3

Measure and Integration

The foundations of integration theory date to the classical Greek period. The most notable contribution from that time is the "method of exhaustion" due to Eudoxos (ca. 408–355 B.C.E.; Asia Minor, now Turkey). Over two thousand years later, Augustin Cauchy stressed the importance of defining an integral as a limit of sums. One's first encounter with a theory of the integral is usually with a variation on Cauchy's definition given by Georg Friedrich Bernhard Riemann (1826–1866; Hanover, now Germany). Though the Riemann integral is attractive for many reasons and is an appropriate integral to learn first, it does have deficiencies. For one, the class of Riemann integrable functions is too small for many purposes. Henri Lebesgue gave, around 1900, another approach to integration. In addition to the integrals of Riemann and Lebesgue, there are yet other integrals, and debate is alive about which one is the best. Arguably, there is no one best integral. Different integrals work for different types of problems. It can be said, however, that Lebesgue's ideas have been extremely successful, and that the Lebesgue integrable functions are the "right" ones for many functional analysts and probabilists. It is no coincidence that the rapid development of functional analysis coincided with the emergence of Lebesgue's work.

As you will discover in this chapter, Lebesgue's ideas on integration are intertwined with his notion of measure. It was this idea — of using measure theory as a platform for integration — that marked a departure from what had been done previously. Lebesgue's measure theory and application to integration theory appear in his doctoral thesis [80]. This thesis is considered one of the greatest mathematical achievements of the twentieth century.

The sole reason that this material on measure and integration is included in this book is because of the important role it plays in functional analysis. Frigyes Riesz was familiar with Lebesgue's work, and also with David Hilbert's work on integral equations and Maurice Fréchet's work on abstract function spaces. He combined these elements brilliantly, developing theories now considered basic to the field of functional analysis. Riesz deserves much credit for recognizing the importance of Lebesgue's ideas and drawing attention to them. Riesz's work will be introduced in the last section of this chapter, and further explored in the next chapter, culminating with the celebrated Riesz–Fischer theorem.

We start our study of measure by considering some problems from probability. Probability theory had, in some sense, been in the mathematical world since the mid-seventeenth century. However, it wasn't until the 1930s, when Andrei Kolmogorov (1903–1987; Russia) laid the foundation for the theory using Lebesgue's measure theory, that probability was truly viewed as a branch of pure mathematics. It is therefore historically inaccurate to use probability to motivate measure theory. Nonetheless, the applications of measure theory to probability theory are beautiful, and they provide very good source of inspiration for students about to embark on their first journey into the rather technical field of measure theory.

3.1 Probability Theory as Motivation

In this section we give an informal introduction to Lebesgue's theory of measure, using an example from probability as inspiration. The ideas for the presentation of this material come from [1]. Consider a sequence of coin tosses of a fair coin. We represent such a sequence by, for example,

$$\text{THHTTTHTHTT}\ldots. \tag{3.1}$$

Let

$$s_n = \text{the number of heads in } n \text{ tosses.}$$

The law of large numbers asserts, in some sense or other, that the ratio $\frac{s_n}{n}$ approaches $\frac{1}{2}$ as n gets larger. The goal of this section is to rephrase this law in measure-theoretic language, thus indicating that measure theory provides a "framework" for probability. Although measure theory did not arise because of probability theorists "looking for a language," one could argue that probability theory helped to ensure measure theory's importance as a branch of pure mathematics worthy of research efforts. We will let this discussion of very basic probability theory serve as a source of motivation for learning about measure theory, and not for continuing the discussion about probability theory. A further investigation of probability theory is a worthy endeavor, but it would take us too far afield to do it here.

The law of large numbers was first stated in the seventeenth century by James (=Jakob=Jacques) Bernoulli (1654–1705; Switzerland). In his honor a sequence, like (3.1) of independent trials with two possible outcomes is called a *Bernoulli sequence*. Let $\mathcal{B}$ denote the collection of all Bernoulli sequences. This is the so-called sample space from probability theory, and is often denoted by Ω in that context.

We note that $\mathcal{B}$ is uncountable. This can be shown using a Cantor diagonalization argument. An alternative, and more useful, proof proceeds by showing that all but a countable number of the elements of $\mathcal{B}$ can be put in to one-to-one correspondence with the interval (0, 1]. Specifically, let $\mathcal{B}_T$ denote the set of Bernoulli trials that are constantly T after a while. Then $\mathcal{B}_T$ is countable (left as an exercise), and so $\mathcal{B} \setminus \mathcal{B}_T$ is still uncountable if $\mathcal{B}$ is. The reader is asked to write out the details of both proofs that $\mathcal{B}$ is uncountable in Exercises 3.1.1 and 3.1.2. To a Bernoulli sequence associate the real number ω whose binary expansion is $\omega = .a_1a_2a_3\ldots$, where $a_i = 1$ if the ith toss in the sequence is a head, and $a_i = 0$ if the ith toss in the sequence is a tail. In the case that ω has two binary expansions, exactly one will be nonterminating, and we use this one. Thus,

$$\omega = \sum_{i=1}^{\infty} \frac{a_i}{2^i}.$$

This allows us to identify subsets E of $\mathcal{B}$ ("events") with subsets B_E of (0, 1].

