is a measure on $\mathcal{A}$, then the triple $(X, \mathcal{A}, \mu)$ is called a **measure space.** The sets belonging to $\mathcal{A}$ are called **measurable sets** because, for them, the measure is defined.

Remark 2.1.2. A simple example of a measure space is the finite set $X = \{x_1, \ldots, x_N\}$, in which the σ-algebra is all possible subsets of X and the measure is defined by ascribing to each element $x_i \in X$ a nonnegative number, say p_i. From this it follows that the measure of a subset $\{x_{\alpha_1}, \ldots, x_{\alpha_k}\}$ of X is just $p_{\alpha_1} + \cdots + p_{\alpha_k}$. If $p_i = 1$, then the measure is called a **counting measure** because it counts the number of elements in the set. □

Remark 2.1.3. If $X = [0, 1]$ or R, the real line, then the most natural σ-algebra is the σ-algebra $\mathcal{B}$ of Borel sets (the **Borel σ-algebra**), which, by definition, is the smallest σ-algebra containing intervals. (The word smallest means that any other σ-algebra that contains intervals also contains any set contained in $\mathcal{B}$.) It can be proved that on the Borel σ-algebra there exists a unique measure μ, called the **Borel measure,** such that $\mu([a, b]) = b - a$. □

As presented, Definition 2.1.3 is extremely general. In almost all applications a more specific measure space is adequate, as follows:

Definition 2.1.4 A measure space $(X, \mathcal{A}, \mu)$ is called **σ-finite** if there is a sequence $\{A_k\}$, $A_k \in \mathcal{A}$, satisfying

$$X = \bigcup_{k=1}^{\infty} A_k \quad \text{and} \quad \mu(A_k) < \infty \qquad \text{for all k.}$$

Remark 2.1.4. If $X = R$, the real line, and μ is the Borel measure, then the A_k may be chosen as intervals of the form $[-k, k]$. In the d-dimensional space R^d, the A_k may be chosen as balls of radius k. □

Definition 2.1.5. A measure space $(X, \mathcal{A}, \mu)$ is called **finite** if $\mu(X) < \infty$. In particular, if $\mu(X) = 1$, then the measure space is said to be **normalized** or **probabilistic.**

Remark 2.1.5. We have defined a hierarchy of measure spaces from the most general (Definition 2.1.3) down to the most specific (Definition 2.1.5). *Throughout this book, unless it is specifically stated to the contrary, a measure space will always be understood to be σ-finite.* □

Remark 2.1.6. If a certain property involving the points of a measure space is true except for a subset of that space having measure zero, then we say that property is true **almost everywhere** (abbreviated as a.e.). □

2.2 Lebesgue integration

In the material we deal with it is often necessary to use a type of integration more general than the customary Riemann integration. In this section we introduce the Lebesgue integral, which is defined for abstract measure spaces in which no other structures except a σ-algebra $\mathcal{A}$ and a measure μ must be introduced.

Definition 2.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space. A real-valued function $f: X \to R$ is **measurable** if $f^{-1}(\Delta) \in \mathcal{A}$ for every interval $\Delta \subset R$.

In developing the concept of the Lebesgue integral, we need the notation

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

(see Figure 2.2.1). Observe that

$$f(x) = f^+(x) - f^-(x) \quad \text{and} \quad |f(x)| = f^+(x) + f^-(x).$$

Before presenting the formal definitions for the Lebesgue integral of a function, consider the following. Let $f: X \to R$ be a bounded, nonnegative measurable function, $0 \leq f(x) < M < \infty$. Take the partition of the interval $[0, M]$, $0 = a_0 < a_1 < \cdots < a_n = M$, $a_i = Mi/n$, $i = 0, \ldots, n$, and define the sets A_i by

$$A_i = \{x : f(x) \in [a_i, a_{i+1})\}, \qquad i = 0, \ldots, n-1.$$

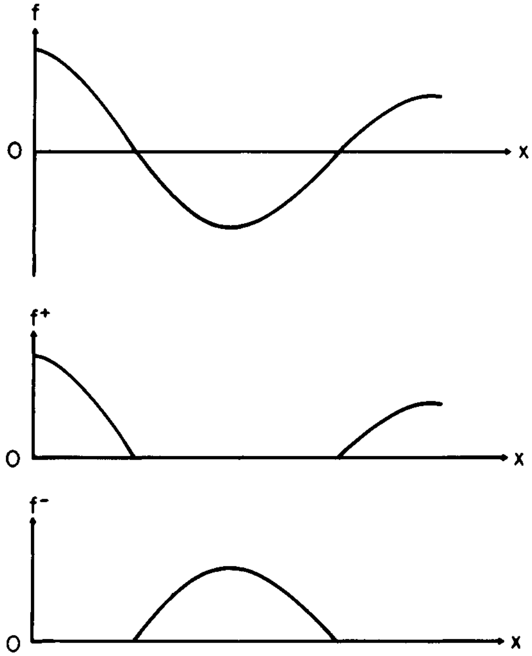

Figure 2.2.1. Illustration of the notation $f^+(x)$ and $f^-(x)$.

