Chapter 1

Measure and Integration

1.1 INTRODUCTION

The most important analytic tool used in this book is integration. The student of analysis meets this concept in a calculus course where an integral is defined as a Riemann integral. While this point of view of integration may be historically grounded and useful in many areas of mathematics, it is far from being adequate for the requirements of modern analysis. The difficulty with the Riemann integral is that it can be defined only for a special class of functions and this class is not closed under the process of taking pointwise limits of sequences (not even monotonic sequences) of functions in this class. Analysis, it has been said, is the art of taking limits, and the constraint of having to deal with an integration theory that does not allow taking limits is much like having to do mathematics only with rational numbers and excluding the irrational ones.

If we think of the graph of a real-valued function of n variables, the integral of the function is supposed to be the $(n+1)$-dimensional volume under the graph. The question is how to define this volume. The Riemann integral attempts to define it as 'base times height' for small, predetermined n-dimensional cubes as bases, with the height being some 'typical' value of the function as the variables range over that cube. The difficulty is that it may be impossible to define this height properly if the function is sufficiently discontinuous.

The useful and far-reaching idea of Lebesgue and others was to compute the $(n+1)$-dimensional volume 'in the other direction' by first computing the n-dimensional volume of the set where the function is greater than some number y. This volume is a well-behaved, monotone nonincreasing function of the number y, which then can be integrated in the manner of Riemann.

This method of integration not only works for a large class of functions (which is closed under taking pointwise limits), but it also greatly simplifies a problem that used to plague analysts: Is it permissible to exchange limits and integration?

In this chapter we shall first sketch in the briefest possible way the ideas about measure that are needed in order to define integrals. Then we shall prove the most important convergence theorems which permit us to interchange limits and integration. Many measure-theoretic details are not given here because the subject is lengthy and complicated and is presented in any number of texts, e.g. [Rudin, 1987]. The most important reason for omitting the measure theory is that the intricacies of its development are not needed for its exploitation. For instance, we all know the tremendously important fact that

$$\int (f+g) = \left(\int f\right) + \left(\int g\right),$$

and we can use it happily without remembering the proof (which actually does require some thought); the interested reader can carry out the proof, however, in Exercise 9. Nevertheless we want to emphasize that this theory is one of the great triumphs of twentieth century mathematics and it is the culmination of a long struggle to find the right perspective from which to view integration theory. We recommend its study to the reader because it is the foundation on which this book ultimately rests.

Before dealing with integration, let us review some elementary facts and notation that will be needed. The real numbers are denoted by $\mathbb{R}$, while the complex numbers are denoted by $\mathbb{C}$ and $\bar{z}$ is the complex conjugate of z. It will be assumed that the reader is equipped with a knowledge of the fundamentals of the calculus on **n-dimensional Euclidean space**

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : \text{each } x_i \text{ is in } \mathbb{R}\}.$$

The **Euclidean distance** between two points y and z in $\mathbb{R}^n$ is defined to be $|y-z|$ where, for $x \in \mathbb{R}^n$,

$$|x| := \left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$

(The symbols $a := b$ and $b =: a$ mean that a is defined by b.) We expect the reader to know some elementary inequalities such as the **triangle inequality**,

$$|x| + |y| \geq |x-y|.$$

The definition of **open sets** (a set, each of whose points is at the center of some ball contained in the set), **closed sets** (the complement of an open set), **compact sets** (closed and bounded subsets of $\mathbb{R}^n$), **connected sets** (see Exercise 1.23), limits, the Riemann integral and differentiable functions are among the concepts we assume known. $[a, b]$ denotes the **closed interval** in $\mathbb{R}$, $a \leq x \leq b$, while (a, b) denotes the **open interval** $a < x < b$. The notation $\{a : b\}$ means, of course, the set of all things of type a that satisfy condition b. We introduce here the useful notation

$$C^k(\Omega)$$

to describe the complex-valued functions on some open set $\Omega \subset \mathbb{R}^n$ that are k times continuously differentiable (i.e., the partial derivatives $\partial^k f/\partial x_{i_1}, \dots, \partial x_{i_k}$ exist at all points $x \in \Omega$ and are continuous functions on Ω). If a function f is in $C^k(\Omega)$ for all k, then we write $f \in C^\infty(\Omega)$.

In general, if f is a function from some set A (e.g., some subset of $\mathbb{R}^n$) with values in some set B (e.g., the real numbers), we denote this fact by $f \colon A \to B$. If $x \in A$, we write $x \mapsto f(x)$, the bar on the arrow serving to distinguish the image of a single point x from the image of the whole set A.

An important class of functions consists of the **characteristic functions** of sets. If A is a set we define

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{1}$$

These will serve as building blocks for more general functions (see Sect. 1.13, Layer cake representation). Note that $\chi_A \chi_B = \chi_{A \cap B}$.

Recall that the **closure of a set** $A \subset \mathbb{R}^n$ is the smallest closed set in $\mathbb{R}^n$ that contains A. We denote the closure by $\overline{A}$. Thus, $\overline{\overline{A}} = \overline{A}$. The **support of a continuous function** $f : \mathbb{R}^n \to \mathbb{C}$, denoted by $\operatorname{supp}\{f\}$, is the *closure* of the set of points $x \in \mathbb{R}^n$ where $f(x)$ is nonzero, i.e.,

$$\operatorname{supp}\{f\} = \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}.$$

It is important to keep in mind that the above definition is a topological notion. Later, in Sect. 1.5, we shall give a definition of *essential* support for measurable functions. We denote the set of functions in $C^\infty(\Omega)$ whose support is bounded and contained in Ω by $C_c^\infty(\Omega)$. The subscript c stands for 'compact' since a set is closed and bounded if and only if it is compact.

Here is a classic example of a compactly supported, infinitely differentiable function on $\mathbb{R}^n$; its support is the unit ball $\{x \in \mathbb{R}^n : |x| \leq 1\}$:

$$j(x) = \begin{cases} \exp\left[-\frac{1}{1-|x|^2}\right] & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1. \end{cases} \tag{2}$$

The verification that j is actually in $C^\infty(\mathbb{R}^n)$ is left as an exercise.

This example can be used to prove a version of what is known as **Urysohn's lemma** in the $\mathbb{R}^n$ setting. Let $\Omega \subset \mathbb{R}^n$ be an open set and let $K \subset \Omega$ be a compact set. Then there exists a nonnegative function $\psi \in C_c^\infty(\Omega)$ with $\psi(x) = 1$ for $x \in K$. An outline of the proof is given in Exercise 15.

1.2 BASIC NOTIONS OF MEASURE THEORY

Before trying to define a measure of a set one must first study the structure of sets that are **measurable**, i.e., those sets for which it will prove to be possible to associate a numerical value in an unambiguous way. Not necessarily all sets will be measurable.

We begin, generally, with a set Ω whose elements are called **points**. For orientation one might think of Ω as a subset of $\mathbb{R}^n$, but it might be a much more general set than that, e.g., the set of paths in a path-space on which we are trying to define a 'functional integral'.

A distinguished collection, Σ, of subsets of Ω is called a **sigma-algebra** if the following axioms are satisfied:

(i) If $A \in \Sigma$, then $A^c \in \Sigma$, where $A^c := \Omega \sim A$ is the **complement** of A in Ω. (Generally, $B \sim A := B \cap A^c$.)

(ii) If $A_1, A_2, \ldots$ is a countable family of sets in Σ, then their union $\bigcup_{i=1}^\infty A_i$ is also in Σ.

(iii) $\Omega \in \Sigma$.

Note that these assumptions imply that the empty set $\varnothing$ is in Σ and that Σ is also closed under countable intersections, i.e., if $A_1, A_2, \ldots \in \Sigma$, then $\bigcap_{i=1}^\infty A_i \in \Sigma$. Also, $A_1 \sim A_2$ is in Σ.

It is a trivial fact that *any* family $\mathcal{F}$ of subsets of Ω can be extended to a sigma-algebra (just take the sigma-algebra consisting of all subsets of Ω). Among all these extensions there is a special one. Consider all the sigma-algebras that contain $\mathcal{F}$ and take their intersection, which we call Σ, i.e., a subset $A \subset \Omega$ is in Σ if and only if A is in every sigma-algebra containing $\mathcal{F}$. It is easy to check that Σ is indeed a sigma-algebra. Indeed it is the **smallest sigma-algebra containing** $\mathcal{F}$; it is also called the **sigma-algebra generated by** $\mathcal{F}$. An important example is the sigma-algebra $\mathcal{B}$ of **Borel sets** of $\mathbb{R}^n$ which is generated by the open subsets of $\mathbb{R}^n$. Alternatively, it is generated by the **open balls** of $\mathbb{R}^n$, i.e., the family of sets of the form

$$B_{x,R} = \{y \in \mathbb{R}^n : |x - y| < R\}. \tag{1}$$

It is a fact that this Borel sigma-algebra contains the closed sets by (i) above. With the help of the axiom of choice one can prove that $\mathcal{B}$ does *not* contain *all* subsets of $\mathbb{R}^n$, but we emphasize that the reader does not need to know either this fact or the axiom of choice.

A **measure** (sometimes also called a **positive measure** for emphasis) μ, defined on a sigma-algebra Σ, is a function from Σ into the nonnegative real numbers (including infinity) such that $\mu(\varnothing) = 0$ and with the following crucial property of **countable additivity**. If $A_1, A_2, \ldots$ is a sequence of *disjoint sets* in Σ, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{2}$$

The big breakthrough, historically, was the realization that countable additivity is an essential requirement. It is, and was, easy to construct finitely additive measures (i.e., where (2) holds with ∞ replaced by an arbitrary finite number), but a satisfactory theory of integration cannot be developed this way. Since $\mu(\varnothing) = 0$, equation (2) includes finite additivity as a special case. Three other important consequences of (2) are

$$\mu(A) \leq \mu(B) \qquad \text{if } A \subset B, \tag{3}$$

$$\lim_{j\to\infty} \mu(A_j) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \qquad \text{if } A_1 \subset A_2 \subset A_3 \subset \cdots, \tag{4}$$

$$\lim_{j\to\infty} \mu(A_j) = \mu\left(\bigcap_{i=1}^{\infty} A_i\right) \qquad \text{if } A_1 \supset A_2 \supset \cdots \text{ and } \mu(A_1) < \infty. \tag{5}$$

The reader can easily prove (3)–(5) using the properties of a sigma-algebra.

A **measure space** thus has three parts: A set Ω, a sigma-algebra Σ and a measure μ. If $\Omega = \mathbb{R}^n$ (or, more generally, if Ω has open subsets, so that $\mathcal{B}$ can be defined) and if $\Sigma = \mathcal{B}$, then μ is said to be a **Borel measure**. We often refer to the elements of Σ as the **measurable sets**. Note that whenever Ω' is a measurable subset of Ω we can always define the measure subspace (Ω', Σ', μ), in which Σ' consists of the measurable subsets of Ω'. This is called the **restriction of μ to Ω'**.

A simple and important example in $\mathbb{R}^n$ is the **Dirac delta-measure**, δ_y, located at some arbitrary, but fixed, point $y \in \mathbb{R}^n$:

$$\delta_y(A) = \begin{cases} 1 & \text{if } y \in A, \\ 0 & \text{if } y \notin A. \end{cases} \tag{6}$$

In other words, using the definition of characteristic functions in 1.1(1),

$$\delta_y(A) = \chi_A(y). \tag{7}$$

Here, the sigma-algebra can be taken to be $\mathcal{B}$ or it can be taken to be *all* subsets of $\mathbb{R}^n$.

The second, and for us most important, example is **Lebesgue measure** on $\mathbb{R}^n$. Its construction is not easy, but it has the property of correctly giving the Euclidean volume of 'nice' sets. We do not give the construction because it can be found in many, many books, e.g., [Rudin, 1987]. However, the determined reader will be invited to construct Lebesgue measure as Exercise 5 in Chapter 6, with the aid of Theorem 6.22 (positive distributions are positive measures). Σ is taken to be $\mathcal{B}$ and the measure (or volume) of a set $A \in \mathcal{B}$ is denoted by $\mathcal{L}^n(A)$ or by the symbol

$$|A| := \mathcal{L}^n(A).$$

The Lebesgue measure of a ball is

$$\mathcal{L}^n(B_{x,r}) = |B_{0,1}|r^n = \frac{2\pi^{n/2}r^n}{n\Gamma(n/2)} = \frac{1}{n}|\mathbb{S}^{n-1}|r^n, \tag{8}$$

where

$$|\mathbb{S}^{n-1}| = 2\pi^{n/2}/\Gamma(n/2)$$

is the area of $\mathbb{S}^{n-1}$, which is the sphere of radius 1 in $\mathbb{R}^n$.