In the general theory of probability, the sample space Ω can be any set. If Ω is finite or countably infinite, Ω is called a *discrete probability space*, and the probability theory is relatively straightforward. However, if Ω is uncountably infinite, as $\mathcal{B}$ is, then the sets B_E can be quite complicated. It is for determining the "size" of these sets B_E that we require measure theory.

Heuristically, a *measure* μ on a space Ω should be a nonnegative function defined on certain subsets of Ω (hereafter referred to as *measurable sets*). For a measurable set A, $\mu(A)$ denotes the measure of A. As a guiding principle of sorts, we require that

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mu(A_i)$$

for every finite collection $\{A_i\}_{i=1}^n$ of measurable sets satisfying $A_i \cap A_j = \emptyset$ $(i \neq j)$.

We will begin our study of abstract measures with a specific example of a measure: *Lebesgue measure*, often denoted by μ_L or m, on certain subsets of $\Omega = \mathbb{R}$. We will require that $m(I) = b - a$ if I is any one of the four intervals (a, b), $(a, b]$, $[a, b)$, and $[a, b]$. It follows from this that each finite subset of $\mathbb{R}$ has Lebesgue measure zero. Every countable set will be seen to have Lebesgue measure zero, and we will see that there are even uncountably infinite sets with Lebesgue measure zero (the Cantor set is such a set, see Exercise 3.2.10).

We now discuss the connection between probability and measure theory. We associate subsets E of $\mathcal{B}$ with subsets B_E of the interval (0, 1], as before. We then *define* the *probability that the event E occurs*, Prob(E), to be the Lebesgue measure $m(B_E)$ of the set B_E. Using this, let us look at two very basic examples and see that the value of $m(B_E)$ agrees with what we would expect Prob(E) to be from everyday experience.

Example 1. Let E be the event that a head is thrown on the first toss. We know that Prob(E) should equal $\frac{1}{2}$. Let's now figure out what the set B_E is, and then see whether this set has Lebesgue measure $\frac{1}{2}$. A number ω is in B_E if and only if $\omega = 0.1a_2a_3\ldots$. Therefore, ω is in B_E if and only if $\omega \geq 0.1000\ldots$ and $\omega \leq 0.1111\ldots$. That is, $B_E = [\frac{1}{2}, 1]$. Then $m(B_E) = \frac{1}{2}$, as desired.

EXAMPLE 2. In the first example we considered the event that the first toss is prescribed. This time we let E be the event that the first n tosses are prescribed. Let us say that these first n tosses are $a_1, a_2, a_3, \ldots, a_n$. We know that Prob$(E)$ should equal $\left(\frac{1}{2}\right)^n$. As in the first example, we now try to identify the set B_E, and figure out its Lebesgue measure. If we let $s = 0.a_1a_2a_3\ldots a_n00000\ldots$, then ω is in B_E if and only $\omega \geq s$ and $\omega \leq 0.a_1a_2a_3\ldots a_n11111\ldots$. But

$$0.a_1a_2a_3\ldots a_n11111\ldots = s + \sum_{i=n+1}^{\infty} \frac{a_i}{2^i} = s + \left(\frac{1}{2}\right)^n.$$

Therefore, $B_E = \left[s, s + \left(\frac{1}{2}\right)^n\right]$, and $m(B_E) = \left(\frac{1}{2}\right)^n$, as desired.

We now return to the law of large numbers. We will give two versions of this result. Since the material of this section is primarily offered to motivate a study of measure, proofs are not included. First, for $\omega \in (0, 1]$, we define $s_n(\omega) = a_1 + \cdots + a_n$, where $0.a_1a_2\ldots$ is the binary representation for ω.

Theorem (Weak Law of Large Numbers). *Fix $\epsilon > 0$ and define, for each positive integer n, the set*

$$B_n = \left\{\omega \in (0, 1] \,\middle|\, \left|\frac{s_n(\omega)}{n} - \frac{1}{2}\right| > \epsilon\right\}.$$

This subset of $(0, 1]$ corresponds to the event that "after the first n tosses, the number of heads is not close to $\frac{1}{2}$." The weak law of large numbers states that $m(B_n) \to 0$ as $n \to 0$.

Theorem (Strong Law of Large Numbers). *Let*

$$S = \left\{\omega \in (0, 1] \,\middle|\, \lim_{n\to\infty} \frac{s_n(\omega)}{n} = \frac{1}{2}\right\}.$$

The strong law of large numbers states that $m((0, 1] \setminus S) = 0$.

We end this section by remarking that the set $(0, 1] \setminus S$ in the strong law of large numbers is uncountable. After the Cantor set, this is our second example of an uncountable set of Lebesgue measure zero.