Then it is clear that the A_i are measurable and

$$\left| f(x) - \sum_{i=0}^{n-1} a_i 1_{A_i}(x) \right| \leqslant \frac{M}{n}.$$

Therefore, we must conclude that every bounded nonnegative measurable function can be approximated by a finite linear combination of characteristic functions. This observation is crucial to our development of the **Lebesgue integral** embodied in the following four definitions.

Definition 2.2.2. Let $(X, \mathcal{A}, \mu)$ be a measure space, and the sets $A_i \in \mathcal{A}$, $i = 1, \ldots, n$ be such that $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then the Lebesgue integral of the function

$$g(x) = \sum_i \lambda_i 1_{A_i}(x), \qquad \lambda_i \in R, \tag{2.2.1}$$

is defined as

$$\int_X g(x)\, \mu(dx) = \sum_i \lambda_i \mu(A_i).$$

A function g of the form (2.2.1) is called a **simple function.**

Definition 2.2.3. Let $(X, \mathcal{A}, \mu)$ be a measure space, $f: X \to R$ an arbitrary nonnegative bounded measurable function, and $\{g_n\}$ a sequence of simple functions converging uniformly to f. Then the Lebesgue integral of f is defined as

$$\int_X f(x)\,\mu(dx) = \lim_{n\to\infty} \int_X g_n(x)\,\mu(dx).$$

Remark 2.2.1. It can be shown that the limit in Definition 2.2.3 exists and is independent of the choice of the sequence of simple functions $\{g_n\}$ as long as they converge uniformly to f. □

Definition 2.2.4. Let $(X, \mathcal{A}, \mu)$ be a measure space, $f: X \to$ R a nonnegative unbounded measurable function, and define

$$f_M(x) = \begin{cases} f(x) & \text{if } 0 \leq f(x) \leq M \\ M & \text{if } M < f(x). \end{cases}$$

Then the Lebesgue integral of f is defined by

$$\int_X f(x)\,\mu(dx) = \lim_{M\to\infty} \int_X f_M(x)\,\mu(dx).$$

Remark 2.2.2. Note that $\int_X f_M(x)\,\mu(dx)$ is an increasing function of M so that the limit in Definition 2.2.4 always exists even though it might be infinite. □

Definition 2.2.5. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f: X \to R$ a measurable function. Then the Lebesgue integral of f is defined by

$$\int_X f(x)\,\mu(dx) = \int_X f^+(x)\,\mu(dx) - \int_X f^-(x)\,\mu(dx)$$

if at least one of the terms

$$\int_X f^+(x)\,\mu(dx), \quad \int_X f^-(x)\,\mu(dx)$$

is finite. If both of these terms are finite then the function f is called **integrable**.

Remark 2.2.3. The four Definitions 2.2.2–5 are for the Lebesgue integral of f over the entire space X. For $A \in \mathcal{A}$ we have, by definition,

$$\int_A f(x)\,\mu(dx) = \int_X f(x) 1_A(x)\,\mu(dx). \quad \square$$

The Lebesgue integral has some important properties that we will often use. We state them without proof. Throughout a measure space $(X, \mathcal{A}, \mu)$ is assumed.

(L1) If $f, g: X \to R$ are measurable, g is integrable, and $|f(x)| \leqslant g(x)$, then f is integrable and

$$\left|\int_X f(x)\,\mu(dx)\right| \leqslant \int_X g(x)\,\mu(dx).$$

(L2) $\displaystyle\int_X |f(x)|\,\mu(dx) = 0$ if and only if $f(x) = 0$ a.e.

(L3) If $f_1, f_2: X \to R$ are integrable functions, then for $\lambda_1, \lambda_2 \in R$ the linear combination $\lambda_1 f_1 + \lambda_2 f_2$ is integrable and

$$\int_X [\lambda_1 f_1(x) + \lambda_2 f_2(x)]\,\mu(dx)$$
$$= \lambda_1 \int_X f_1(x)\,\mu(dx) + \lambda_2 \int_X f_2(x)\,\mu(dx).$$

(L4) Let $f, g: X \to R$ be measurable functions and $f_n: X \to R$ be measurable functions such that $|f_n(x)| \leqslant g(x)$ and $f_n(x)$ converges to $f(x)$ almost everywhere. If g is integrable, then f and f_n are also integrable and

$$\lim_{n\to\infty} \int_X f_n(x)\,\mu(dx) = \int_X f(x)\,\mu(dx).$$

The last formula is also true if the assumption $|f_n(x)| \leqslant g(x)$ with an integrable g is replaced by $0 \leqslant f_1(x) \leqslant f_2(x)\ldots$. In this case, however, the integrals could be infinite.

Remark 2.2.4. The properties described in (L4) are often referred to as the **Lebesgue dominated convergence theorem** ($|f_n(x)| \leqslant g(x)$) and the **Lebesgue monotone convergence theorem** ($0 \leqslant f_1(x) \leqslant \cdots$). □