This measure is translation invariant—meaning that for every fixed $y \in \mathbb{R}^n$, $\mathcal{L}^n(A) = \mathcal{L}^n(\{x + y : x \in A\})$. Up to an over-all constant *it is the only translation invariant measure on* $\mathbb{R}^n$. The fact that the classical measure (8) can be extended in a *countably additive way* to a *sigma-algebra* containing all balls is a triumph which, having been achieved, makes integration theory relatively painless.

A small annoyance is connected with sets of measure zero, and is caused by the fact that a subset of a set of measure zero might not be measurable. An example is produced in the following fashion: Take a line ℓ in the plane $\mathbb{R}^2$. This set is a Borel set and $\mathcal{L}^2(\ell) = 0$. Now take any subset $\gamma \subset \ell$ that is not a Borel set in the one-dimensional sense. One can show that γ is also not a Borel set in the two-dimensional sense and therefore it is meaningless to say that $\mathcal{L}^2(\gamma) = 0$. One can get around this difficulty by declaring all subsets of sets of zero measure to be measurable and to have zero measure. But then, for consistency, these new sets have to be added to, and subtracted from, the Borel sets in $\mathcal{B}$. In this way Lebesgue measure can be extended to a larger class than $\mathcal{B}$, and it is easy to see that this class forms a sigma-algebra (Exercise 10). While this extension (called the **completion**) has its merits, we shall not use it in this book for it has no real value for us and causes problems, notably that the intersection of a measurable set in $\mathbb{R}^n$ with a hyperplane may not be measurable. For us, $\mathcal{L}^n$ is defined only on $\mathcal{B}$.

There is, however, one way in which subsets of sets of zero measure play a role. Given (Ω, Σ, μ) we say that some property holds **μ-almost everywhere** (or μ-a.e., or simply a.e. if μ is understood) whenever the subset of Ω for which the property fails to hold is a subset of a set of measure zero.

Lebesgue measure has two important properties called **inner regularity** and **outer regularity**. (See Theorem 6.22 and Exercise 6.5.) For every Borel set A

$$\mathcal{L}^n(A) = \inf\{\mathcal{L}^n(O) : A \subset O \text{ and } O \text{ is open}\} \qquad \textit{outer regularity,} \quad (9)$$

$$\mathcal{L}^n(A) = \sup\{\mathcal{L}^n(C) : C \subset A \text{ and } C \text{ is compact}\} \quad \textit{inner regularity.} \quad (10)$$

The reader will be asked to prove equations (9) and (10) in Exercise 26, with the help of Theorem 1.3 (Monotone class theorem) and ideas similar to those used in the proof of Theorem 1.18.

Another important property of Lebesgue measure is its **sigma-finiteness**. A measure space (Ω, Σ, μ) is sigma-finite if there are countably many sets $A_1, A_2, \dots$ such that $\mu(A_i) < \infty$ for all $i = 1, 2, \dots$ and such that $\Omega = \bigcup_{i=1}^{\infty} A_i$. If sigma-finiteness holds it is easy to prove that the A_i's can be taken to be disjoint. In the case of $\mathcal{L}^n$ we can, for instance, take the A_i's to be cubes of unit edge length.

As a final topic in this section we explain **product sigma-algebras** and **product measures**. Given two spaces Ω_1, Ω_2 with sigma-algebras Σ_1 and Σ_2 we can form the **product space**

$$\Omega = \Omega_1 \times \Omega_2 = \{(x_1, x_2) : x_1 \in \Omega_1, \quad x_2 \in \Omega_2\}.$$

A good example is to think of Ω_1 as $\mathbb{R}^m$ and Ω_2 as $\mathbb{R}^n$ and $\Omega = \mathbb{R}^{m+n}$. The product sigma-algebra $\Sigma = \Sigma_1 \times \Sigma_2$ of sets in Ω is defined by first declaring all **rectangles** to be members of Σ. A rectangle is a set of the form

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

where A_1 and A_2 are members of Σ_1 and Σ_2. Then $\Sigma = \Sigma_1 \times \Sigma_2$ is *defined* to be the *smallest* sigma-algebra containing all these rectangles, i.e., the sigma-algebra generated by all these rectangles. We shall see that the fact that Σ is the *smallest* sigma-algebra is important for Fubini's theorem (see Sects. 1.10 and 1.12).

Next suppose that $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are two measure spaces. It is a basic and nontrivial fact that there exists a unique measure μ on the product sigma-algebra Σ of Ω with the 'product property' that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

for all rectangles. This measure μ is called the *product measure* and is denoted by $\mu_1 \times \mu_2$. It will be constructed in Theorem 1.10 (product measure). The sigma-algebra Σ has the **section property** that if we take an arbitrary $A \in \Sigma$ and form the set $A_1(x_2) \subset \Omega_1$ defined by $A_1(x_2) = \{x_1 \in \Omega_1 : (x_1, x_2) \in A\}$, then $A_1(x_2)$ is in Σ_1 for *every* choice of x_2. An analogous property holds with 1 and 2 interchanged.

The section property depends crucially on the fact that Σ is defined to be the *smallest sigma-algebra that contains all rectangles.* To prove the section property one reasons as follows. Let $\Sigma' \subset \Sigma$ be the set of all those measurable sets $A \in \Sigma$ that *do have* the section property. Certainly, $\varnothing$ is in Σ' and $\Omega_1 \times \Omega_2$ is also in Σ'. Moreover, all rectangles are in Σ'. From the identity

$$\left(\bigcup_i A^i\right)_2(x_1) = \left(\bigcup_i A_2^i(x_1)\right)$$

which holds for any family of sets it follows that countable unions of sets in Σ' also have the section property. And from $A_2^c(x_1) = (A_2(x_1))^c$ one infers that if $A \in \Sigma'$, then $A^c \in \Sigma'$. Hence $\Sigma' \subset \Sigma$ is a sigma-algebra and since it contains all the rectangles it must be equal to the minimal sigma-algebra Σ. This way of reasoning will be used again in the proof of Theorem 1.10.

In the same fashion one easily proves that for any three sigma-algebras $\Sigma_1, \Sigma_2, \Sigma_3$ the smallest sigma-algebra $\Sigma = \Sigma_1 \times \Sigma_2 \times \Sigma_3$ that contains all cubes also has the section property, i.e., for $A \in \Sigma$,

$$A_{23}(x_1) = \{(x_2, x_3) : (x_1, x_2, x_3) \in A\} \in \Sigma_2 \times \Sigma_3$$

for every $x_1 \in \Omega_1$, etc. By cubes we understand sets of the form $A_1 \times A_2 \times A_3$ where $A_i \in \Sigma_i$, $i = 1, 2, 3$.

If we turn to Lebesgue measure, then we find that if $\mathcal{B}^m$ is the Borel sigma-algebra of $\mathbb{R}^m$ then $\mathcal{B}^m \times \mathcal{B}^n = \mathcal{B}^{m+n}$. Note, however, that if we first extend Lebesgue measure to the nonmeasurable sets contained in Borel sets of measure zero, as described above, then the section property does not hold. A counterexample was mentioned earlier, namely a nonmeasurable subset of the real line is, when viewed as a subset of the plane, a subset of a set of measure zero. This failure of the section property is our chief reason for restricting the Lebesgue measure to the Borel sigma-algebra. It also shows that the product of the completion of the Borel sigma-algebra with itself is not complete; if it were complete it would contain the set mentioned above, but then it would fail to have the section property which, as we proved above, the product always has. On the other hand, if we take the completion of the product, then the section property can be shown to hold for *almost* every section.

● Up to now we have avoided proving any difficult theorems in measure theory. The following Theorem 1.3, however, is central to the subject and will be needed later in Sect. 1.10 on the product measure and for the proof of Fubini's theorem in 1.12. Because of its importance, and as an example of a 'pure measure theory' proof, we give it in some detail. The proof, but not the content, of Theorem 1.3 can be skipped on a first reading.

A **monotone class** $\mathcal{M}$ is a collection of sets with two properties:

if $A_i \in \mathcal{M}$ for $i = 1, 2, \dots$, and if $A_1 \subset A_2 \subset \cdots$, then $\bigcup_i A_i \in \mathcal{M}$;

if $B_i \in \mathcal{M}$ for $i = 1, 2, \dots$, and if $B_1 \supset B_2 \supset \cdots$, then $\bigcap_i B_i \in \mathcal{M}$.

Obviously any sigma-algebra is a monotone class, and the collection of all subsets of a set Ω is again a monotone class. Thus any collection of subsets is *contained* in a monotone class.

A collection of sets, $\mathcal{A}$, is said to form an **algebra of sets** if for every A and B in $\mathcal{A}$ the differences $A \sim B$, $B \sim A$ and the union $A \cup B$ are in $\mathcal{A}$. A sigma-algebra is then an algebra that is closed under countably many operations of this kind. Note that passage from an algebra, $\mathcal{A}$, to a sigma-algebra amounts to incorporation of countable unions of subsets of $\mathcal{A}$, thereby yielding some collection of sets, $\mathcal{A}_1$, which is no longer closed under taking intersections. Next, we incorporate countable intersections of sets in $\mathcal{A}_1$. This yields a collection of sets $\mathcal{A}_2$ which is not closed under taking unions. Proceeding this way one can arrive at a sigma-algebra by 'transfinite induction', which is enough to cause goose-bumps. The following theorem avoids this and simply states that sigma-algebras are monotone 'limits' of algebras. The key word in the following is 'sigma-algebra'.

1.3 THEOREM (Monotone class theorem)

Let Ω be a set and let $\mathcal{A}$ be an algebra of subsets of Ω such that Ω is in $\mathcal{A}$ and the empty set $\varnothing$ is also in $\mathcal{A}$. Then there exists a smallest monotone class $\mathcal{S}$ that contains $\mathcal{A}$. That class, $\mathcal{S}$, is also the smallest sigma-algebra that contains $\mathcal{A}$.

PROOF. Let $\mathcal{S}$ be the intersection of all monotone classes that contain $\mathcal{A}$, i.e., $Y \in \mathcal{S}$ if and only if Y is in every monotone class containing $\mathcal{A}$. We leave it as an exercise to the reader to show that $\mathcal{S}$ is a monotone class containing $\mathcal{A}$. By definition, it is then the smallest such monotone class.

We first note that it suffices to show that $\mathcal{S}$ is closed under forming complements and finite unions. Assuming this closure for the moment, we have, with $A_1, A_2, \dots$ in $\mathcal{S}$, that $B_n := \bigcup_{i=1}^n A_i$ is a monotone increasing sequence of sets in $\mathcal{S}$. Since $\mathcal{S}$ is a monotone class $\bigcup_{i=1}^\infty A_i$ is in $\mathcal{S}$. Thus $\mathcal{S}$

is necessarily closed under forming countable unions. The formula

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \left(\bigcup_{i=1}^{\infty} A_i^c\right)$$

implies that $\mathcal{S}$, being closed under forming complements, contains also countable intersections of its members. Thus $\mathcal{S}$ is a sigma-algebra and since any sigma-algebra is a monotone class, $\mathcal{S}$ is the smallest sigma-algebra that contains $\mathcal{A}$.

Next, we show that $\mathcal{S}$ is indeed closed under finite unions. Fix a set $A \in \mathcal{A}$ and consider the collection $\mathcal{C}(A) = \{B \in \mathcal{S} : B \cup A \in \mathcal{S}\}$. Since $\mathcal{A}$ is an algebra, $\mathcal{C}(A)$ contains $\mathcal{A}$. For any increasing sequence of sets B_n in $\mathcal{C}(A)$, $A \cup B_i$ is an increasing sequence of sets in $\mathcal{S}$. Since $\mathcal{S}$ is a monotone class,

$$A \cup \left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} A \cup B_i$$

is in $\mathcal{S}$ and therefore $\bigcup_{i=1}^{\infty} B_i$ is in $\mathcal{C}(A)$. The reader can show that $\mathcal{C}(A)$ is closed under countable intersections of decreasing sets, and we then conclude that $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$. Since $\mathcal{C}(A) \subset \mathcal{S}$ and $\mathcal{S}$ is the smallest monotone class that contains $\mathcal{A}$, $\mathcal{C}(A) = \mathcal{S}$.

Again, fix a set A, but this time an arbitrary one in $\mathcal{S}$, and consider the collection $\mathcal{C}(A) = \{B \in \mathcal{S} : B \cup A \in \mathcal{S}\}$. From the previous argument we know that $\mathcal{A}$ is a subset of $\mathcal{C}(A)$. A verbatim repetition of that argument to this new collection $\mathcal{C}(A)$ will convince the reader that $\mathcal{C}(A)$ is a monotone class and hence $\mathcal{C}(A) = \mathcal{S}$. Thus $\mathcal{S}$ is closed under finite unions, as claimed.