3.2 Lebesgue Measure on Euclidean Space

Before we give a formal treatment of Lebesgue measure on $\mathbb{R}^n$, we give a few general definitions. A family $\mathcal{R}$ of sets is called a *ring* if $A \in \mathcal{R}$ and $B \in \mathcal{R}$ imply $A \cup B \in \mathcal{R}$ and $A \setminus B \in \mathcal{R}$ (we remark that B need not be a subset of A in order to define $A \setminus B$). A ring $\mathcal{R}$ is called a *σ-ring* if $A_k \in \mathcal{R}$, $k = 1, 2, \ldots$, implies

$\bigcup_{k=1}^{\infty} A_k \in \mathcal{R}$. We consider functions μ defined on a ring or σ-ring $\mathcal{R}$ and taking values in $\mathbb{R} \cup \{\pm\infty\}$. Such a function μ is *additive* if

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

whenever $A \cap B = \emptyset$. If for each sequence $A_k \in \mathcal{R}$, $k = 1, 2, \ldots$, with $\bigcup_{k=1}^{\infty} A_k \in \mathcal{R}$ we have

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k)$$

whenever $A_k \cap A_j = \emptyset$ $(k \neq j)$, we say that μ is *countably additive*. We must assume that the range of μ does not contain both ∞ and $-\infty$, or else the right side of $\mu(A \cup B) = \mu(A) + \mu(B)$ might not make sense. A countably additive, nonnegative function μ defined on a ring $\mathcal{R}$ is called a *measure*. The elements of $\mathcal{R}$ are subsets of some set X; X is called the *measure space*.

In general, measures can (and do) exist on rings consisting of subsets of any set. We will see examples of such abstract measures in the last section of this chapter. Until then we restrict ourselves to Euclidean space $\mathbb{R}^n$, and develop Lebesgue measure on a (yet-to-be-specified) ring of subsets of $\mathbb{R}^n$.

We consider subsets of $\mathbb{R}^n$ of form

$$\{(x_1, \ldots, x_n) \,\big|\, a_k \leq x_k \leq b_k, k = 1, \ldots, n\},$$

where $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are fixed elements in $\mathbb{R}^n$ with each $a_k \leq b_k$. Often we use the notation

$$[a_1, b_1] \times \cdots \times [a_n, b_n]$$

to denote the set just described. Any or all of the $\leq$ signs may be replaced by $<$, with corresponding changes made in the interval notation. Such subsets of $\mathbb{R}^n$ are called the *intervals* of $\mathbb{R}^n$. For an interval I, we define the Lebesgue measure $m(I)$ of I by

$$m(I) = \prod_{k=1}^{n} (b_k - a_k).$$

This definition is independent of whether $\leq$s or $<$s appear in the definition of I. It should be noted that if $n = 1, 2$, or 3, then m is the length, area, or volume of I. We can extend m to $\mathcal{E}$, the collection of all finite unions of disjoint intervals, by requiring that m be additive. In Exercise 3.2.4 you are asked to show that $\mathcal{E}$ is a ring. Note that $m(A) < \infty$ for any $A \in \mathcal{E}$.

Lemma 3.1. *If $A \in \mathcal{E}$ and $\epsilon > 0$, then there exists a closed set $F \in \mathcal{E}$ and an open set $G \in \mathcal{E}$ such that $F \subseteq A \subseteq G$ and*

$$m(F) \geq m(A) - \epsilon \quad \text{and} \quad m(G) \leq m(A) + \epsilon.$$

PROOF. Left as Exercise 3.2.5. □

Theorem 3.2. *m is a measure on $\mathcal{E}$.*

PROOF. All that needs to be shown is that m is countably additive. Suppose that $\{A_k\}_{k=1}^{\infty}$ is a disjoint collection of sets in $\mathcal{E}$ and that $A = \bigcup_{k=1}^{\infty} A_k$ is also in $\mathcal{E}$. For each N, $\bigcup_{k=1}^{N} A_k \subseteq A$, and so (by Exercise 3.2.2(a))

$$m(A) \geq m\Big(\bigcup_{k=1}^{N} A_k\Big) = \sum_{k=1}^{N} m(A_k).$$

Since this holds for each N,

$$m(A) \geq \sum_{k=1}^{\infty} m(A_k).$$

We now aim to show that the other inequality also holds. Choose a closed set F corresponding to A as in Lemma 3.1 to satisfy

$$m(F) \geq m(A) - \epsilon.$$

Choose an open set G_k for each A_k as in Lemma 3.1 satisfying

$$m(G_k) \leq m(A_k) + \frac{\epsilon}{2^k}.$$

Notice that F is closed and bounded, and thus is compact by the Heine–Borel theorem. Since $\{G_k\}_{k=1}^{\infty}$ is an open cover for F, there exists an integer N such that

$$F \subseteq G_1 \cup G_2 \cup \cdots \cup G_N.$$

Then

$$\begin{aligned} m(A) - \epsilon \leq m(F) &\leq m(G_1 \cup \cdots \cup G_N) \leq m(G_1) + \cdots + m(G_N) \\ &\leq m(A_1) + \cdots + m(A_N) + \epsilon. \end{aligned}$$