(L5) Let $f: X \to R$ be an integrable function and the sets $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$, be disjoint. If $A = \bigcup_i A_i$, then

$$\sum_i \int_{A_i} f(x)\,\mu(dx) = \int_A f(x)\,\mu(dx).$$

Remark 2.2.5. Observe that f is integrable if and only if $|f|$ is integrable. This is easy to see since $|f| = f^+ + f^-$. If f is integrable, f^+ and f^- are also and thus

$$\int_X |f(x)|\,\mu(dx) = \int_X f^+(x)\,\mu(dx) + \int_X f^-(x)\,\mu(dx)$$

is finite. Hence $|f|$ is integrable. The converse is equally easy to prove. □

***Remark* 2.2.6.** Our definition of the Lebesgue integral was stated in four distinct steps. It should be evident from this construction that for every integrable function f there is a sequence of simple functions

$$f_n(x) = \sum_i \lambda_{i,n} 1_{A_{i,n}}(x)$$

such that

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ a.e.} \quad \text{and} \quad |f_n(x)| \leqslant |f(x)|.$$

Thus, by the Lebesgue dominated convergence theorem (L4), we have

$$\lim_{n\to\infty} \int_X f_n(x)\,\mu(dx) = \int_X f(x)\,\mu(dx).$$

This observation will be used many times in simplifying proofs since it enables us to reduce our arguments to two steps: First, we must only verify some formula for simple functions and then, second, pass to the limit. □

***Remark* 2.2.7.** The notion of the Lebesgue integral is quite important since it is defined for very abstract measure spaces $(X, \mathcal{A}, \mu)$ in which no other structures are introduced except for the existence of a σ-algebra $\mathcal{A}$ and a measure μ. In calculus the definition of the Riemann integral is intimately related to the algebraic properties of the real line, and it is easy to establish a connection between the Lebesgue and Riemann integrals. For example, if we define μ as in Remark 2.1.3, then

$$\int_{[a,b]} f(x)\,\mu(dx) = \int_a^b f(x)\,dx$$

where the left-hand side is the Lebesgue integral and the right-hand side is the Riemann integral. This equality is true for any Riemann integrable function f since any Riemann integrable function is automatically Lebesgue integrable. An analogous connection exists in higher dimensions. □

From the properties of the Lebesgue integral it is easy to demonstrate that if $f: X \to R$ is a nonnegative integrable function then $\mu_f(A)$, defined by

$$\mu_f(A) = \int_A f(x)\,\mu(dx),$$

is a finite measure. In fact, by the definition of the Lebesgue integral it is clear that $\mu_f(A)$ is nonnegative and finite, and from property (L5) it is also additive. Further, from (L2) if $\mu(A) = 0$ then

$$\mu_f(A) = \int_A 1_A(x) f(x)\,\mu(dx) = 0$$

since $1_A(x)f(x) = 0$ a.e. Thus $\mu_f(A)$ satisfies all the properties of a measure as detailed in Definition 2.1.2, and $\mu_f(A) = 0$ whenever $\mu(A) = 0$. This observation that every integrable nonnegative function defines a finite measure can be reversed by the following theorem, which is of fundamental importance for the development of the Frobenius–Perron operator.

Theorem 2.2.1. (Radon–Nikodym theorem). Let $(X, \mathcal{A}, \mu)$ be a measure space and let ν be a second finite measure with the property that $\nu(A) = 0$ for all $A \in \mathcal{A}$ such that $\mu(A) = 0$. Then there exists a nonnegative integrable function $f: X \to R$ such that

$$\nu(A) = \int_A f(x)\,\mu(dx) \qquad \text{for all } A \in \mathcal{A}\,.$$

Remark 2.2.8. It should be observed that we have not explicitly stated that $(X, \mathcal{A}, \mu)$ is a σ-finite measure space, which is an important assumption in the Radon–Nikodym theorem. Once again we wish to stress our earlier assumption that *all measure spaces are taken to be σ-finite* unless a contrary assumption is made. □

Although we omit the proof of the Radon–Nikodym theorem, it is easy to show that the function f is in some sense unique. To see this, assume that there are two functions $f_1, f_2: X \to R$ such that

$$\nu(A) = \int_A f_1(x)\,\mu(dx) \quad \text{and} \quad \nu(A) = \int_A f_2(x)\,\mu(dx)\,.$$

Then for all $A \in \mathcal{A}$ we have

$$\int_A [f_1(x) - f_2(x)]\,\mu(dx) = 0\,.$$

Define two sets A_1 and A_2 by

$$A_1 = \{x: f_1(x) > f_2(x)\} \quad \text{and} \quad A_2 = \{x: f_1(x) \leq f_2(x)\}\,.$$

Then

$$\begin{aligned} 0 &= \int_{A_1} [f_1(x) - f_2(x)]\,\mu(dx) - \int_{A_2} [f_1(x) - f_2(x)]\,\mu(dx) \\ &= \int_{A_1 \cup A_2} |f_1(x) - f_2(x)|\,\mu(dx) \\ &= \int_X |f_1(x) - f_2(x)|\,\mu(dx)\,. \end{aligned}$$

Hence, from property (L2) of Lebesgue integrals, we have $|f_1(x) - f_2(x)| = 0$ a.e., so that $f_1(x)$ and $f_2(x)$ differ only on a set of measure zero.

Observe that our argument is quite general, and we have in fact proved the following.

Proposition 2.2.1. If f_1 and f_2 are integrable functions such that

$$\int_A f_1(x)\,\mu(dx) = \int_A f_2(x)\,\mu(dx) \qquad \text{for } A \in \mathcal{A}$$

then $f_1 = f_2$ a.e.