Finally, we address the complementation question. Let $\mathcal{C} = \{B \in \mathcal{S} : B^c \in \mathcal{S}\}$. This set contains $\mathcal{A}$ since $\mathcal{A}$ is an algebra. For any increasing sequence of sets $B_i \in \mathcal{C}$, $i = 1, 2, \ldots$, B_i^c is a decreasing sequence of sets in $\mathcal{S}$. Since $\mathcal{S}$ is a monotone class,

$$\left(\bigcup_{i=1}^{\infty} B_i\right)^c = \bigcap_{i=1}^{\infty} B_i^c$$

is in $\mathcal{S}$. Similarly for any decreasing sequence of sets $B_i \in \mathcal{C}$, $i = 1, 2, \ldots$, B_i^c is an increasing sequence of sets in $\mathcal{S}$ and hence

$$\left(\bigcap_{i=1}^{\infty} B_i\right)^c = \bigcup_{i=1}^{\infty} B_i^c$$

is in $\mathcal{S}$. Again $\mathcal{C} = \mathcal{S}$.

Thus $\mathcal{S}$ is closed under finite intersections and complementation. ∎

As an application of the monotone class theorem we present a uniqueness theorem for measures. It demonstrates a typical way of using the monotone class theorem and it will be handy in Sect. 1.10 on product measures.

1.4 THEOREM (Uniqueness of measures)

Let Ω be a set, $\mathcal{A}$ an algebra of subsets of Ω and Σ the smallest sigma-algebra that contains $\mathcal{A}$. Let μ_1 be a sigma-finite measure in the stronger sense that there exists a sequence of sets $A_i \in \mathcal{A}$ (and not merely $A_i \in \Sigma$), $i = 1, 2, \ldots$, each having finite μ_1 measure, such that $\bigcup_{i=1}^{\infty} A_i = \Omega$. If μ_2 is a measure that coincides with μ_1 on $\mathcal{A}$, then $\mu_1 = \mu_2$ on all of Σ.

PROOF. First we prove the theorem under the assumption that μ_1 is a finite measure on Ω. Consider the set

$$\mathcal{M} = \{A \in \Sigma : \mu_1(A) = \mu_2(A)\}.$$

Clearly this collection of sets contains $\mathcal{A}$ and we shall show that $\mathcal{M}$ is a monotone class. By the previous Theorem 1.3 we then conclude that $\mathcal{M} = \Sigma$. Let $A_1 \subset A_2 \subset \cdots$ be an increasing sequence of sets in $\mathcal{M}$. Define $B_1 = A_1, B_2 = A_2 \sim A_1, \ldots, B_n = A_n \sim A_{n-1}, \ldots$. These sets are mutually disjoint and $\bigcup_{i=1}^{n} B_i = A_n$, in particular

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

By the countable additivity of measures,

$$\begin{aligned}\mu_1\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} \mu_1(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu_1(B_i) \\ &= \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2\left(\bigcup_{i=1}^{\infty} A_i\right).\end{aligned}$$

Hence $\bigcup_{i=1}^{\infty} A_i$ is in $\mathcal{M}$. Now, with $A \in \mathcal{M}$, its complement A^c is also in $\mathcal{M}$, which follows from the fact that $\mu_i(A^c) = \mu_i(\Omega) - \mu_i(A)$, $i = 1, 2$, and that $\mu_1(\Omega) = \mu_2(\Omega) < \infty$. From this, it is easy to show that $\mathcal{M}$ is a monotone class. We leave the details to the reader.

Next, we return to the sigma-finite case. The theorem for the finite case implies that $\mu_1(B \cap A_0) = \mu_2(B \cap A_0)$ for every $A_0 \in \mathcal{A}$ with $\mu(A_0) < \infty$ and every $B \in \Sigma$. To see this, simply note that $A_0 \cap \Sigma$ is a sigma-algebra on A_0

which is the smallest one that contains the algebra $A_0 \cap \mathcal{A}$. (Why?) Recall that, by assumption, there exists a sequence of sets $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$, each having finite μ_1 measure, such that $\bigcup_{i=1}^{\infty} A_i = \Omega$. Without loss of generality we may assume that these sets are disjoint. (Why?) Now for $B \in \Sigma$,

$$\mu_1(B) = \mu_1\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right) = \sum_{i=1}^{\infty} \mu_1(A_i \cap B) = \sum_{i=1}^{\infty} \mu_2(A_i \cap B) = \mu_2(B). \quad \blacksquare$$

1.5 DEFINITION OF MEASURABLE FUNCTIONS AND INTEGRALS

Suppose that $f : \Omega \to \mathbb{R}$ is a real-valued function on Ω. Given a sigma-algebra Σ, we say that f is a **measurable function** (with respect to Σ) if for every number t the **level set**

$$S_f(t) := \{x \in \Omega : f(x) > t\} \tag{1}$$

is measurable, i.e., $S_f(t) \in \Sigma$. The phrase f is Σ-measurable or, with an abuse of terminology, f is μ-measurable (in case there is a measure μ on Σ) is often used to denote measurability. Note, however, that measurability does not require a measure!

More generally, if $f : \Omega \to \mathbb{C}$ is complex-valued, we say that f is measurable if its real and imaginary parts, $\operatorname{Re} f$ and $\operatorname{Im} f$, are measurable.

REMARK. Instead of the $>$ sign in (1) we could have chosen $\geq$, $\leq$ or $<$. All these definitions are in fact equivalent. To see this, one notes, for example, that

$$\{x \in \Omega : f(x) > t\} = \bigcup_{j=1}^{\infty} \{x \in \Omega : f(x) \geq t + 1/j\}.$$

If Σ is the Borel sigma-algebra $\mathcal{B}$ on $\mathbb{R}^n$, it is evident that every continuous function is Borel measurable, in fact $S_f(t)$ is then open. Other examples of Borel measurable functions are upper and lower semicontinuous functions. Recall that a real-valued function f is **lower semicontinuous** if $S_f(t)$ is open and it is **upper semicontinuous** if $\{x \in \Omega : f(x) < t\}$ is open. f is continuous if it is both upper and lower semicontinuous. To prove measurability when f is upper semicontinuous, note that the set $\{x : f(x) < t + 1/j\}$ is measurable. Since

$$\{x \in \Omega : f(x) \leq t\} = \bigcap_{j=1}^{\infty} \{x : f(x) < t + 1/j\},$$

the set $\{x : f(x) \leq t\}$ is measurable. Therefore

$$S_f(t) = \Omega \sim \{x : f(x) \leq t\}$$

is also measurable.

By pursuing the above reasoning a little further, one can show that for *any* Borel set $A \subset \mathbb{R}$ the set $\{x : f(x) \in A\}$ is Σ-measurable whenever f is Σ-measurable.

An amusing exercise (see Exercises 3, 4, 18) is to prove the facts that whenever f and g are measurable functions then so are the functions $x \mapsto \lambda f(x)+\gamma g(x)$ for λ and $\gamma \in \mathbb{C}$, $x \mapsto f(x)g(x)$, $x \mapsto |f(x)|$ and $x \mapsto \phi(f(x))$, where ϕ is any Borel measurable function from $\mathbb{C}$ to $\mathbb{C}$. In the same vein $x \mapsto \max\{f(x), g(x)\}$ and $x \mapsto \min\{f(x), g(x)\}$ are measurable functions. Moreover, when $f^1, f^2, f^3, \ldots$ is a sequence of measurable functions then the functions $\limsup_{j\to\infty} f^j(x)$ and $\liminf_{j\to\infty} f^j(x)$ are measurable.

Hence, *if a sequence $f^j(x)$ has a limit $f(x)$ for μ-almost every x, then f is a measurable function.* (More precisely, f can be redefined on a set of measure zero so that it becomes measurable.) The reader is urged to prove all these assertions or at least look them up in any standard text.

That a measurable function is defined only almost everywhere can cause some difficulties with some concepts, e.g., with the notion of strict positivity of a function. To remedy this we say that a nonnegative measurable function f is a **strictly positive measurable function** on a measurable set A, if the set $\{x \in A : f(x) = 0\}$ has zero measure.

Similar difficulties arise in the definition of the support of a measurable function. For a given Borel measure μ let f be a Borel measurable function on $\mathbb{R}^n$, or on any topological space for that matter. Recall that the open sets are measurable, i.e., they are members of the sigma-algebra. Consider the collection Ω of *open* subsets ω with the property that $f(x) = 0$ for μ-almost every $x \in \omega$ and let the open set ω^* be the union of all the ω's in Ω. Note that Ω and ω^* might be empty. Now we define the **essential support of** f, $\operatorname{ess\,supp}\{f\}$, to be the complement of ω^*. Thus, $\operatorname{ess\,supp}\{f\}$ is a closed, and hence measurable, set. Consider, e.g., the function f on $\mathbb{R}$, defined by $f(x) = 1$, x rational, and $f(x) = 0$, x not rational, and with μ being Lebesgue measure. Obviously $f(x) = 0$ for a.e. $x \in \mathbb{R}$, and hence $\operatorname{ess\,supp}\{f\} = \varnothing$. Note also that $\operatorname{ess\,supp}\{f\}$ depends on the measure μ and not just on the sigma-algebra. It is a simple exercise to verify that for μ being Lebesgue measure and f continuous, $\operatorname{ess\,supp}\{f\}$ coincides with $\operatorname{supp}\{f\}$, defined in Sect. 1.1.

In the remainder of this book we shall, for simplicity, use $\operatorname{supp}\{f\}$ to mean $\operatorname{ess\,supp}\{f\}$.

Our next task is to use a measure μ to define integrals of measurable functions. (Recall that the concept of measurability has nothing to do with a measure.)

First, suppose that $f : \Omega \to \mathbb{R}^+$ is a nonnegative real-valued, Σ-measurable function on Ω. (Our notation throughout will be that $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$.) We then define

$$F_f(t) = \mu(S_f(t)),$$

i.e., $F_f(t)$ is the measure of the set on which $f > t$. Evidently $F_f(t)$ is a nonincreasing function of t since $S_f(t_1) \subset S_f(t_2)$ for $t_1 \geq t_2$. Thus $F_f(t) : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotone nonincreasing function and it is an elementary calculus exercise (and a fundamental part of the theory of Riemann integration) to verify that the Riemann integral of such functions is always well defined (although its value might be $+\infty$). This Riemann integral *defines* **the integral of f over Ω**, i.e.,

$$\int_\Omega f(x)\mu(\mathrm{d}x) := \int_0^\infty F_f(t)\,\mathrm{d}t. \tag{2}$$

(Notation: sometimes we abbreviate this integral as $\int f$ or $\int f\,\mathrm{d}\mu$. The symbol $\mu(\mathrm{d}x)$ is intended to display the underlying measure, μ. Some authors use $\mathrm{d}\mu(x)$ while others use just $\mathrm{d}\mu x$. When μ is Lebesgue measure, $\mathrm{d}x$ is used in place of $\mathcal{L}^n(\mathrm{d}x)$.) A heuristic verification of the reason that (2) agrees with the usual definition can be given by introducing Heaviside's step-function $\Theta(s) = 1$ if $s > 0$ and $\Theta(s) = 0$ otherwise. Then, *formally*,

$$\begin{aligned} \int_0^\infty F_f(t)\,\mathrm{d}t &= \int_0^\infty \left\{ \int_\Omega \Theta(f(x) - t)\mu(\mathrm{d}x) \right\} \mathrm{d}t \\ &= \int_\Omega \left\{ \int_0^{f(x)} \mathrm{d}t \right\} \mu(\mathrm{d}x) = \int_\Omega f(x)\mu(\mathrm{d}x). \end{aligned} \tag{3}$$

If f is measurable and nonnegative and if $\int f\,\mathrm{d}\mu < \infty$, we say that f is a **summable** (or **integrable**) **function**.

It is an important fact (which we shall not need, and therefore not prove here) that if the function f is **Riemann integrable**, then its **Riemann integral** coincides with the value given in (2). See, however, Exercise 21 for a special case which will be used in Chapter 6.