From this,

$$m(A) \leq \sum_{k=1}^{N} m(A_k) + 2\epsilon \leq \sum_{nk=1}^{\infty} m(A_k) + 2\epsilon.$$

Since $\epsilon > 0$ was arbitrary,

$$m(A) \leq \sum_{k=1}^{\infty} m(A_k),$$

completing the proof. □

We have now constructed Lebesgue measure m on $\mathcal{E}$ consisting of certain subsets of $\mathbb{R}^n$. The reader should verify that $\mathcal{E}$ is a ring, but not a σ-ring (Exercise 3.2.4). We would like to extend m to a much larger (σ-)ring of subsets of $\mathbb{R}^n$. First, we point out that for any set X, the collection of all subsets of X is a ring (and is even a σ-ring). This ring is often denoted by 2^X. To extend m to a larger ring than $\mathcal{E}$ we proceed by first extending m to an "outer measure" m^* defined on all of the ring $2^{(\mathbb{R}^n)}$. Unfortunately, m^* will not actually be a measure on $2^{(\mathbb{R}^n)}$(hence the new name "outer measure"). We will then take a certain collection $\mathcal{M}$ such that

$\mathcal{E} \subseteq \mathcal{M} \subseteq 2^{(\mathbb{R}^n)}$. Happily, $\mathcal{M}$ will be big enough to be a σ-ring, and small enough so that m^* restricted to $\mathcal{M}$ will be a measure.

Let A be any subset of $\mathbb{R}^n$ and consider a countable covering of A with intervals I_k such that $A \subseteq \bigcup_{k=1}^{\infty} I_k$. We define the *outer measure* $m^*(A)$ of A by

$$m^*(A) = \inf \sum_{k=1}^{\infty} m(I_k),$$

where the infimum is taken over all such coverings of A. We call m^* the outer measure corresponding to m.

Note that m^* is defined on all of $2^{(\mathbb{R}^n)}$. It should be clear to the reader that m^* is nonnegative and monotone (that is, $A \subseteq B$ implies $m^*(A) \leq m^*(B)$), and that $m(A) = m^*(A) < \infty$ for $A \in \mathcal{E}$. Also, the reader should check (and is given the opportunity in Exercise 3.2.6) that m^* is *countably subadditive*, that is,

$$m^*\Big(\bigcup_{k=1}^{\infty} A_k\Big) \leq \sum_{k=1}^{\infty} m^*(A_k),$$

whenever $A_1, A_2, \ldots$ are subsets of $\mathbb{R}^n$.

For two sets A, B in $\mathbb{R}^n$, we define their *symmetric difference*

$$S(A, B) = (A \setminus B) \cup (B \setminus A)$$

and the distance from A to B by

$$D(A, B) = m^*\Big(S(A, B)\Big).$$

We let $\mathcal{M}_{\mathcal{F}}$ denote the collection of subsets A of $\mathbb{R}^n$ such that $D(A_k, A) \to 0$ as $k \to \infty$ for some sequence of sets $A_k \in \mathcal{E}$. We let $\mathcal{M}$ denote the collection of subsets of $\mathbb{R}^n$ that can be written as a countable union of sets in $\mathcal{M}_{\mathcal{F}}$. It should be evident that $\mathcal{M}_{\mathcal{F}} \subseteq \mathcal{M}$.

As a precursor to the next lemma, we point out that m^* satisfies a sort of continuity condition. Consider two subsets A, B of $\mathbb{R}^n$ with at least one of $m^*(A)$ and $m^*(B)$ finite; we assume that $m^*(B) < \infty$ and that $m^*(B) < m^*(A)$. Then

$$m^*(A) = D(A, \emptyset) \leq D(B, \emptyset) + D(A, B) = m^*(B) + D(A, B).$$

Therefore,

$$|m^*(A) - m^*(B)| \leq D(A, B).$$

Lemma 3.3. *m^* is additive on $\mathcal{M}_{\mathcal{F}}$.*

Proof. Let A and B be disjoint sets in $\mathcal{M}_{\mathcal{F}}$. We aim to prove that

$$m^*(A \cup B) = m^*(A) + m^*(B).$$

We choose sets A_k, B_k in $\mathcal{E}$ such that $D(A_k, A) \to 0$ and $D(B_k, B) \to 0$ as $k \to \infty$. From Exercise 3.2.7(e) we have

$$D(A_k \cup B_k, A \cup B) \to 0 \qquad \text{as } n \to \infty.$$

Since m^* restricted to $\mathcal{E}$ coincides with m, we can make two observations (see Theorem 3.2):

$$m^*(A_k \cup B_k) = m^*(A_k) + m^*(B_k)$$

for each k, and all three terms in this equation are finite. Using these two observations and the continuity property of m^* we have

$$\begin{aligned}
&|m^*(A \cup B) - m^*(A) - m^*(B)| \\
&\quad \le |m^*(A \cup B) - m^*(A_k \cup B_k)| + |m^*(A_k) - m^*(A)| + |m^*(B_k) - m^*(B)| \\
&\quad \le D(A_k \cup B_k, A \cup B) + D(A_k, A) + D(B_k, B).
\end{aligned}$$