Also from property (L2) of the Lebesgue integral it is clear that two measurable functions, f_1 and f_2, differing from one another only on a set of measure zero, cannot be distinguished by calculating integrals. Thus we say that in the **space of measurable functions,** every two functions f_1, f_2, differing only on a set of measure zero, represent the same **element** of that space. However, to simplify our notation, we will often write "measurable function" instead of "an element of the space of measurable functions." Because of property (L2) this should not lead to any confusion.

With these remarks in mind, we now introduce the concept of an L^p space.

Definition 2.2.6. Let $(X, \mathcal{A}, \mu)$ be a measure space and p a real number, $1 \leq p < \infty$. The family of all possible real-valued measurable functions $f: X \to R$ satisfying

$$\int_X |f(x)|^p\,\mu(dx) < \infty \tag{2.2.2}$$

is the $L^p(X, \mathcal{A}, \mu)$ **space**. Here we use the term "measurable function" to mean "an element of the space of measurable functions."

We shall sometimes write L^p instead of $L^p(X, \mathcal{A}, \mu)$ if the measure space is understood, or $L^p(X)$ if $\mathcal{A}$ and μ are understood. Note that if $p = 1$ then the L^1 space consists of all possible integrable functions.

The integral appearing in (2.2.2) is very important for an element $f \in L^p$. Thus it is assigned the special notation

$$\|f\|_{L^p} = \left[\int_X |f(x)|^p\,\mu(dx)\right]^{1/p} \tag{2.2.3}$$

and is called the L^p **norm** of f. When property (L2) of the Lebesgue integral is applied to $|f|^p$, it immediately follows that the condition $\|f\|_{L^p} = 0$ is equivalent to $f(x) = 0$ a.e. Or, more precisely, $\|f\|_{L^p} = 0$ if and only if f is a zero element

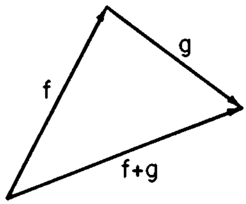

Figure 2.2.2. A geometric interpretation of the triangle inequality (2.2.5).

in L^p (which is an element represented by all functions equal to zero almost everywhere).

Two other important properties of the norm are

$$\|\alpha f\|_{L^p} = |\alpha| \cdot \|f\|_{L^p} \qquad \text{for } f \in L^p,\ \alpha \in R \tag{2.2.4}$$

and

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p} \qquad \text{for } f, g \in L^p \tag{2.2.5}$$

The first condition, (2.2.4), simply says that the norm is homogeneous. The second is called the **triangle inequality.** As shown in Figure 2.2.2, if we think of f, g, and $f + g$ as vectors, we can consider a triangle with sides f, g, and $f + g$. Then, by equation, (2.2.5), the length of the side $(f + g)$ is shorter than the sum of the lengths of the other two sides.

From (2.2.4) it follows that for every $f \in L^1$ and real α, the product αf belongs to L^p. Further, from (2.2.5) it follows that for every $f, g \in L^p$ the sum $f + g$ is also an element of L^p. This is denoted by saying that L^p is a **vector space.**

Because the value of $\|f\|_{L^p}$ is interpreted as the length of f, we say that

$$\|f - g\|_{L^p} = \left[\int_X |f(x) - g(x)|^p \, \mu(dx) \right]^{1/p}$$

is the L^p **distance** between f and g.

It is important to note that the product fg of two functions $f, g \in L^p$ is not necessarily in L^p, for example, $f(x) = x^{-1/2}$ is integrable on $[0, 1]$ but $[f(x)]^2 = x^{-1}$ is not.

This leads us to define the space adjoint to L^p.

Definition 2.2.7. Let $(X, \mathcal{A}, \mu)$ be a measure space. The **space adjoint** to $L^p(X, \mathcal{A}, \mu)$ is $L^{p'}(X, \mathcal{A}, \mu)$, where

$$(1/p) + (1/p') = 1 .$$

Remark 2.2.9. If $p = 1$, Definition 2.2.7 of adjoint space fails. The adjoint space, in the case $p = 1$, by definition, consists of all bounded almost everywhere measurable functions and is denoted by L^∞. Functions that differ only on a set of measure zero are considered to represent the same element. □

It is well known that if $f \in L^p$ and $g \in L^{p'}$, then fg is integrable, and hence we define the **scalar product** of two functions by

$$\langle f, g\rangle = \int_X f(x)g(x)\,\mu(dx).$$

An important relation we will often use is the **Cauchy–Hölder inequality.** Thus, if $f \in L^p$ and $g \in L^{p'}$, then

$$|\langle f, g\rangle| \leqslant \|f\|_{L^p} \cdot \|g\|_{L^{p'}}.$$

For this inequality to make sense when $f \in L^1$, $g \in L^\infty$, we take the L^∞ norm of g to be the smallest constant c such that

$$|g(x)| \leqslant c$$

for almost all $x \in X$. This constant is denoted by ess sup$|g(x)|$, called the **essential supremum** of g.

Remark 2.2.10. As we almost always work in L^1, we will not indicate the space in which the norm is taken unless it is not L^1. Thus we will write $\|f\|$ instead of $\|f\|_{L^1}$. Observe that in L^1 the norm has the exceptional property that the triangle inequality is sometimes an equality. To see this, note from property (L3) that

$$\|f + g\| = \|f\| + \|g\| \qquad \text{for } f \geqslant 0,\ g \geqslant 0;\ f, g \in L^1.$$

Thus geometrical intuition in some abstract spaces may be misleading. □

The concept of the L^1 space simplifies the Radon–Nikodym theorem as shown by the following corollary.