More generally, suppose $f : \Omega \to \mathbb{C}$ is a complex-valued function on Ω. Then f consists of two real-valued functions, because we can write $f(x) = g(x) + ih(x)$, with g and h real-valued. In turn, each of these two functions can be thought of as the difference of two nonnegative functions, e.g.,

$$g(x) = g_+(x) - g_-(x) \quad \text{where} \tag{4}$$

$$g_+(x) = \begin{cases} g(x) & \text{if } g(x) > 0, \\ 0 & \text{if } g(x) \leq 0. \end{cases} \tag{5}$$

Alternatively, $g_+(x) = \max(g(x), 0)$ and $g_-(x) = -\min(g(x), 0)$. These are called the **positive and negative parts** of g. If f is measurable, then all four functions are measurable by the earlier remark. If all four functions g_+, g_-, h_+, h_- are summable, we say that f is summable and we define

$$\int f := \int g_+ - \int g_- + i \int h_+ - i \int h_-. \tag{6}$$

Equivalently, f is summable if and only if $x \mapsto |f(x)| \in \mathbb{R}^+$ is a summable function. It is to be emphasized that the integral of f can be defined *only* if f is summable. To attempt to integrate a function that is not summable is to open a Pandora's box of possibly false conclusions and paradoxes. There is, however, a noteworthy exception to this rule: If f is nonnegative we shall often abuse notation slightly by writing $\int f = +\infty$ when f is not summable. With this convention a relation such as $\int g < \int f$ (for $f \geq 0$ and $g \geq 0$) is meant to imply that when g is not summable, then f is also not summable. This convention saves some pedantic verbiage.

Another amusing (and not so trivial) exercise (see Exercise 9) is the verification of the linearity of integration. If f and g are summable, then $\lambda f + \gamma g$ are summable (for any λ and $\gamma \in \mathbb{C}$) and

$$\int_\Omega (\lambda f + \gamma g)\, \mathrm{d}\mu = \lambda \int_\Omega f\, \mathrm{d}\mu + \gamma \int_\Omega g\, \mathrm{d}\mu. \tag{7}$$

The difficulty here lies in computing the level sets of linear combinations of summable functions.

An important class of measurable functions consists of the characteristic functions of measurable sets, as defined in 1.1(1). Clearly,

$$\int_\Omega \chi_A\, \mathrm{d}\mu = \mu(A)$$

and hence χ_A is summable if and only if $\mu(A) < \infty$.

Sometimes we shall use the notation $\chi_{\{\cdots\}}$, where $\{\cdots\}$ denotes a set that is specified by condition $\cdots$. For example, if f is a measurable function, $\chi_{\{f>t\}}$ is the characteristic function of the set $S_f(t)$, whence $\int \chi_{\{f>t\}}$ is precisely $F_f(t)$ for $t \geq 0$.

For later use we now show that $\chi_{\{f>t\}}$ is a *jointly* measurable function of x and t. We have to show that the level sets of $\chi_{\{f>t\}}$ are $\Sigma \times \mathcal{B}^1$-measurable, where $\mathcal{B}^1$ is the Borel sigma-algebra on the half line $\mathbb{R}^+$. The level sets in (x, t)-space are parametrized by $s \geq 0$ and have the form

$$\{(x, t) \in \Omega \times \mathbb{R}^+ : \chi_{\{f>t\}}(x) > s\}.$$

If $s \geq 1$, then the level set is empty and hence measurable. For $0 \leq s < 1$ the level set does not depend on s since $\chi_{\{f>t\}}$ takes only the values zero or one. In fact it is the set 'under the graph of f', i.e., the set $G = \{(x,t) \in \Omega \times \mathbb{R}^+ : 0 \leq t < f(x)\}$. This set is the union of sets of the form $S_f(r) \times [0, r]$ for rational r. (Recall that $[a, b]$ denotes the closed interval $a \leq x \leq b$ while (a, b) denotes the open interval $a < x < b$.) Since the rationals are countable we see that G is the countable union of rectangles and hence is measurable. Another way to prove that $G \subset \mathbb{R}^{n+1}$ is measurable, but which is secretly the same as the previous proof, is to note that

$$G = \{(x, t) : f(x) - t \geq 0\} \cap \{t : t > 0\},$$

and this is a measurable set since the set on which a measurable function ($f(x) - t$, in this case) is nonnegative is measurable by definition. (Why is $f(x) - t$ $\mathcal{L}^{n+1}$-measurable?)

Our definition of the integral suggests that it should be interpreted as the '$\mu \times \mathcal{L}^1$' measure of the set G which is in $\Sigma \times \mathcal{B}^1$. It is reasonable to *define*

$$(\mu \times \mathcal{L}^1)(G) := \int_0^\infty \int_\Omega \chi_{\{f>a\}}(x)\mu(\mathrm{d}x)\ \mathrm{d}a = \int_\Omega f(x)\mu(\mathrm{d}x). \tag{8}$$

A necessary condition for this to be a good definition is that it should not matter whether we integrate first over a or over x. In fact, since for every $x \in \Omega$, $\int_0^\infty \chi_{\{f>a\}}(x)\,\mathrm{d}a = f(x)$ (even for nonmeasurable functions), we have (recalling the definition of the integral) that

$$\int_0^\infty \int_\Omega \chi_{\{f>a\}}(x)\mu(\mathrm{d}x)\,\mathrm{d}a = \int_\Omega \int_0^\infty \chi_{\{f>a\}}(x)\,\mathrm{d}a\mu(\mathrm{d}x). \tag{9}$$

This is a first elementary instance of **Fubini's theorem** about the interchange of integration. We shall see later in Theorem 1.10 that this interchange of integration is valid for *any* set $A \in \Sigma \times \mathcal{B}^1$ and we shall *define* $(\mu \times \mathcal{L}^1)(A)$ to be $\int_{\mathbb{R}} \mu(\{x : (x, a) \in A\})\,\mathrm{d}a$. We shall also see that $\mu \times \mathcal{L}^1$ defined this way is a measure on $\Sigma \times \mathcal{B}^1$.

● With this brief sketch of the fundamentals behind us, we are now ready to prove one of the basic convergence theorems in the subject. It is due to Levi and Lebesgue. (Here and in the following the measure space (Ω, Σ, μ) will be understood.)

Suppose that $f^1, f^2, f^3, \ldots$ is an increasing sequence of summable functions on (Ω, Σ, μ), i.e., for each j, $f^{j+1}(x) \geq f^j(x)$ for μ-almost every $x \in \Omega$. Because a countable union of sets of measure zero also has measure zero, it then follows that the sequence of numbers $f^1(x), f^2(x), \ldots$ is nondecreasing for almost every x. This monotonicity allows us to define

$$f(x) := \lim_{j\to\infty} f^j(x)$$

for almost every x, and we can define $f(x) := 0$ on the set of x's for which the above limit does not exist. This limit can, of course, be $+\infty$, but it is well defined a.e. It is also clear that the numbers $I_j := \int_\Omega f^j \,\mathrm{d}\mu$ are also nondecreasing and we can define

$$I := \lim_{j\to\infty} I_j.$$

1.6 THEOREM (Monotone convergence)

Let $f^1, f^2, f^3, \ldots$ be an increasing sequence of summable functions on (Ω, Σ, μ), with f and I as defined above. Then f is measurable and, moreover, I is finite if and only if f is summable, in which case $I = \int_\Omega f \,\mathrm{d}\mu$. In other words,

$$\lim_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) = \int_\Omega \lim_{j\to\infty} f^j(x)\mu(\mathrm{d}x), \tag{1}$$

with the understanding that the left side of (1) *is $+\infty$ when f is not summable.*

PROOF. We can assume that the f^j are nonnegative; otherwise, we can replace f^j by $f^j - f^1$ and use the summability of f^1. To compute $\int f^j$ we must first compute

$$F_{f^j}(t) = \mu(\{x : f^j(x) > t\}).$$

Note that, by definition, the set $\{x : f(x) > t\}$ equals the union of the increasing, countable family of sets $\{x : f^j(x) > t\}$. Hence, by 1.2(4), $\lim_{j\to\infty} F_{f^j}(t) = F_f(t)$ for every t. Moreover, this convergence is plainly monotone.

To prove our theorem, it then suffices to prove the corresponding theorem for the Riemann integral of monotone functions. That is,

$$\lim_{j\to\infty} \int_0^\infty F_{f^j}(t)\,\mathrm{d}t = \int_0^\infty F_f(t)\,\mathrm{d}t \tag{2}$$

given that each function $F_{f^j}(t)$ is monotone (in t), and the family is monotone in the index j, with the pointwise limit $F_f(t)$. This is an easy exercise; all that is needed is to note that the upper and lower Riemann sums converge. ■

● The previous theorem can be paraphrased as saying that the functional $f \mapsto \int f$ on nonnegative functions behaves like a continuous functional with respect to sequences that converge pointwise and *monotonically.* It is easy to see that $f \mapsto \int f$ is not continuous in general, i.e., if f^j is a sequence of positive functions and if $f^j \to f$ pointwise a.e. it is not true in general that $\lim_{j\to\infty} \int f^j = \int f$, or even that the limit exists (see the Remark after the next lemma). What is true, however, is that $f \mapsto \int f$ is pointwise lower semicontinuous, i.e., $\liminf_{j\to\infty} \int f^j \geq \int f$ if $f^j \to f$ pointwise (see Exercise 2). The precise enunciation of that fact is the lemma of Fatou.

1.7 LEMMA (Fatou's lemma)

Let $f^1, f^2, \ldots$ be a sequence of nonnegative, summable functions on (Ω, Σ, μ). Then $f(x) := \liminf_{j\to\infty} f^j(x)$ is measurable and

$$\liminf_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) \geq \int_\Omega f(x)\mu(\mathrm{d}x).$$

in the sense that the finiteness of the left side implies that f is summable.

► *Caution:* The word 'nonnegative' is crucial.

PROOF. Define $F^k(x) = \inf_{j\geq k} f^j(x)$. Since

$$\{x : F^k(x) \geq t\} = \bigcap_{j\geq k} \{x : f^j(x) \geq t\},$$

we see that $F^k(x)$ is measurable for all $k = 1, 2, \ldots$ by the Remark in 1.5. Moreover $F^k(x)$ is summable since $F^k(x) \leq f^k(x)$. The sequence F^k is obviously increasing and its limit is given by $\sup_{k\geq 1} \inf_{j\geq k} f^j(x)$ which is, by definition, $\liminf_{j\to\infty} f^j(x)$. We have that

$$\begin{aligned}\liminf_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) &:= \sup_{k\geq 1} \inf_{j\geq k} \int_\Omega f^j(x)\mu(\mathrm{d}x) \\ &\geq \lim_{k\to\infty} \int_\Omega F^k(x)\mu(\mathrm{d}x) = \int_\Omega f(x)\mu(\mathrm{d}x).\end{aligned}$$

The last equality holds by monotone convergence and shows that f is summable if the left side is finite. The first equality is a definition. The middle inequality comes from the general fact that $\inf_j \int h^j \geq \inf_j \int(\inf_j h^j) = \int(\inf_j h^j)$, since $(\inf_j h^j)$ does not depend on j. ■

REMARK. In case $f^j(x)$ converges to $f(x)$ for almost every $x \in \Omega$ the lemma says that

$$\liminf_j \int_\Omega f^j(x)\mu(\mathrm{d}x) \geq \int_\Omega f(x)\mu(\mathrm{d}x).$$

Even in this case the inequality can be strict. To give an example, consider on $\mathbb{R}$ the sequence of functions $f^j(x) = 1/j$ for $|x| \leq j$ and $f^j(x) = 0$ otherwise. Obviously $\int_\mathbb{R} f^j(x)\,\mathrm{d}x = 2$ for all j but $f^j(x) \to 0$ pointwise for all x.

● So far we have only considered the interchange of limits and integrals for nonnegative functions. The following theorem, again due to Lebesgue, is the one that is usually used for applications and takes care of this limitation. It is one of the most important theorems in analysis. It is equivalent to the monotone convergence theorem in the sense that each can be simply derived from the other.

1.8 THEOREM (Dominated convergence)

Let $f^1, f^2, \ldots$ be a sequence of complex-valued summable functions on (Ω, Σ, μ) and assume that these functions converge to a function f pointwise a.e. If there exists a summable, nonnegative function $G(x)$ on (Ω, Σ, μ) such that $|f^j(x)| \leq G(x)$ for all $j = 1, 2, \ldots$, then $|f(x)| \leq G(x)$ and

$$\lim_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) = \int_\Omega f(x)\mu(\mathrm{d}x).$$

► *Caution*: The existence of the dominating G is crucial!