Since all three terms of this sum tend to zero as $k \to \infty$, we are done. □

Lemma 3.4. *$\mathcal{M}_{\mathcal{F}}$ is a ring.*

PROOF. Consider A, B in $\mathcal{M}_{\mathcal{F}}$ and A_k, B_k in $\mathcal{E}$ such that $D(A_k, A) \to 0$ and $D(B_k, B) \to 0$ as $k \to \infty$. Then, for each k, $A_k \cup B_k \in \mathcal{E}$ by Exercise 3.2.4, and by Exercise 3.2.7(e),

$$D(A_k \cup B_k, A \cup B) \to 0 \qquad \text{as } k \to \infty,$$

showing that $A \cup B$ is in $\mathcal{M}_{\mathcal{F}}$. It remains to be seen that $A \setminus B$ is $\mathcal{M}_{\mathcal{F}}$. We now do this; this proof should give an idea of how to do Exercise 3.2.7(d). Again, since $\mathcal{E}$ is a ring, $A_k \setminus B_k$ is in $\mathcal{E}$. From

$$\begin{aligned}
S(A_k \setminus B_k, A \setminus B) &= S(A_k \cap B_k^c, A \cap B^c) \\
&\subseteq S(A_k, A) \cup S(B_k^c, B^c) \\
&= S(A_k, A) \cup S(B_k, B)
\end{aligned}$$

it follows that

$$\begin{aligned}
D(A_k \setminus B_k, A \setminus B) &= m^*\Big(S(A_k \setminus B_k, A \setminus B)\Big) \\
&\le m^*\Big(S(A_k, A)\Big) + m^*\Big(S(B_k, B)\Big) \\
&= D(A_k, A) + D(B_k, B).
\end{aligned}$$

This proves that $A \setminus B$ is $\mathcal{M}_{\mathcal{F}}$, as desired. □

Lemma 3.5. *Let $A \in \mathcal{M}$. Then $A \in \mathcal{M}_{\mathcal{F}}$ if and only if $m^*(A) < \infty$.*

PROOF. First, assume that $A \in \mathcal{M}_{\mathcal{F}}$. Then there exists a sequence of sets $A_k \in \mathcal{E}$ satisfying $D(A_k, A) \to 0$ as $k \to \infty$. We choose N to satisfy $D(A_N, A) < 1$. From Exercise 3.2.7(d) it follows that

$$D(A, \emptyset) \le D(A, A_N) + D(A_N, \emptyset),$$

or

$$m^*(A) \le D(A, A_N) + m^*(A_N) < 1 + m^*(A_N) < \infty.$$

To prove the converse we assume that $A \in \mathcal{M}$ and that $m^*(A) < \infty$. We aim to show that $A \in \mathcal{M}_{\mathcal{F}}$. Since $A \in \mathcal{M}$, we can write $A = \bigcup_{k=1}^{\infty} B_k$, where $B_k \in \mathcal{M}_{\mathcal{F}}$

for each k. Letting $A_1 = B_1$, and $A_k = B_k \setminus \big(\bigcup_{j=1}^{k-1} B_j\big)$ for $k \geq 2$, we note that $A_k \in \mathcal{M}_{\mathcal{F}}$ for each k, and we have rewritten A as the union $A = \bigcup_{k=1}^{\infty} A_k$ of disjoint sets $A_1, A_2, \ldots$.

Countable subadditivity of m^* (see Exercise 3.2.6) yields

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

We claim that this is actually an equality. To see this, note that for each N,

$$\bigcup_{k=1}^{N} A_k \subseteq A,$$

and so, by monotonicity,

$$m^*\Big(\bigcup_{k=1}^{N} A_k\Big) \leq m^*(A).$$

Lemma 3.3 asserts that m^* is additive when restricted to $\mathcal{M}_{\mathcal{F}}$. Therefore,

$$\sum_{k=1}^{N} m^*(A_k) = m^*\Big(\bigcup_{k=1}^{N} A_k\Big).$$

Since now

$$\sum_{k=1}^{N} m^*(A_k) \leq m^*(A)$$

for each N, we have shown that

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

We are assuming that $m^*(A) < \infty$, and hence the series on the left converges. Therefore, given $\epsilon > 0$, there exists an N such that $\sum_{k=N+1}^{\infty} m^*(A_k) < \epsilon$. Then

$$\begin{aligned} D\Big(A, \bigcup_{k=1}^{N} A_k\Big) &= m^*\Big(S\Big(A, \bigcup_{k=1}^{N} A_k\Big)\Big) \\ &= m^*\Big(A \setminus \bigcup_{k=1}^{N} A_k\Big) \\ &= m^*\Big(\bigcup_{k=N+1}^{\infty} A_k\Big) \leq \sum_{k=N+1}^{\infty} m^*(A_k) < \epsilon. \end{aligned}$$