Corollary 2.2.1. If $(X, \mathcal{A}, \mu)$ is a measure space and ν is a finite measure on $\mathcal{A}$ such that $\nu(A) = 0$ whenever $\mu(A) = 0$, then there exists a unique element $f \in L^1$ such that

$$\nu(A) = \int_A f(x)\,\mu(dx) \qquad \text{for } A \in \mathcal{A}.$$

One of the most important notions in analysis, measure theory, and topology, as well as other areas of mathematics, is that of the Cartesian product. To introduce this concept we start with a definition.

Definition 2.2.8. Given two arbitrary sets A_1 and A_2, the **Cartesian product** of A_1 and A_2 (note that the order is important) is the set of all pairs (x_1, x_2) such that $x_1 \in A_1$ and $x_2 \in A_2$. This is customarily written as

$$A_1 \times A_2 = \{(x_1, x_2): x_1 \in A_1,\ x_2 \in A_2\}.$$

In a natural way this concept may be extended to more than two sets. Thus the Cartesian product of the sets $A_1, \ldots, A_d$ is the set of all sequences $(x_1, \ldots, x_d)$ such that $x_i \in A_i$, $i = 1, \ldots, d$, or

$$A_1 \times \cdots \times A_d = \{(x_1, \ldots, x_d): x_i \in A_i \text{ for } i = 1, \ldots, d\}.$$

An important consequence following from the concept of the Cartesian product is that if a structure is defined on each of the factors A_i, for example, a measure, then it is possible to extend that property to the Cartesian product. Thus, given d measure spaces $(X_i, \mathcal{A}_i, \mu_i)$, $i = 1, \ldots, d$, we define

$$X = X_1 \times \cdots \times X_d, \tag{2.2.6}$$

$\mathcal{A}$ to be the smallest σ-algebra of subsets of X containing all sets of the form

$$A_1 \times \cdots \times A_d \qquad \text{with } A_i \in \mathcal{A}_i,\ i = 1, \ldots, d, \tag{2.2.7}$$

and

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \cdots \mu_d(A_d) \qquad \text{for } A_i \in \mathcal{A}_i. \tag{2.2.8}$$

Unfortunately, by themselves they do not define a measure space $(X, \mathcal{A}, \mu)$. There is no problem with either X or $\mathcal{A}$, but μ is defined only on special sets, namely $A = A_1 \times \cdots \times A_d$, that do not form a σ-algebra. To show that μ, as defined by (2.2.8), can be extended to the entire σ-algebra $\mathcal{A}$ requires the following theorem.

Theorem 2.2.2. If measure spaces $(X_i, \mathcal{A}_i, \mu_i)$, $i = 1, \ldots, d$ are given and X, $\mathcal{A}$, and μ are defined by equations (2.2.6), (2.2.7), and (2.2.8), respectively, then there exists a unique extension of μ to a measure defined on $\mathcal{A}$.

The measure space $(X, \mathcal{A}, \mu)$, whose existence is guaranteed by Theorem 2.2.2, is called the **product of the measure spaces** $(X_1, \mathcal{A}_1, \mu_1), \ldots, (X_d, \mathcal{A}_d, \mu_d)$, or more briefly a **product space.** The measure μ is called the **product measure.**

Observe that from equation (2.2.8) it follows that

$$\mu(X_1 \times \cdots \times X_d) = \mu(X_1) \cdots \mu(X_d).$$

Thus, if all the measure spaces $(X_i, \mathcal{A}_i, \mu_i)$ are finite or probabilistic, then $(X, \mathcal{A}, \mu)$ will also be finite or probabilistic.

Theorem 2.2.2 allows us to define integration on the product space $(X, \mathcal{A}, \mu)$ since it is also a measure space. A function $f: X \to R$ may be written as a function of d variables because every point $x \in X$ is a sequence $x = (x_1, \dots, x_d)$, $x_i \in X_i$. Thus it is customary to write integrals on X either as

$$\int_X f(x)\,\mu(dx)\,,$$

where it is implicitly understood that $x = (x_1, \dots, x_d)$ and $X = X_1 \times \cdots \times X_d$, or in the more explicit form

$$\int_{X_1} \cdots \int_{X_d} f(x_1, \dots, x_d)\,\mu(dx_1 \cdots dx_d)\,.$$

Integrals on the product of measure spaces are related to integrals on the individual factors by a theorem associated with the name of Fubini. For simplicity, we first formulate it for product spaces containing only two factors.