PROOF. It is obvious that the real and imaginary parts of f^j, R^j and I^j, satisfy the same assumptions as f^j itself. The same is true for the positive and negative parts of R^j and I^j. Thus it suffices to prove the theorem for nonnegative functions f^j and f. By Fatou's lemma

$$\liminf_{j\to\infty} \int_\Omega f^j \geq \int_\Omega f.$$

Again by Fatou's lemma

$$\liminf_{j\to\infty} \int_\Omega (G(x) - f^j(x))\mu(\mathrm{d}x) \geq \int_\Omega (G(x) - f(x))\mu(\mathrm{d}x),$$

since $G(x) - f^j(x) \geq 0$ for all j and all $x \in \Omega$. Summarizing these two inequalities we obtain

$$\liminf_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) \geq \int_\Omega f(x)\mu(\mathrm{d}x) \geq \limsup_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x),$$

which proves the theorem. ■

REMARK. The previous theorem allows a slight, but useful, generalization in which the dominating function $G(x)$ is replaced by a *sequence* $G^j(x)$ with the property that there exists a summable G such that

$$\int_\Omega |G(x) - G^j(x)|\mu(\mathrm{d}x) \to 0 \quad \text{as } j \to \infty$$

and such that $0 \leq |f^j(x)| \leq G^j(x)$. Again, if $f^j(x)$ converges pointwise a.e. to f the limit and the integral can be interchanged, i.e.,

$$\lim_{j\to\infty} \int_\Omega f^j(x)\mu(\mathrm{d}x) = \int_\Omega f(x)\mu(\mathrm{d}x).$$

To see this assume first that $f^j(x) \geq 0$ and note that

$$\int (G - f^j)_+ \to \int (G - f)_+ \quad \text{as } j \to \infty$$

since $(G - f^j)_+ \leq G$, using dominated convergence. Next observe that

$$\int (G - f^j)_- = \int (G - G^j + G^j - f^j)_- \leq \int (G - G^j)_-$$

since $G^j - f^j \geq 0$. See 1.5(5). The last integral however tends to zero as $j \to \infty$, by assumption. Thus we obtain

$$\lim_{j\to\infty} \int (G - f^j) = \int (G - f)_+ = \int (G - f)$$

since clearly $f(x) \leq G(x)$. The generalization in which f takes complex values is straightforward.

● Theorem 1.8 was proved using Fatou's lemma. It is interesting to note that Theorem 1.8 can be used, in turn, to prove the following generalization of Fatou's lemma. Suppose that f^j is a sequence of nonnegative functions that converges pointwise to a function f. As we have seen in the Remark after Lemma 1.7, limit and integral cannot be interchanged since, intuitively, the sequence f^j might 'leak out to infinity'. The next theorem taken from [Brézis-Lieb] makes this intuition precise and provides us with a correction term that changes Fatou's lemma from an inequality to an equality. While it is not going to be used in this book, it is of intrinsic interest as a theorem in measure theory and has been used effectively to solve some problems in the calculus of variations. We shall state a simple version of the theorem; the reader can consult the original paper for the general version in which, among other things, $f \mapsto |f|^p$ is replaced by a larger class of functions, $f \mapsto j(f)$.

1.9 THEOREM (Missing term in Fatou's lemma)

Let f^j be a sequence of complex-valued functions on a measure space that converges pointwise a.e. to a function f (which is measurable by the remarks in 1.5). *Assume, also, that the f^j's are uniformly p^{th} power summable for some fixed $0 < p < \infty$, i.e.,*

$$\int_\Omega |f^j(x)|^p \mu(\mathrm{d}x) < C \quad \textit{for } j = 1, 2, \ldots$$

and for some constant C. Then

$$\lim_{j\to\infty} \int_\Omega \big||f^j(x)|^p - |f^j(x) - f(x)|^p - |f(x)|^p\big| \mu(\mathrm{d}x) = 0. \qquad (1)$$

REMARKS. (1) By Fatou's lemma, $\int |f|^p \le C$.

(2) By applying the triangle inequality to (1) we can conclude that

$$\int |f^j|^p = \int |f|^p + \int |f - f^j|^p + o(1), \qquad (2)$$

where $o(1)$ indicates a quantity that vanishes as $j \to \infty$. Thus the correction term is $\int |f - f^j|^p$, which measures the 'leakage' of the sequence f^j. One obvious consequence of (2), for all $0 < p < \infty$, is that if $\int |f - f^j|^p \to 0$ and if $f^j \to f$ a.e., then

$$\int |f^j|^p \to \int |f|^p.$$

(In fact, this can be proved directly under the sole assumption that $\int |f - f^j|^p \to 0$. When $1 \le p < \infty$ this a trivial consequence of the triangle inequality in 2.4(2). When $0 < p < 1$ it follows from the elementary inequality $|a + b|^p \le |a|^p + |b|^p$ for all complex a and b.) Another consequence of (2), for all $0 < p < \infty$, is that if $\int |f^j|^p \to \int |f|^p$ and $f^j \to f$ a.e., then

$$\int |f - f^j|^p \to 0.$$

PROOF. Assume, for the moment, that the following family of inequalities, (3), is true: For any $\varepsilon > 0$ there is a constant C_ε such that for all numbers $a, b \in \mathbb{C}$

$$\big||a+b|^p - |b|^p\big| \le \varepsilon |b|^p + C_\varepsilon |a|^p. \tag{3}$$

Next, write $f^j = f + g^j$ so that $g^j \to 0$ pointwise a.e. by assumption. We claim that the quantity

$$G_\varepsilon^j = \left(\big||f+g^j|^p - |g^j|^p - |f|^p\big| - \varepsilon |g^j|^p\right)_+ \tag{4}$$

satisfies $\lim_{j\to\infty}\int G_\varepsilon^j = 0$. Here $(h)_+$ denotes as usual the positive part of a function h. To see this, note first that

$$\begin{aligned}&\big||f+g^j|^p - |g^j|^p - |f|^p\big| \\ &\qquad \le \big||f+g^j|^p - |g^j|^p\big| + |f|^p \le \varepsilon |g^j|^p + (1+C_\varepsilon)|f|^p\end{aligned}$$

and hence $G_\varepsilon^j \le (1+C_\varepsilon)|f|^p$. Moreover $G_\varepsilon^j \to 0$ pointwise a.e. and hence the claim follows by Theorem 1.8 (dominated convergence). Now

$$\int \big||f+g^j|^p - |g^j|^p - |f|^p\big| \le \varepsilon \int |g^j|^p + \int G_\varepsilon^j.$$

We have to show $\int |g^j|^p$ is uniformly bounded. Indeed,

$$\int |g^j|^p = \int |f - f^j|^p \le 2^p \int (|f|^p + |f^j|^p) \le 2^{p+1}C.$$

Therefore,

$$\limsup_{j\to\infty} \int \big||f+g^j|^p - |g^j|^p - |f|^p\big| \le \varepsilon D.$$

Since ε was arbitrary the theorem is proved.

It remains to prove (3). The function $t \mapsto |t|^p$ is convex if $p > 1$. Hence $|a+b|^p \le (|a|+|b|)^p \le (1-\lambda)^{1-p}|a|^p + \lambda^{1-p}|b|^p$ for any $0 < \lambda < 1$. The choice $\lambda = (1+\varepsilon)^{-1/(p-1)}$ yields (3) in the case where $p > 1$. If $0 < p \le 1$ we have the simple inequality $|a+b|^p - |b|^p \le |a|^p$ whose proof is left to the reader. ■

● With these convergence tools at our disposal we turn to the question of proving Fubini's theorem, 1.12. Our strategy to prove Fubini's theorem in full generality will be the following: First, we prove the 'easy' form in Theorem 1.10; this will imply 1.5(9). Then we use a small generalization of Theorem 1.10 to establish the general case in Theorem 1.12.

1.10 THEOREM (Product measure)

Let $(\Omega_1, \Sigma_1, \mu_1), (\Omega_2, \Sigma_2, \mu_2)$ be two sigma-finite measure spaces. Let A be a measurable set in $\Sigma_1 \times \Sigma_2$ and, for every $x \in \Omega_2$, set $f(x) := \mu_1(A_1(x))$ and, for every $y \in \Omega_1$, $g(y) := \mu_2(A_2(y))$. (Note that by the considerations at the end of Sect. 1.2 the sections are measurable and hence these quantities are defined). Then f is Σ_2-measurable, g is Σ_1-measurable and

$$(\mu_1 \times \mu_2)(A) := \int_{\Omega_2} f(x)\mu_2(\mathrm{d}x) = \int_{\Omega_1} g(y)\mu_1(\mathrm{d}y). \tag{1}$$

Moreover, $\mu_1 \times \mu_2$, the product of the measures μ_1 and μ_2, defined in (1), is a sigma-finite measure on $\Sigma_1 \times \Sigma_2$.

PROOF. The measurability of f and g parallels the proof of the section property in Sect. 1.2 and uses the Monotone Class Theorem; it is left to Exercise 22.

Consider any collection of disjoint sets A^i, $i = 1, 2, \dots$, in $\Sigma_1 \times \Sigma_2$. Clearly their sections $A_1^i(x)$, $i = 1, 2, \dots$, which are measurable (see Sect. 1.2), are also disjoint and hence

$$\mu_1\left(\left(\bigcup_{i=1}^{\infty} A^i\right)_1(x)\right) = \sum_{i=1}^{\infty} \mu_1(A_1^i(x))$$

The monotone convergence theorem then yields the countable additivity of $\mu_1 \times \mu_2$. Similarly, the second integral in (1) also defines a countably additive measure.

We now verify the assumptions of Theorem 1.4 (uniqueness of measures). Define $\mathcal{A}$ to be the set of finite unions of rectangles, with $\Omega_1 \times \Omega_2$ and the empty set included. It is easy to see that this set is an algebra since the difference of two sets in $\mathcal{A}$ can be written again as a union of rectangles. Simply use the identities

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$$

and

$$(A_1 \times B_1) \sim (A_2 \times B_2) = [(A_1 \sim A_2) \times B_1] \cup [(A_1 \cap A_2) \times (B_1 \sim B_2)].$$

By assumption there exists a collection of sets $A_i \subset \Omega_1$ with $\mu_1(A_i) < \infty$ for $i = 1, 2, \dots$ and with

$$\bigcup_{i=1}^{\infty} A_i = \Omega_1.$$

Similarly, there exists a collection $B_j \subset \Omega_2$ with $\mu_2(B_j) < \infty$ for $j = 1, 2, \ldots$ and with

$$\bigcup_{j=1}^{\infty} B_j = \Omega_2.$$

Clearly the collection of rectangles $A_i \times B_j$ is countable, covers $\Omega_1 \times \Omega_2$ and

$$(\mu_1 \times \mu_2)(A_i \times B_j) = \mu_1(A_i)\mu_2(B_j) < \infty.$$

Thus, the two measures defined by the two integrals in (1) are sigma-finite in the stronger sense of Theorem 1.4. Now, note that the two integrals in (1) coincide on $\mathcal{A}$. Since, by definition, $\Sigma_1 \times \Sigma_2$ is the smallest sigma-algebra that contains $\mathcal{A}$, Theorem 1.4 yields (1) on all of $\Sigma_1 \times \Sigma_2$. ■

● The following generalization of the previous theorem is useful and is an important step in proving Fubini's theorem.

1.11 COROLLARY (Commutativity and associativity of product measures)

Let $(\Omega_i, \Sigma_i, \mu_i)$ for $i = 1, 2, 3$ be sigma-finite measure spaces. For $A \in \Sigma_1 \times \Sigma_2$ define the reflected set

$$RA := \{(x, y) : (y, x) \in A\}.$$

This defines a one-to-one correspondence between $\Sigma_1 \times \Sigma_2$ and $\Sigma_2 \times \Sigma_1$. Then the formation of the product measure $\mu_1 \times \mu_2$ is commutative in the sense that

$$(\mu_2 \times \mu_1)(RA) = (\mu_1 \times \mu_2)(A)$$

for every $A \in \Sigma_1 \times \Sigma_2$. Moreover, the formation of product measures is associative, i.e.

$$(\mu_1 \times \mu_2) \times \mu_3 = \mu_1 \times (\mu_2 \times \mu_3). \tag{1}$$

PROOF. The proof of the commutativity is an obvious consequence of the previous theorem. To see the associativity, simply note that the sigma-algebras associated with $(\mu_1 \times \mu_2) \times \mu_3$ and $\mu_1 \times (\mu_2 \times \mu_3)$ are the smallest monotone classes that contain unions of cubes. Hence (1) follows, since the two measures coincide on cubes. ■

1.12 THEOREM (Fubini's theorem)

Consider two sigma-finite *measure spaces* $(\Omega_i, \Sigma_i, \mu_i)$, $i = 1, 2$, *and let* f *be a* $\Sigma_1 \times \Sigma_2$ *measurable function on* $\Omega_1 \times \Omega_2$. *If* $f \geq 0$, *then the following three integrals are equal (in the sense that all three can be infinite):*

$$\int_{\Omega_1 \times \Omega_2} f(x,y)(\mu_1 \times \mu_2)(\mathrm{d}x\,\mathrm{d}y), \tag{1}$$

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(x,y)\mu_2(\mathrm{d}y) \right) \mu_1(\mathrm{d}x), \tag{2}$$

$$\int_{\Omega_2} \left(\int_{\Omega_1} f(x,y)\mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y). \tag{3}$$

If f *is complex-valued, then the above holds if one assumes in addition that*

$$\int_{\Omega_1 \times \Omega_2} |f(x,y)|(\mu_1 \times \mu_2)(\mathrm{d}x\,\mathrm{d}y) < \infty. \tag{4}$$

REMARK. Sigma-finiteness is essential! In Exercise 19 we ask the reader to construct a counterexample.