Since ϵ was chosen arbitrarily, this proves that $A \in \mathcal{M}_{\mathcal{F}}$. □

Theorem 3.6. *$\mathcal{M}$ is a σ-ring, and m^* is countably additive on $\mathcal{M}$.*

PROOF. First, we prove that $\mathcal{M}$ is a σ-ring.

If $A_1, A_2, \ldots \in \mathcal{M}$, then their union can be seen to be in $\mathcal{M}$ via a standard diagonalization argument.

Let $A, B \in \mathcal{M}$. We can then write

$$A = \bigcup_{k=1}^{\infty} A_k, \quad \text{and} \quad B = \bigcup_{k=1}^{\infty} B_k,$$

for some $A_k, B_k \in \mathcal{M}_{\mathcal{F}}$, $k = 1, 2, \ldots$. The reader should check that the identity

$$A_k \cap B = \bigcup_{j=1}^{\infty} (A_k \cap B_j), \quad k = 1, 2 \ldots,$$

holds. From this it follows that $A_k \cap B \in \mathcal{M}$ for each k. Since

$$m^*(A_k \cap B) \leq m^*(A_k) < \infty,$$

Lemma 3.5 implies that $A_k \cap B \in \mathcal{M}_{\mathcal{F}}$ for each n. Lemma 3.4 then implies that

$$A_k \setminus B = A_k \setminus (A_k \cap B) \in \mathcal{M}_{\mathcal{F}}, \quad k = 1, 2 \ldots.$$

Finally we have our desired result, that

$$A \setminus B = \bigcup_{k=1}^{\infty} \Big(A_k \setminus B \Big) \in \mathcal{M}.$$

Second, we prove that m^* is countably additive on $\mathcal{M}$. We consider $A = \bigcup_{k=1}^{\infty} B_k$, where $B_k \in \mathcal{M}$ for each k (and so also $A \in \mathcal{M}$ by the first part of the theorem). Letting $A_1 = B_1$, and $A_k = B_k \setminus \left(\bigcup_{j=1}^{k-1} B_j \right)$ for $k \geq 2$, we can rewrite A as the union $A = \bigcup_{k=1}^{\infty} A_k$ of disjoint sets $A_1, A_2, \ldots$. In Exercise 3.2.6 you are asked to prove that m^* is countably subadditive, and therefore

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

On the other hand, for each positive integer N,

$$\bigcup_{k=1}^{N} A_k \subseteq A,$$

so the additivity of m^* on $\mathcal{M}_{\mathcal{F}}$ (Lemma 3.3) and the monotonicity of m^* on $2^{(\mathbb{R}^n)}$ together imply that

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

Since this holds for each positive integer N,

$$\sum_{k=1}^{\infty} m^*(A_k) = m^*(A),$$

as desired. $\square$

We now have rings $\mathcal{E} \subseteq \mathcal{M} \subseteq 2^{(\mathbb{R}^n)}$ and an outer measure m^* defined on $2^{(\mathbb{R}^n)}$. Further, m^* restricted to the σ-ring $\mathcal{M}$ is a measure. The elements of $\mathcal{M}$ are called the *Lebesgue measurable subsets of* $\mathbb{R}^n$. The restriction of m^* to $\mathcal{M}$ is called *Lebesgue measure*, and is (again) denoted by m. It is important to figure out which subsets of $\mathbb{R}^n$ are Lebesgue measurable. Exercises 3.2.9 and 3.2.10 give some answers. After doing that exercise you may well wonder whether there are any sets in $\mathbb{R}^n$ that are not Lebesgue measurable, and if there are, just how bizarre they must be. There are indeed such sets. A discussion of nonmeasurable sets is deferred to the Section 6.4. Because the Lebesgue measurable sets are hard to describe, people often choose to work with Lebesgue measure on the smaller σ-ring of all "Borel" sets. The Borel sets are defined, and briefly discussed, in the second example of Section 6 of this chapter.

FIGURE 3.1. Henri Lebesgue.

Henri Léon Lebesgue was born on June 28, 1875, in Beauvais, France (Figure 3.1). His father was a typographical worker, and his mother was an elementary-school teacher; both were intellectually motivated people. In 1897 Lebesgue graduated from the École Normale Supérieure in Paris and then worked for two years in their library. During these two years, he published his first four mathematical papers. His first paper gave a simpler proof of the Weierstrass approximation theorem (discussed in detail in Section 6.1).