Theorem 2.2.3 (Fubini's theorem). Let $(X, \mathcal{A}, \mu)$ be the product space formed by $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$, and let a μ integrable function $f: X \to R$ be given. Then, for almost every x_1, the function $f(x_1, x_2)$ is μ_2 integrable with respect to x_2. Furthermore the function

$$\int_{X_2} f(x_1, x_2)\,\mu_2(dx_2)$$

of the variable x_1 is μ_1 integrable and

$$\int_{X_1} \left\{ \int_{X_2} f(x_1, x_2)\,\mu_2(dx_2) \right\} \mu_1(dx_1) = \iint_X f(x_1, x_2)\,\mu(dx_1\,dx_2)\,. \qquad (2.2.9)$$

Theorem 2.2.3 extends, in a natural way, to product spaces with an arbitrary number of factors. If $(X, \mathcal{A}, \mu)$ is the product of the measure spaces $(X_i, \mathcal{A}_i, \mu_i)$, $i = 1, \dots, d$, and $f: X \to R$ is μ integrable, then

$$\int \cdots \int_X f(x_1, \dots, x_d)\,\mu(dx_1 \cdots dx_d)$$
$$= \int_{X_1} \left\{ \cdots \int_{X_{d-1}} \left[\int_{X_d} f(x_1, \dots, x_d)\,\mu_d(dx_d) \right] \mu_{d-1}(dx_{d-1}) \cdots \right\} \mu_1(dx_1)\,. \qquad (2.2.10)$$

Remark 2.2.11. As we noted in Remark 2.1.3, the "natural" Borel measure on the real line R is defined on the smallest σ-algebra $\mathcal{B}$ that contains all intervals.

For every interval $[a,b]$ this measure satisfies $\mu([a,b]) = b - a$. Having the structure $(R, \mathcal{B}, \mu)$, we define by Theorem 2.2.2 the product space $(R^d, \mathcal{B}^d, \mu^d)$, where

$$R^d = R \times \overset{d}{\cdots} \times R,$$

$\mathcal{B}^d$ is the smallest σ-algebra containing all sets of the form

$$A_1 \times \cdots \times A_d \qquad \text{with } A_i \in \mathcal{B},$$

and

$$\mu^d(A_1 \times \cdots \times A_d) = \mu(A_1) \cdots \mu(A_d). \tag{2.2.11}$$

The measure μ^d is again called the Borel measure. It is easily verified that $\mathcal{B}^d$ may be alternately defined as either the smallest σ-algebra containing all the rectangles

$$[a_1, b_1] \times \cdots \times [a_d, b_d],$$

or as the smallest σ-algebra containing all the open subsets of R^d. From (2.2.11) it follows that

$$\mu^d([\mathrm{a}_1, \mathrm{b}_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d),$$

which is the classical formula for the volume of an n-dimensional box.

The same construction may be repeated by starting, not from the whole real line R, but from the unit interval $[0,1]$ or from any other finite interval. Thus, from Theorem 2.2.2, we will obtain the Borel measure on the unit square $[0,1] \times [0,1]$ or on the d-dimensional cube

$$[0,1]^d = [0,1] \times \overset{d}{\cdots} \times [0,1].$$

In all cases $(R^d, [0,1]^d,$ etc.) we will omit the superscript d on $\mathcal{B}^d$ and μ^d and write $(R^d, \mathcal{B}, \mu)$ instead of $(R^d, \mathcal{B}^d, \mu^d)$. Furthermore, in all cases when the space is R, R^d, or any subset of these ($[0,1]$, $[0,1]^d$, $R^+ = [0, \infty)$, etc.) and the measure and σ-algebra are not specified, we will assume that the measure space is taken with the Borel σ-algebra and Borel measure. Finally, all the integrals on R or R^d taken with respect to the Borel measure will be written with dx instead of $\mu(dx)$. □

Remark 2.2.12. From the additivity property of a measure (Definition 2.1.2c) it follows that every measure is **monotonic**, that is, if A and B are measurable sets and $A \subset B$ then $\mu(A) \leq \mu(B)$. This follows directly from

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

Thus, if $\mu(B) = 0$ and $A \subset B$, then $\mu(A) = 0$. However, it could happen that $A \subset B$ and B is a measurable set while A is not. In this case, if $\mu(B) = 0$, then it does not follow that $\mu(A) = 0$, because $\mu(A)$ is not defined, which is a peculiar situation.

It is rather natural, therefore, to require that a "good" measure have the property that subsets of measurable sets of measure zero should also be measurable with, of course, measure zero. If a measure has this property it is called **complete**. Indeed, it can be proved that, if $(X, \mathcal{A}, \mu)$ is a measure space, then there exists a smallest σ-algebra $\mathcal{A}_1 \supset \mathcal{A}$ and a measure μ_1 on $\mathcal{A}_1$ identical with μ on $\mathcal{A}$ such that $(X_1, \mathcal{A}_1, \mu_1)$ is complete.

Every Borel measure on R (or R^d, $[0, 1]$, $[0, 1]^d$, etc.) can be completed. This complete measure is called the **Lebesgue measure.** However, when working in R (or R^d, etc.), we will use the Borel measure and *not* the Lebesgue measure, because, with the Lebesgue measure, we encounter problems with the measurability of the composition of measurable functions that are avoided with the Borel measure. □

2.3 Convergence of sequences of functions

Having defined L^p spaces and introduced the notions of norms and scalar products, we now consider three different types of convergence for a sequence of functions.