PROOF. The second part of the statement follows from the first applied to the positive and negative parts of the $\operatorname{Re} f$ and $\operatorname{Im} f$ separately. As for (1), (2), (3), recall that by Theorem 1.10 (product measure) and the considerations at the end of Sect. 1.5 the value of the integral in (1) is given by

$$(\mu_1 \times \mu_2 \times \mathcal{L}^1)(G) \tag{5}$$

where $G = \{(x, y, t) \in \Omega_1 \times \Omega_2 \times \mathbb{R} : 0 \leq t < f(x, y)\}$, i.e., G is the set under the graph of f. Note that by the previous corollary the sequence of the factors in (5) is of no concern. Hence one can interpret (5) in three ways as

$$(\mathcal{L}^1 \times (\mu_1 \times \mu_2))(G), \qquad (\mu_1 \times (\mathcal{L}^1 \times \mu_2))(R_1 G)$$

and

$$(\mu_2 \times (\mathcal{L}^1 \times \mu_1))(R_2 G)$$

where R_1 and R_2 are the appropriate reflections. By the previous corollary these numbers are all equal and thus the theorem follows from the definitions

$$\int_{\Omega_1 \times \Omega_2} f(x,y)(\mu_1 \times \mu_2)(\mathrm{d}x\,\mathrm{d}y) = \int_0^\infty (\mu_1 \times \mu_2)(\chi_{f>t})\,\mathrm{d}t,$$

$$\int_{\Omega_1} \mu_1(\mathrm{d}x) \int_{\Omega_2} f(x,y)\mu(\mathrm{d}y) = \int_{\Omega_1} \mu_1(\mathrm{d}x) \int_0^\infty \mu_2(\chi_{f(x,\cdot)>t})\,\mathrm{d}t,$$

and similarly with μ_1 and μ_2 interchanged. ■

● The next theorem is an elementary illustration of the use of Fubini's theorem. It is also extremely useful in practice because it permits us, in many cases, to reduce a problem about an integral of a general function to a problem about the integration of characteristic functions, i.e., functions that take only the values 0 or 1.

1.13 THEOREM (Layer cake representation)

Let ν be a measure on the Borel sets of the positive real line $[0,\infty)$ such that

$$\phi(t) := \nu([0,t)) \tag{1}$$

is finite for every $t > 0$. (Note that $\phi(0) = 0$ and that ϕ, being monotone, is Borel measurable.) Now let (Ω, Σ, μ) be a measure space and f any nonnegative measurable function on Ω. Then

$$\int_\Omega \phi(f(x))\mu(\mathrm{d}x) = \int_0^\infty \mu(\{x : f(x) > t\})\nu(\mathrm{d}t). \tag{2}$$

In particular, by choosing $\nu(\mathrm{d}t) = pt^{p-1}\,\mathrm{d}t$ for $p > 0$, we have

$$\int_\Omega f(x)^p \mu(\mathrm{d}x) = p \int_0^\infty t^{p-1}\mu(\{x : f(x) > t\})\,\mathrm{d}t. \tag{3}$$

By choosing μ to be the Dirac measure at some point $x \in \mathbb{R}^n$ and $p = 1$ we have

$$f(x) = \int_0^\infty \chi_{\{f>t\}}(x)\,\mathrm{d}t. \tag{4}$$

REMARKS. (1) It is formula (4) that we call the layer cake representation of f. (Approximate the $\mathrm{d}t$ integral by a Riemann sum and the allusion will be obvious.)

(2) The theorem can easily be generalized to the case in which ν is replaced by the difference of two (positive) measures, i.e., $\nu = \nu_1 - \nu_2$. Such a difference is called a **signed measure**. The functions ϕ that can be written as in (1) with this ν are called functions of **bounded variation**. The additional assumption needed for the theorem is that for the given f, and each of the measures ν_1 and ν_2, one of the integrands in (2) is summable. As an example,

$$\int_\Omega \sin[f(x)]\mu(\mathrm{d}x) = \int_0^\infty (\cos t)\mu(\{x : f(x) > t\})\,\mathrm{d}t.$$

(3) In the case where $\phi(t) = t$, equation (2) is just the *definition* of the integral of f.

(4) Our proof uses Fubini's theorem, but the theorem can also be proved by appealing to the original definition of the integral and computing the μ-measure of the set $\{x : \phi(f(x)) > t\}$. This can be tedious (we leave this to the reader) in case ϕ is not strictly monotone.

PROOF. Recall that

$$\int_0^\infty \mu(\{x : f(x) > t\})\nu(\mathrm{d}t) = \int_0^\infty \int_\Omega \chi_{\{f>t\}}(x)\mu(\mathrm{d}x)\nu(\mathrm{d}t)$$

and that $\chi_{\{f>t\}}(x)$ is jointly measurable as discussed in Sect. 1.5. By applying Theorem 1.12 (Fubini's theorem) the right side equals

$$\int_\Omega \left(\int_0^\infty \chi_{\{f>t\}}(x)\nu(\mathrm{d}t) \right) \mu(\mathrm{d}x).$$

The result follows by observing that

$$\int_0^\infty \chi_{\{f>t\}}(x)\nu(\mathrm{d}t) = \int_0^{f(x)} \nu(\mathrm{d}t) = \phi(f(x)). \qquad \blacksquare$$

● Another application of the notion of level sets is the 'bathtub principle'. It solves a simple minimization problem — one that arises from time to time, but which sometimes appears confusing until the problem is viewed in the correct light (see, e.g., Sects. 12.2 and 12.8). The proof, which we leave to the reader, is an easy exercise in manipulating level sets.

1.14 THEOREM (Bathtub principle)

Let (Ω, Σ, μ) be a measure space and let f be a real-valued, measurable function on Ω such that $\mu(\{x : f(x) < t\})$ is finite for all $t \in \mathbb{R}$. Let the number $G > 0$ be given and define a class of measurable functions on Ω by

$$\mathcal{C} = \left\{ g : 0 \le g(x) \le 1 \text{ for all } x \text{ and } \int_\Omega g(x)\mu(\mathrm{d}x) = G \right\}.$$

Then the minimization problem

$$I = \inf_{g \in \mathcal{C}} \int_\Omega f(x) g(x) \mu(\mathrm{d}x) \tag{1}$$

is solved by

$$g(x) = \chi_{\{f<s\}}(x) + c\chi_{\{f=s\}}(x) \tag{2}$$

and

$$I = \int_{f<s} f(x)\mu(\mathrm{d}x) + cs\mu(\{x : f(x) = s\}), \tag{3}$$

where

$$s = \sup\{t : \mu(\{x : f(x) < t\}) \le G\}, \tag{4}$$

$$c\mu(\{x : f(x) = s\}) = G - \mu(\{x : f(x) < s\}). \tag{5}$$

The minimizer given in (2) is unique if $G = \mu(\{x : f(x) < s\})$ or if $G = \mu(\{x : f(x) \le s\})$.

In order to understand why this is like filling a bathtub (and also for the purpose of constructing a proof of Theorem 1.14) think of the graph of f as a bathtub, take μ to be Lebesgue measure, and think of filling this bathtub with a fluid whose density g is not allowed to be greater than 1, but whose total mass, G, is given.

● The following theorem can be skipped at first reading for it will not be needed until Chapter 6 in the proof of Theorem 6.22 (positive distributions are measures). It provides a tool for constructing measures. Usually one is given a 'measure' on some collection of sets that is only finitely additive. The first step is to extend this 'measure' to an outer measure (defined by (i), (ii) and (iii) in Theorem 1.15 below) on *all* subsets. (Note: an outer measure is not necessarily finitely additive.) The second step is to restrict this outer measure to a class of sets that form a sigma-algebra in such a way that it is countably additive there. This construction is very general and the idea is due to Carathéodory.

1.15 THEOREM (Constructing a measure from an outer measure)

Let Ω be a set and let μ be an **outer measure** *on the collection of subsets of Ω, i.e., a nonnegative set function satisfying*

(i) $\mu(\varnothing) = 0$,

(ii) $\mu(A) \le \mu(B)$ *if* $A \subset B$,

(iii)
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu(A_i)$$
for any countable collection of subsets of Ω.

Define Σ to be the collection of sets satisfying **Carathéodory's criterion**, *namely $A \in \Sigma$ if*

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) \tag{1}$$

for every *set $E \subset \Omega$.*

Then Σ is a sigma-algebra and the restriction of μ to Σ is a countably additive measure. The sets in Σ are called the measurable sets.

PROOF. Clearly Σ is not empty since $\varnothing \in \Sigma$ and $\Omega \in \Sigma$. Obviously with $A \in \Sigma, A^c \in \Sigma$. It is an instructive exercise for the reader to show that any finite union and any finite intersection of measurable sets is measurable (see Exercise 8). Thus Σ is an algebra.

We show next that μ is a *finitely* additive measure on Σ. Let E be any set in Ω and let $B_1, B_2, \dots, B_m$ be a collection of disjoint measurable sets. Then

$$\begin{aligned} \mu(E) &= \mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)\right) + \mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)^c\right) \\ &\le \sum_{i=1}^{m} \mu(E \cap B_i) + \mu\left(E \cap \left(\bigcap_{i=1}^{m} B_i^c\right)\right). \end{aligned} \tag{2}$$

The equality holds since, by the above, finite unions of measurable sets are measurable and the inequality holds because of (iii). Further, since the B_i's are disjoint, we have for every $i = 1, 2, \dots,$

$$E \cap B_i = E \cap \left(\bigcap_{j<i} B_j^c\right) \cap B_i$$

and hence the right side of (2) equals

$$\sum_{i=1}^{m-1} \mu\left(E \cap \left(\bigcap_{j<i} B_j^c\right) \cap B_i\right) + \mu\left(E \cap \left(\bigcap_{j<m} B_j^c\right) \cap B_m\right)$$
$$+ \mu\left(E \cap \left(\bigcap_{j=1}^{m} B_j^c\right)\right). \tag{3}$$

By the measurability of B_m the sum of the last two terms in (3) equals

$$\mu\left(E \cap \left(\bigcap_{j=1}^{m-1} B_j^c\right)\right), \tag{4}$$

and hence the right side of (2) is not changed when m is replaced by $m-1$. By peeling off the sets $B_j, j = m, m-1, \ldots, 1$ in this fashion, we see that the right side of (2) equals $\mu(E)$. Hence,

$$\mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)\right) = \sum_{i=1}^{m} \mu(E \cap B_i). \tag{5}$$

In particular, with $E = \Omega$, (5) establishes finite additivity.

Now, for a *countable collection* of disjoint sets $B_1, B_2, \ldots$

$$\mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mu\left(\bigcup_{i=1}^{\infty} (E \cap B_i)\right) \leq \sum_{i=1}^{\infty} \mu(E \cap B_i)$$

by (iii). Thus, by (ii),

$$\mu\left(E \cap \bigcup_{i=1}^{m} B_i\right)$$

is an increasing sequence and

$$\lim_{m\to\infty} \mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)\right) \leq \mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) \leq \sum_{i=1}^{\infty} \mu(E \cap B_i).$$

From this and (5) we conclude that

$$\lim_{m\to\infty} \mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)\right) = \mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right)$$
$$= \sum_{i=1}^{\infty} \mu(E \cap B_i). \tag{6}$$

Since

$$\mu\left(E \cap \left(\bigcup_{i=1}^{m} B_i\right)^c\right) \geq \mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)^c\right),$$

the measurability of $\bigcup_{i=1}^{m} B_i$ together with (5) and (6) yields

$$\mu(E) \geq \mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) + \mu\left(E \cap \left(\bigcup_{i=1}^{\infty} B_i\right)^c\right). \tag{7}$$

In case $\mu(E) = \infty$ equation (1) holds for any set A by (iii) and, in particular, for *any* union (countable or not) of sets. Equation (5) is trivial in case $\mu(E \cap \bigcup_{i=1}^{m} B_i) = \infty$. If $\mu(E \cap \bigcup_{i=1}^{m} B_i)$ is finite, simply replace E by $E' := E \cap \bigcup_{i=1}^{m} B_i$, and then the case $\mu(E) < \infty$ applied to E' yields (5). Thus, (6) and (7) hold generally and, by (iii), $\bigcup_{i=1}^{\infty} B_i$ is measurable.