Lebesgue can be said to have made two huge contributions to mathematics: He helped to sort out the correct definition of the term "function," and he developed complete, and to date the most successful, theories of measure and integration. For historical perspective, [93] is a good reference for the former contribution, and [61] is recommended for the latter. In his obituary of Lebesgue [29], J.C. Burkill concludes, "His work lay almost entirely in one field — the theory of real functions; in that field he is supreme."

Between 1899 and 1902 Lebesgue was teaching at the lycée in Nancy, and also working on his thesis. During these three years he published six papers. The last five of these were then incorporated to form his doctoral thesis. He received his Ph.D. from the Sorbonne in 1902. His dissertation is considered to be one of the best mathematics theses ever written. The first chapter develops his theory of measure; the second chapter develops his integral; the third chapter discusses length, area, and certain surfaces; the fourth chapter is on Plateau's problem about minimal surfaces. Many of the important properties of the Riemann integral were generalized by Lebesgue to his integral in the second chapter.

Probably the most notable exception to this is that what is now often referred to as the first version of fundamental theorem of calculus (that $\frac{d}{dx}\left(\int_a^x f(t)dt\right) = f(x)$ almost everywhere) does not appear. Lebesgue was aware that he wanted this statement and, in fact, was unable to prove it for his thesis. He was able to prove it later, and it appeared in print one year later [81]. Several other unresolved issues in his thesis were also resolved by Lebesgue himself during the two years following his thesis work.

At the end of the nineteenth century, only continuous functions could be dealt with in a satisfactory manner, and there was still much debate over what the definition of a function should be. By the end of the first decade of the twentieth century, the treatment of discontinuous functions was fully incorporated. This ten-year revolution, culminating in the modern theory of real functions, was led by Lebesgue. Lebesgue's ideas can be seen to be very strongly and most directly influenced by the works of René Baire, Emile Borel, and Camille Jordan.[1] Baire's work gave deep insights into the behavior of discontinuous functions, while the work of Borel and Jordan focused on measuring the size of sets. The ideas of Baire, Borel, and Jordan had, of course, interested others as well. In particular, the works of Giuseppe Vitali (1875–1932; Italy) and William Henry Young (1863–1942; England) should be noted in the context of the development of the measure and integral credited to Lebesgue (see [61]).

In 1904 Lebesgue published his book ***Leçons sur l'intégration et la recherche des fonctions primitives***. This book reached a large number of readers, and it did not take long for Lebesgue's integral to become the integral of choice for most practitioners. It was taught to undergraduates as early as 1914, at the Rice Institute (now Rice University, in Texas). The Lebesgue integral has had remarkable success in applications, and its staying power is really because of these applications. Lebesgue himself applied his integral to problems having to do with trigonometric series, problems that had arisen in Fourier's work. As discussed in the opening paragraphs of Chapter 3, it is Riesz who deserves much credit for drawing attention to the importance of Lebesgue's ideas by showing their value for solving problems in the new field of functional analysis. Indeed, if it were not for Riesz's applications of Lebesgue's ideas, functional analysis would not have developed as it did and might look very different today. And as we have seen, the field of probability would not be the same without the notion of the Lebesgue integral.

By 1922, Lebesgue had published dozens of papers on set theory, integration, measure, trigonometric series, polynomial approximation, topology, and geometry. Over the next twenty years he continued to write, but the focus of his papers shifted toward the expository, often treating historical, philosophical, or pedagogical topics and reflecting his great interest and strong views on teaching.

Henri Lebesgue died on July 26, 1941, in Paris.

---

[1] Borel we have already encountered. Jordan was a French mathematician who lived from 1838 to 1922; the Jordan canonical form in matrix theory and the Jordan curve theorem in topology are two results named for him. Baire was also French, and lived from 1874 to 1932. One of Baire's results is the subject of Section 6.2.

3.3 Measurable and Lebesgue Integrable Functions on Euclidean Space

We will be using Euclidean space $\mathbb{R}^n$ as our measure space X, the Lebesgue measurable sets $\mathcal{M}$ as our σ-ring $\mathcal{R}$, and Lebesgue measure m as our measure μ. Everything that we say in this and in the next section for the triple $(\mathbb{R}^n, \mathcal{M}, m)$ can be said for the more general measure space $(X, \mathcal{R}, \mu)$. Note that the use of the phrase "measure space" introduced in the last sentence is different from the prior usage, when it was used to refer to X alone.

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is called *measurable* if the set

$$\{x \,|\, f(x) > a\}$$

is measurable for each $a \in \mathbb{R}$.

Theorem 3.7. *The following are equivalent statements:*

(a) $\{x \,|\, f(x) > a\}$ *is measurable for every* $a \in \mathbb{R}$.
(b) $\{x \,|\, f(x) \geq a\}$ *is measurable for every* $a \in \mathbb{R}$.
(c) $\{x \,|\, f(x) < a\}$ *is measurable for every* $a \in \mathbb{R}$.
(d) $\{x \,|\, f(x) \leq a\}$ *is measurable for every* $a \in \mathbb{R}$.