Definition 2.3.1. A sequence of functions $\{f_n\}$, $f_n \in L^p$, $1 \leqslant p < \infty$, is **(weakly) Cesaro convergent** to $f \in L^p$ if

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \langle f_k, g \rangle = \langle f, g \rangle \qquad \text{for all } g \in L^{p'}. \tag{2.3.1}$$

Definition 2.3.2. A sequence of functions $\{f_n\}, f_n \in L^p$, $1 \leqslant p < \infty$, is **weakly convergent** to $f \in L^p$ if

$$\lim_{n \to \infty} \langle f_n, g \rangle = \langle f, g \rangle \qquad \text{for all } g \in L^{p'}. \tag{2.3.2}$$

Definition 2.3.3. A sequence of functions $\{f_n\}, f_n \in L^p$, $1 \leqslant p \leqslant \infty$, is **strongly convergent** to $f \in L^p$ if

$$\lim_{n \to \infty} \|f_n - f\|_{L^p} = 0\,. \tag{2.3.3}$$

From the Cauchy–Hölder inequality, we have

$$|\langle f_n - f, g \rangle| \leqslant \|f_n - f\|_{L^p} \cdot \|g\|_{L^{p'}}\,,$$

and, thus, if $\|f_n - f\|_{L^p}$ converges to zero, so must $\langle f_n - f, g \rangle$. Hence strong convergence implies weak convergence, and the condition for strong convergence is relatively straightforward to check. However, the condition for weak convergence requires a demonstration that it holds for all $g \in L^{p'}$, which seems difficult to do at first glance. In some special and important spaces, it is sufficient to check weak convergence for a restricted class of functions, defined as follows.

Definition 2.3.4. A subset $K \subset L^p$ is called **linearly dense** if for each $f \in L^p$ and $\varepsilon > 0$ there are $g_1, \ldots, g_n \in K$ and constants $\lambda_1, \ldots, \lambda_n$, such that

$$\|f - g\|_{L^p} < \varepsilon,$$

where

$$g = \sum_{i=1}^{n} \lambda_i g_i .$$

By using the notion of linearly dense sets, it is possible to simplify the proof of weak convergence. If the sequence $\{f_n\}$ is bounded in norm, that is, $\|f_n\|_{L^p} \leqslant c < \infty$, and if K is linearly dense in $L^{p'}$, then it is sufficient to check weak convergence in Definition 2.3.2 for any $g \in K$.

It is well known that in the space $L^p([0,1])$ $(1 \leqslant p < \infty)$ the following sets are linearly dense:

$$K_1 = \{\text{the set of characteristic functions } 1_\Delta(x) \text{ of the Borel sets } \Delta \subset [0,1]\},$$

$$K_2 = \{\text{the set of continuous functions on } [0,1]\},$$

$$K_3 = \{\sin(n\pi x);\ n = 1, 2, \ldots\}.$$

In K_1 it is enough to take a family of sets Δ that are generators of Borel sets on $[0,1]$, for example, $\{\Delta\}$ could be the family of subintervals of $[0,1]$. Observe that the linear density of K_3 follows from the Fourier expansion theorem. In higher dimensions, for instance on a square in the plane, we may take analogous sets K_1 and K_2 but replace K_3 with

$$K_3' = \{\sin(m\pi x)\,\sin(n\pi y)\colon n, m = 1, 2, \ldots\}.$$

Example 2.3.1. Consider the sequence of functions $f_n(x) = \sin(nx)$ on $L^2([0,1])$. We are going to show that $\{f_n\}$ converges weakly to $f \equiv 0$. First observe that

$$\|f_n\|_{L^2} = \left(\int_0^1 \sin^2 nx\,dx\right)^{1/2} = \left|\frac{1}{2} - \frac{\sin 2n}{4n}\right|^{1/2} \leqslant 1,$$

and hence the sequence $\{\|f_n\|_{L^2}\}$ is bounded. Now take an arbitrary function $g(x) = \sin(m\pi x)$ from K_3. We have

$$\langle f_n, g\rangle = \int_0^1 \sin(nx)\,\sin(m\pi x)\,dx$$
$$= \frac{\sin(n - m\pi)}{2(n - m\pi)} - \frac{\sin(n + m\pi)}{2(n + m\pi)}$$

so that

$$\lim_{n\to\infty}\langle f_n, g\rangle = \langle 0, g\rangle = 0, \qquad \text{for } g \in K_3$$

and $\{f_n\}$ thus converges weakly to $f = 0$. □

We have seen that, in a given L^p space, strong convergence implies weak convergence. It also turns out that we may compare convergence in different L^p spaces using the following proposition.

Proposition 2.3.1. If $(X, \mathscr{A}, \mu)$ is a finite measure space and $1 \leqslant p_1 < p_2 \leqslant \infty$, then

$$\|f\|_{L^{p_1}} \leqslant c\|f\|_{L^{p_2}} \qquad \text{for every } f \in L^{p_2} \tag{2.3.4}$$

where c depends on $\mu(X)$. Thus every element of L^{p_2} belongs to L^{p_1}, and strong convergence in L^{p_2} implies strong convergence in L^{p_1}.