By setting $E = \Omega$ in (6) we obtain the countable additivity, i.e.,

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i). \tag{8}$$

Having established that countable unions of disjoint measurable sets are measurable, it is straightforward to show that Σ is a sigma-algebra and μ is a countably additive measure on Σ. ■

● Several theorems in this chapter and the next are concerned with the pointwise convergence of a sequence of measurable functions. One might expect that such convergence can be quite 'wild' and irregular, and this is certainly possible. Uniform convergence, as would be appropriate for suitable sequences of continuous functions, is the exception rather than the rule. Nevertheless, a remarkable and useful theorem of [Egoroff] asserts that if the space has finite measure, and if one is prepared to ignore a subset of arbitrarily small measure, then pointwise convergence is *always* uniform.

1.16 THEOREM (Uniform convergence except on small sets)

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) < \infty$, let $f, f^1, f^2, \dots$ be complex-valued, measurable functions on Ω, and assume $f^j(x) \to f(x)$ as $j \to \infty$ for almost every $x \in \Omega$. Then, for every $\varepsilon > 0$ there is a set $A_\varepsilon \subset \Omega$ with $\mu(A_\varepsilon) > \mu(\Omega) - \varepsilon$ such that $f_j(x)$ converges to $f(x)$ uniformly on A_ε. That is, for every $\delta > 0$ there is an N_δ such that when $j > N_\delta$ we have $|f^j(x) - f(x)| < \delta$ for every $x \in A_\varepsilon$.

PROOF. Choose $\delta > 0$. Pointwise convergence at x means that there is an integer $M(\delta, x)$ such that $|f^j(x) - f(x)| < \delta$ for all $j > M(\delta, x)$. For integer N define the sets $S(\delta, N) = \{x : M(\delta, x) \leq N\}$, which obviously are nondecreasing with respect to N and δ. These sets are measurable since $\{x : M(\delta, x) \leq N\} = \bigcup_{M=1}^{N} \bigcap_{j>M} B_j$, where $B_j = \{x : |f^j(x) - f(x)| < \delta\}$. Next, we define $S(\delta) = \bigcup_N S(\delta, N)$. Since almost every x is in some $S(\delta, N)$, we have that $\mu(S(\delta)) = \mu(\Omega)$. Countable additivity is crucial here.

Thus, for every $\delta > 0$ and $\tau > 0$ there is an N such that $\mu(S(\delta, N)) > \mu(\Omega) - \tau$. Let $\delta_1 > \delta_2 > \cdots$ be a sequence of δ's tending to 0, and let N_j be such that $\mu(S(\delta_j, N_j)) > \mu(\Omega) - 2^{-j}\varepsilon$. Set $A_\varepsilon := \bigcap_j S(\delta_j, N_j)$. Obviously, by construction, f^j converges to f uniformly on A_ε.

To complete the proof we have to show that $\mu(A_\varepsilon^c) \leq \varepsilon$. This is an immediate consequence of de Morgan's law, $\left(\bigcap_j S(\delta_j, N_j)\right)^c = \bigcup_j S(\delta_j, N_j)^c$, and the fact that the measure of the right side is less than ε. ■

1.17 SIMPLE FUNCTIONS AND REALLY SIMPLE FUNCTIONS

The beauty and power of measure theory and the Lebesgue integral allows us to deal with functions and their limits economically and elegantly. Nevertheless, Theorem 1.16 suggests that the expanded concept of measurable functions has not really taken us far from the kinds of functions, mostly continuous, that mathematicians thought about in the nineteenth century. We shall explore this idea a little further and also say a little about the connection between our presentation of integration theory and the more customary approach via simple functions. In fact, we shall take a step even further in that direction by tracing the path back to 'really simple functions' — a concept we learned from E. Carlen.

Given a measure space (Ω, Σ, μ), we know what a measurable function is, what a measurable set is, and what the characteristic function of such a set is. The integral of a characteristic function of a measurable set is defined to be the measure of the set. Next, we can define a **simple function** f to be a measurable function that takes on only finitely many values. I.e., $f(x) = \sum_{j=1}^{N} C_j \chi_j(x)$ where $C_j \in \mathbb{C}$ and χ_j is the characteristic function of some measurable set A_j. (Since such an f can be thus written in several ways, it is customary to require the A_j to be disjoint sets and the C_j to be all different; this makes the representation unique but it is often advantageous not to do so — and we shall not impose this requirement.) We can, in any case, define $\int_\Omega f \mathrm{d}\mu = \sum_{j=1}^{N} C_j \mu(A_j)$, and check that this 'definition' is independent of the representation. Finally, the integral of a nonnegative, measurable function, f, is defined to be the supremum of the integrals of simple functions, g, with the property that $0 \leq g(x) \leq f(x)$ for all x. Evidently this definition agrees with the one in 1.5(2); it is only necessary to look at simple functions whose sets A_j are the sets $S_f(t)$ (see 1.5(1)) for suitably chosen values of t. The equivalence of the two definitions stems from the fact that the integral on the right side of 1.5(2) is a Riemann integral and thus can be approximated by a finite sum. We also note that any nonnegative function f can be approximated from below by an increasing sequence of nonnegative simple functions f^j, i.e., $f \geq f^{j+1} \geq f^j \geq 0$.

This way of developing integration theory is not without its advantages. For instance, it makes it easier to prove that $\int(f+g) = \int f + \int g$. One is still left with the problem of understanding measurable sets, however. A measurable set can be weird but, as we shall see, it is not far from a 'nice' set — in the sense of measure.

Let us recall that we start with an algebra of sets $\mathcal{A}$ (containing Ω and the empty set; see the end of Sect. 1.2) and then define the sigma-algebra Σ to be the smallest sigma-algebra containing $\mathcal{A}$. The monotone-class theorem identifies Σ as a more 'natural' object — the smallest monotone class containing $\mathcal{A}$, but it would be helpful if we could define integration in terms of $\mathcal{A}$ *directly.* To this end we define a **really simple function** f to be

$$f(x) = \sum_{j=1}^{N} C_j \chi_j(x),$$

where $C_j \in \mathbb{C}$ and χ_j is the characteristic function of some set A_j in the algebra $\mathcal{A}$. (Again, we can, if we wish, choose the A_j to be disjoint sets and the C_j to all be different.)

An important example is $\Omega = \mathbb{R}^n$ and a member of $\mathcal{A}$ is a set consisting of a finite union (including the empty set) of **half open rectangles**, by which we mean sets of the form

$$A = \{x \in \mathbb{R}^n \, : \, a_i < x_i \leq b_i, \quad 1 \leq i \leq n\} \tag{1}$$

with $a_i < b_i$ for all $1 \leq i \leq n$. Finite unions of such sets form an algebra (why?) but not a sigma-algebra, and confusion about this distinction caused problems in times past. We can even make $\mathcal{A}$ into a countable algebra by requiring the a_i, b_i to be rational. The sigma-algebra generated by $\mathcal{A}$ is the Borel sigma-algebra. (This sigma-algebra is also generated by open sets, but the collection of open sets in $\mathbb{R}^n$ is *not* an algebra. If we want to make an algebra out of the open sets, without going to the full Σ-algebra, we can do so by taking all open sets and all closed sets and their finite unions and intersections. Unlike (1), this algebra has the virtue that it can be defined for general metric spaces, for example, but this algebra is not as easy to picture as (1).) We can take the measure to be Lebesgue measure $\mathcal{L}^n$, whose definition for a set in $\mathcal{A}$ is evident, but we can also consider any other measure μ defined on this sigma-algebra.

In the general case we suppose that a set Ω and an algebra $\mathcal{A}$ — and hence Σ — are given. We suppose also that the measure μ is given, but we make the additional assumption that Ω is sigma-finite in the strong sense of Theorem 1.4 (uniqueness of measures), namely that Ω can be covered by countably many sets in $\mathcal{A}$ of finite measure (without using other sets in Σ). This is certainly true of $\mathbb{R}^n$ with Lebesgue measure and the algebra $\mathcal{A}$ just mentioned. For the purposes of what we want to do in the following, it is convenient to replace $\mathcal{A}$ by the subalgebra consisting of those sets in $\mathcal{A}$ that have finite μ-measure. Thus, we shall assume henceforth that

$$\mu(A) < \infty \qquad \text{for all } A \in \mathcal{A}. \tag{2}$$

Sigma-finiteness in the strong sense means now that Ω can be covered by countably many sets in $\mathcal{A}$ (since all sets in $\mathcal{A}$ now have finite measure). All really simple functions are bounded and summable.

The question to be answered is whether summable functions can be approximated by really simple functions in the sense of integrals (or, to use the terminology of the next chapter, in the $L^1(\Omega)$ sense). The next theorem answers this affirmatively, and the heuristic implication of this is that while there may be many more sets in Σ than in $\mathcal{A}$, the additional sets are not critically important for evaluating an integral.

1.18 THEOREM (Approximation by really simple functions)

Let (Ω, Σ, μ) be a measure space with Σ generated by an algebra $\mathcal{A}$. Assume that Ω is sigma-finite in the strong sense mentioned above. Let f be a complex-valued summable function and let $\varepsilon > 0$. Then there is a really simple function h_ε such that

$$\int_\Omega |f - h_\varepsilon| \mathrm{d}\mu < \varepsilon. \tag{1}$$

PROOF. The proof will show, once again, the utility of Theorem 1.3 (monotone class theorem). Without loss of generality we can suppose that f is real-valued and $f \geq 0$ (why?). In view of what was said in Sect. 1.17 about the fact that there is a simple function f_ε for which $\int_\Omega |f - f_\varepsilon| \, \mathrm{d}\mu < \varepsilon$ for any $\varepsilon > 0$, it suffices to prove (1) when f is the characteristic function of some measurable set C of finite μ-measure.

Let us define $\mathcal{B}$ to be the family of sets $B \in \Sigma$ such that $\mu(B) < \infty$ and such that for *every* $\varepsilon > 0$ there is an $A_\varepsilon \in \mathcal{A}$ satisfying the condition

$$\mu(B \Delta A_\varepsilon) < \varepsilon, \tag{2}$$

where $X \Delta Y := (X \sim Y) \cup (Y \sim X)$ denotes the **symmetric difference** of the sets X and Y.

Clearly, $\mathcal{A} \subset \mathcal{B}$. Our goal is to show that $\mathcal{B} = \widetilde{\Sigma}$, where $\widetilde{\Sigma}$ denotes the sets in Σ with finite μ-measure.

Assume, provisionally, that $\mu(\Omega) < \infty$. If B_j is an increasing family in $\mathcal{B}$, set $\beta = \bigcup_k B_k$. Since $\mu(\Omega) < \infty$, we have that $\mu(\beta) < \infty$. We want to show that $\mu(\beta \Delta A) \le \varepsilon$ for some $A \in \mathcal{A}$.

We set $\sigma_j := \beta \sim B_j$ and choose j large enough so that $\mu(\sigma_j) < \varepsilon/2$. By definition, we can find an $A_j \in \mathcal{A}$ so that $\mu(B_j \Delta A_j) < \varepsilon/2$. Now we compute the measure of $\beta \Delta A_j = (\beta \sim A_j) \cup (A_j \sim \beta)$. First, we have that $A_j \sim \beta \subset A_j \sim B_j$, so $\mu(A_j \sim \beta) \le \mu(A_j \sim B_j)$. Second, we set $X = B_j \sim A_j$ and $Y = \sigma_j \sim A_j \subset \sigma_j$, so $\beta \sim A_j = X \cup Y$. Then

$$\begin{aligned}\mu(\beta \sim A_j) &\le \mu(X) + \mu(Y) \le \mu(X) + \mu(\sigma_j) \\ &= \mu(B_j \sim A_j) + \mu(\sigma_j) \le \mu(B_j \sim A_j) + \varepsilon/2.\end{aligned}$$

If we add our inequalities for $\mu(A_j \sim \beta)$ and $\mu(\beta \sim A_j)$ we obtain

$$\mu(\beta \Delta A_j) \le \mu(A_j \sim B_j) + \mu(B_j \sim A_j) + \varepsilon/2 \le \mu(B_j \Delta A_j) + \varepsilon/2 \le \varepsilon.$$

Similarly, we can show that the intersection of a decreasing family in $\mathcal{B}$ is in $\mathcal{B}$, and, therefore, $\mathcal{B}$ is a monotone class. If we also assume, provisionally, that Ω is in $\mathcal{A}$, then, by the monotone class theorem, $\mathcal{B} = \Sigma$ and we are done.