Proof: Let $f \in L^{p_2}$ and let $p_2 < \infty$. By setting $g = |f|^{p_1}$, we obtain

$$\|f\|_{L^{p_1}}^{p_1} = \int_X |f|^{p_1}\,\mu(dx) = \langle 1, g\rangle\,.$$

Setting $p' = p_2/p_1$ and denoting by p the number adjoint to p', that is, $(1/p) + (1/p') = 1$, we have

$$\langle 1, g\rangle \leqslant \|1\|_{L^p}\cdot\|g\|_{L^{p'}} = \left[\int_X \mu(dx)\right]^{1/p}\left[\int_X |f|^{p_1p'}\,\mu(dx)\right]^{1/p'}$$
$$= \mu(X)^{1/p}\|f\|_{L^{p_2}}^{p_1}$$

and, consequently,

$$\|f\|_{L^{p_1}}^{p_1} \leqslant \mu(X)^{1/p}\|f\|_{L^{p_2}}^{p_1},$$

which proves equation (2.3.4). Hence, if $\|f\|_{L^{p_2}}$ is finite, then $\|f\|_{L^{p_1}}$ is also finite, proving that L^{p_2} is contained in L^{p_1}. Furthermore, the inequality

$$\|f_n - f\|_{L^{p_1}} \leqslant c\|f_n - f\|_{L^{p_2}}$$

implies that strong convergence in L^{p_2} is stronger than strong convergence in L^{p_1}. If $p_2 = \infty$, the inequality (2.3.4) is obvious, and thus the proof is complete. ■

Observe that the strong convergence of f_n to f in L^1 (with arbitrary measure) as well as the strong convergence of f_n to f in L^p ($p > 1$) with finite measure both imply

$$\lim_{n\to\infty} \int_X f_n \, \mu(dx) = \int_X f \, \mu(dx) .$$

To see this simply note that

$$\left| \int_X f_n \, \mu(dx) - \int_X f \, \mu(dx) \right| \leqslant \int_X |f_n - f| \, \mu(dx) = \|f_n - f\|_{L^1} \leqslant c\|f_n - f\|_{L^p} .$$

It is often necessary to define a function as a limit of a convergent sequence and/or as a sum of a convergent series. Thus the question arises how to show that a sequence $\{f_n\}$ is convergent if the limit is unknown. The famous **Cauchy condition for convergence** provides such a tool. To understand this condition, first assume that $\{f_n\}$, $f_n \in L^p$, is strongly convergent to f. Take $\varepsilon > 0$. Then there is an integer n_0 such that

$$\|f_n - f\|_{L^p} \leqslant \tfrac{1}{2}\varepsilon \qquad \text{for } n \geqslant n_0$$

and, in particular,

$$\|f_{n+k} - f\|_{L^p} \leqslant \tfrac{1}{2}\varepsilon \qquad \text{for } n \geqslant n_0 \quad \text{and} \quad k \geqslant 0 .$$

From this and the triangle inequality, we obtain

$$\|f_{n+k} - f_n\|_{L^p} \leqslant \|f_{n+k} - f\|_{L^p} + \|f - f_n\|_{L^p} \leqslant \varepsilon .$$

Thus we have proved that, if $\{f_n\}$ is strongly convergent in L^p to f, then

$$\lim_{n\to\infty} \|f_{n+k} - f_n\|_{L^p} = 0 \qquad \text{uniformly for all } k \geqslant 0 . \tag{2.3.5}$$

This is the Cauchy condition for convergence.

It can be proved that all L^p spaces ($1 \leqslant p \leqslant \infty$) have the property that condition (2.3.5) is also sufficient for convergence. This is stated more precisely in the following theorem.

Theorem 2.3.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\{f_n\}, f_n \in L^p(X, \mathcal{A}, \mu)$ be a sequence such that equation (2.3.5) holds. Then there exists an element $f \in L^p(X, \mathcal{A}, \mu)$ such that $\{f_n\}$ converges strongly to f, that is, condition (2.3.3) holds.

The fact that Theorem 2.3.1 holds for L^p spaces is referred to by saying that L^p spaces are **complete**.

Theorem 2.3.1 enables us to prove the convergence of series by the use of a **comparison series.** Suppose we have a sequence $\{g_n\} \subset L^p$ and we know the series of norms $\|g_n\|_{L^p}$ is convergent, that is,

$$\sum_{n=0}^{\infty} \|g_n\|_{L^p} < \infty . \tag{2.3.6}$$

Then, using Theorem 2.3.1, it is easy to verify that the series

$$\sum_{n=0}^{\infty} g_n \tag{2.3.7}$$

is also strongly convergent and that its sum is an element of L^p.

To see this note that the convergence of (2.3.7) simply means that the sequence of partial sums

$$s_n = \sum_{m=0}^{n} g_m$$

is convergent. To verify that $\{s_n\}$ is convergent, set

$$\sigma_n = \sum_{m=0}^{n} \|g_m\|_{L^p} .$$

From equation (2.3.6) the sequence of real numbers $\{\sigma_n\}$ is convergent and, therefore, the Cauchy condition holds for this sequence. Thus

$$\lim_{n\to\infty} |\sigma_{n+k} - \sigma_n| = 0 \qquad \text{uniformly for } k \geqslant 0 .$$

Further

$$\|s_{n+k} - s_n\|_{L^p} = \left\| \sum_{m=n+1}^{n+k} g_m \right\|_{L^p} \leqslant \sum_{m=n+1}^{n+k} \|g_m\|_{L^p} = |\sigma_{n+k} - \sigma_n|$$

so finally

$$\lim_{n\to\infty} \|s_{n+k} - s_n\|_{L^p} = 0 \qquad \text{uniformly for } k \geqslant 0 ,$$

which is the Cauchy condition for $\{s_n\}$.