The obstacle to using the monotone class theorem in the general case is the condition $\Omega \in \mathcal{A}$. Recall that we only need to approximate the set C mentioned at the beginning, and that $\mu(C) < \infty$. By assumption, there are sets $A_1, A_2, \ldots$ in $\mathcal{A}$ such that $\Omega = \bigcup_{j=1}^{\infty} A_j$. Therefore, there is a finite number J such that if we define $\Omega' = \bigcup_{j=1}^{J} A_j$, then the set $C' := \Omega' \cap C \subset C$ is close to C in the sense that $\mu(C \sim C') < \varepsilon/2$. We can now carry out the previous proof with the following changes: (1) replace Ω by Ω'; (2) replace C by C'; (3) replace the algebra $\mathcal{A}$ by the *subalgebra* $\mathcal{A}' \subset \mathcal{A}$, consisting of the sets $A \subset \Omega'$ with $A \in \mathcal{A}$. (Check that $\mathcal{A}'$ is an algebra.) Since $\Omega' \in \mathcal{A}'$, we see that we can find an $A \in \mathcal{A}'$ so that $\mu(C' \Delta A) < \varepsilon/2$. ■

1.19 COROLLARY (Approximation by C^∞ functions)

Let Ω be an open subset of $\mathbb{R}^n$ and let μ be a measure on the Borel sigma-algebra of Ω. Let $\mathcal{A}$ be the algebra of half open rectangles of 1.17(1) *and assume that Ω is sigma-finite in the strong sense. Assume, also, that every finite, closed rectangle that is contained in Ω has finite μ-measure. If f is a μ-summable function, then, for each $\varepsilon > 0$, there is a $C^\infty(\mathbb{R}^n)$ function g_ε such that*

$$\int_\Omega |f - g_\varepsilon| \, \mathrm{d}\mu < \varepsilon. \tag{1}$$

REMARKS. (1) Since g_ε is in $C^\infty(\mathbb{R}^n)$, it is automatically in $C^\infty(\Omega)$.

(2) This Corollary gives a different approach to $C^\infty(\mathbb{R}^n)$ approximation than the one presented in Theorem 2.16. Approximation by convolution, as in 2.16, is, however, useful in many contexts.

PROOF. From Theorem 1.18, it suffices to prove that the characteristic function of a half open rectangle $H \subset \Omega$ of finite measure can be approximated to arbitrary accuracy, in the sense of (1), by a $C^\infty(\mathbb{R}^n)$ function. This is easily accomplished. We shall demonstrate it in $\mathbb{R}^1$ for convenience; the extension to $\mathbb{R}^n$ is trivial.

The "rectangle" H is, e.g., the interval $H = (a, b]$. Since Ω is open, it contains some *closed* rectangle $G = [a + \delta, b + \delta]$ and $\mu(G) < \infty$ by assumption.

Let $h_\varepsilon(x) := f(x/\varepsilon)$, where

$$f(x) = \begin{cases} \exp\left[-\left\{\exp\left[x/(1-x)\right] - 1\right\}^{-1}\right], & \text{if } 0 < x < 1, \\ 0, & \text{if } x \le 0, \\ 1, & \text{if } x \ge 1, \end{cases}$$

which is an infinitely differentiable function. Let

$$g_\varepsilon(x) = \begin{cases} h_\varepsilon(x - a - \varepsilon), & \text{if } x \le a + \varepsilon, \\ 1, & \text{if } a + \varepsilon \le x \le b, \\ h_\varepsilon(x - b), & \text{if } x \ge b. \end{cases}$$

It is easy to check that g_ε is infinitely differentiable. As $\varepsilon \to 0$, $g_\varepsilon(x) \to \chi_H(x)$ for *every* x. The convergence is monotone decreasing if $x \ge b$ and monotone increasing if $x < b$, but this is of no consequence. The important point is that $0 \le g_\varepsilon(x) \le \chi_G(x) + \chi_H(x)$ when $\varepsilon < \delta$. Thus, (1) follows by the dominated convergence theorem. ∎

Exercises for Chapter 1

1. Complete the proof of Theorem 1.3 (monotone class theorem).

2. With regard to the remark about continuous functions in Sect. 1.5, show that f is continuous (in the sense of the usual ε, δ definition) if and only if f is both upper and lower semicontinuous. Show that f is upper semicontinuous at x if and only if, for every sequence $x_1, x_2, \ldots$ converging to x, we have $f(x) \geq \limsup_{n\to\infty} f(x_n)$.

3. Prove the assertion made in Sect. 1.5 that for any Borel set $A \subset \mathbb{R}$ and any sigma-algebra Σ the set $\{x : f(x) \in A\}$ is Σ-measurable whenever the function f is Σ-measurable.

4. (Continuation of Problem 3): Let $\phi : \mathbb{C} \to \mathbb{C}$ be a Borel measurable function and let the complex-valued function f be Σ-measurable. Prove that $\phi(f(x))$ is Σ-measurable.

5. Prove equation (2) in Theorem 1.6 (monotone convergence).

6. Give the alternative proof of the layer cake representation, alluded to in Remark (4) of 1.13, that does not make use of Fubini's theorem.

7. Prove Theorem 1.14 (bathtub principle).

8. Prove the statement about finite unions and intersections in the first paragraph of the proof of Theorem 1.15 (constructing a measure from an outer measure).

   ► *Hint.* For any two measurable sets A, B and E arbitrary, show that

   $$\begin{aligned}\mu(E) = \mu(E \cap A \cap B) + \mu(E \cap A^c \cap B) + \mu(E \cap A \cap B^c) \\ + \mu(E \cap A^c \cap B^c).\end{aligned}$$

   Use this to prove that $A \cap B$ is measurable.

9. Verify the linearity of the integral as given in 1.5(7) by completing the steps outlined below. In what follows, f and g are nonnegative summable functions.
   a) Show that $f + g$ is also summable. In fact, by a simple argument $\int(f+g) \leq 2(\int f + \int g)$.
   b) For any integer N find two functions f_N and g_N that take only finitely many values, such that $|\int f - \int f_N| \leq C/N$, $|\int g - \int g_N| \leq C/N$ and

$|\int(f+g) - \int(f_N+g_N)| \leq C/N$ for some constant C independent of N.

c) Show that for f_N and g_N as above $\int(f_N+g_N) = \int f_N + \int g_N$, thus proving the additivity of the integral for nonnegative functions.

d) In a similar fashion, show that for $f, g \geq 0$, $\int(f-g) = \int f - \int g$.

e) Now use c) and d) to prove the linearity of the integral.

10. Prove that when we add and subtract the subsets of sets of zero measure to the sets of a sigma-algebra then the result is again a sigma-algebra and the extended measure is again a measure.

11. Prove that the measure constructed in Theorem 1.15 is complete, i.e., every subset of a measurable set that has measure zero is measurable.

12. Find a simple condition on $f_n(x)$ so that

$$\sum_{n=0}^{\infty} \int_\Omega f_n(x)\mu(\mathrm{d}x) = \int_\Omega \left\{ \sum_{n=0}^{\infty} f_n(x) \right\} \mu(\mathrm{d}x).$$

13. Let f be the function on $\mathbb{R}^n$ defined by $f(x) = |x|^{-p}\chi_{\{|x|<1\}}(x)$. Compute $\int f \,\mathrm{d}\mathcal{L}^n$ in two ways: (i) Use polar coordinates and compute the integral by the standard calculus method. (ii) Compute $\mathcal{L}^n(\{x : f(x) > a\})$ and then use Lebesgue's definition.

14. Prove that $j(x)$, defined in 1.1(2), is infinitely differentiable.

15. **Urysohn's lemma.** Let $\Omega \subset \mathbb{R}^n$ be open and let $K \subset \Omega$ be compact. Prove that there is a $\psi \in C_c^\infty(\Omega)$ with $\psi(x) = 1$ for all $x \in K$.

   ▶ *Hints.* (a) Replace K by a slightly larger compact set K_ε, i.e., $K \subset K_\varepsilon \subset \Omega$; (b) Using the distance function $d(x, K_\varepsilon) = \inf\{|x-y| : y \in K_\varepsilon\}$, construct a function $\psi_\varepsilon \in C_c^0(\Omega)$ with $\psi_\varepsilon = 1$ on K_ε and $\psi_\varepsilon(x) = 0$ for $x \notin K_{2\varepsilon} \subset \Omega$; (c) Take $j_\varepsilon(x) = \varepsilon^{-n} j(x/\varepsilon)$, with j given in Exercise 14 and $\int j = 1$ (here $\int$ denotes the *Riemann* integral from elementary calculus). Define $\psi(x) = \int j_\varepsilon(x-y)\psi_\varepsilon(y)\,\mathrm{d}y$ (again, the Riemann integral); (d) Verify that ψ has the correct properties. To show that $\psi \in C_c^\infty(\Omega)$ it will be necessary to differentiate 'under the integral sign', a process that can be justified with standard theorems from calculus.

16. Let $\Omega \subset \mathbb{R}^n$ be open and $\phi \in C_c^\infty(\Omega)$. Show that there exist nonnegative functions ϕ_1 and ϕ_2, both in $C_c^\infty(\Omega)$, such that $\phi = \phi_1 - \phi_2$.

17. Show that the infimum of a family of continuous functions is upper semicontinuous.

18. Simple facts about measure:
   a) Show that the condition $\{x : f(x) > a\}$ is measurable for all $a \in \mathbb{R}$ holds if and only if it holds for all rational a.
   b) For rational a, show that

$$\{x : f(x) + g(x) > a\} = \bigcup_{b \text{ rational}} \big(\{x : f(x) > b\} \cap \{x : g(x) > a - b\}\big).$$

   c) In a similar way, prove that fg is measurable if f and g are measurable.

19. Give a 'counterexample' to Fubini's theorem in the absence of sigma-finiteness.

   ▶ *Hint.* Take Lebesgue measure on $[0, 1]$ as one space and counting measure on $[0, 1]$ as the other. (**The counting measure** of a set is just the number of elements in the set.)

20. If f and g are two continuous functions on a common open set in $\mathbb{R}^n$ that agree everywhere on the complement of a set of zero Lebesgue measure, then, in fact, f and g agree everywhere.

21. Prove that if $f : \mathbb{R}^n \to \mathbb{C}$ is uniformly continuous and summable, then the Riemann integral of f equals its Lebesgue integral.

22. Theorem 1.10 (product measure) asserts that f and g are measurable functions. Prove this by imitating the proof of the section property in Sect. 1.2 and by using the Monotone Class Theorem.

23. A concept we shall need later on is a connected open set. In elementary topology one learns that there are two notions of a topological space Ω being connected:
   1) **Topologically connected**, i.e., that $\Omega \neq A \cup B$ with $A \cap B = \emptyset$ and where A and B are both open (in the topology of Ω).
   2) **Arcwise connected**, i.e., it is possible to connect any two points of Ω by a continuous curve lying entirely in Ω . Arcwise connectedness implies topological connectedness, but the converse does not hold, generally.
      a) Define "continuous curve".
      b) Prove that if $\Omega \subset \mathbb{R}^n$ is *open*, then topological connectedness implies arcwise connectedness.

   ▶ *Hint.* Arcwise connectedness defines a relation among points.

24. With the same assumptions as in Egoroff's theorem, show that if

$$\int_\Omega |f^j|^2 \, \mathrm{d}\mu < 1 \quad \text{and} \quad \int_\Omega |f|^2 \, \mathrm{d}\mu < \infty,$$

then $\int_\Omega |f^j - f|^p \, d\mu \to 0$ as $j \to \infty$ for any $0 < p < 2$. Construct a counterexample to show that this can fail for $p = 2$.

25. A theorem closely related to Egoroff's theorem is **Lusin's theorem.** *Let μ be a Borel measure on $\mathbb{R}^n$ and let Ω be a measurable subset of $\mathbb{R}^n$ with $\mu(\Omega) < \infty$. Let f be a measurable, complex-valued function on Ω. Then for each $\varepsilon > 0$ there is a continuous function f_ε such that $f_\varepsilon(x) = f(x)$ except on a set of measure less than ε.* Prove this.

    ▶ *Hint.* Urysohn's lemma can be helpful.

26. Using the monotone class theorem, imitate the proof of Theorem 1.18 to prove that Lebesgue measure is inner and outer regular.

27. Referring to Theorem 1.18, it would be false to assert that a measurable set B can be approximated from the inside by a member of the algebra $\mathcal{A}$. Consider $\mathbb{R}^n$ and the half open rectangle algebra in 1.17(1). Find a closed set in $\mathbb{R}^n$ of finite measure that contains *no* member of $\mathcal{A}$.

28. Verify that the sigma-algebra Σ generated by the half open rectangles in 1.17(1) is the Borel sigma-algebra on $\mathbb{R}^n$. Show explicitly that open and closed rectangles are in Σ.