Proof. Theorem 3.6 shows that $\mathcal{M}$ is a σ-ring, and hence $A \in \mathcal{M}$ if and only if $A^c \in \mathcal{M}$. From this, (a) $\Leftrightarrow$ (d) and (b) $\Leftrightarrow$ (c) follow immediately.

That (a) implies (b) follows from Theorem 3.6 and the equalities

$$\{x \,|\, f(x) \geq a\} = \bigcap_{k=1}^{\infty} \left\{x \,\Big|\, f(x) > a - \frac{1}{k}\right\} = \left(\bigcup_{k=1}^{\infty} \left\{x \,\Big|\, f(x) \leq a - \frac{1}{k}\right\}\right)^c.$$

That (b) implies (a) follows from Theorem 3.6 and the equality

$$\{x \,|\, f(x) > a\} = \bigcup_{k=1}^{\infty} \left\{x \,\Big|\, f(x) \geq a + \frac{1}{k}\right\}. \qquad \square$$

Theorem 3.8. *If f is measurable, then $|f|$ is measurable.*

Proof. Left as an Exercise 3.3.4. $\square$

Theorem 3.9. *If f and g are measurable, then so are $f + g$, fg, f_+, and f_-. If $\{f_k\}_{k=1}^{\infty}$ is a sequence of measurable functions, then the four functions*

$$\left(\inf f_k\right)(x) = \inf\{f_k(x) \,|\, 1 \leq k < \infty\},$$

$$\left(\sup f_k\right)(x) = \sup\{f_k(x) \,|\, 1 \leq k < \infty\},$$

$$\left(\liminf f_k\right)(x) = \sup_{j \geq 1} \left(\inf_{k \geq j} f_k(x)\right),$$

$$\left(\limsup f_k\right)(x) = \inf_{j \geq 1} \left(\sup_{k \geq j} f_k(x)\right),$$

are each measurable.

PROOF. First, we prove that $f+g$ is measurable. Observe that

$$(f+g)(x) < a \Leftrightarrow f(x) < a - g(x)$$

and that this is true if and only if there exists a rational number r such that

$$f(x) < r < a - g(x).$$

Therefore,

$$\{x \mid (f+g)(x) < a\} = \bigcup_{r\in\mathbb{Q}} \Big(\{x \mid f(x) < r\} \cap \{x \mid g(x) < a - r\}\Big).$$

Since the right-hand side belongs to $\mathcal{M}$, so does the set on the left.

Next, we next prove that fg is measurable for the special case when $f = g$. We then use the first part of the theorem, the special case, and the so-called polarization identity

$$fg = \frac{1}{4}(f+g)^2 - \frac{1}{4}(f-g)^2,$$

to get the general case. The case $f = g$ is taken care of by noticing that

$$\{x \mid (ff)(x) > a\} = \{x \mid f(x) > \sqrt{a}\} \cup \{x \mid f(x) < -\sqrt{a}\}.$$

We next prove that $\sup f_k$ is measurable. For each $a \in \mathbb{R}$,

$$\{x \mid (\sup f_k)(x) > a\} = \bigcup_{k=1}^{\infty} \{x \mid f_k(x) > a\}.$$

Since each set in the union on the right side of the equation is in $\mathcal{M}$, so is the set on the left side.

The proof that $\inf f_k$ is measurable is similar to the argument for the supremum. Then $\limsup f_k$ and $\liminf f_k$ are measurable from these (applying the argument twice in succession). The facts that f_- and f_+ are measurable follow from the proof for $\sup f_k$, since $f_+ = \max\{f, 0\} = \sup\{f, 0\}$ and $f_- = \max\{-f, 0\} = \sup\{-f, 0\}$. □

This last theorem can be interpreted as saying that the usual ways of combining functions preserve measurability. One way of combining functions is noticeably missing: composition. It is not the case that the composition of two measurable functions is again measurable. See, for example, [1] (page 57) or [70] (page 362) to see what can be said about the measurability of the composition of two functions.

A real-valued function with only a finite number of elements in its range is called a *simple function*. One type of simple function is the *characteristic function*, χ_E, of a set $E \subseteq \mathbb{R}^n$. This is defined by

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

Every simple function can be written as a finite linear combination of characteristic functions. Specifically, if the range of the simple function s is $\{c_1, \dots, c_N\}$, then

$$s(x) = \sum_{k=1}^{N} c_k \chi_{E_k}(x),$$

where $E_k = \{x : s(x) = c_k\}$. The function s is measurable if and only if each set E_k is in $\mathcal{M}$.

What might be more remarkable is that every function defined on $\mathbb{R}^n$ can be well approximated by simple functions. This is the thrust of the next theorem.

Theorem 3.10. *If f is a real-valued function defined on $\mathbb{R}^n$, then there exists a sequence $\{s_k\}_{k=1}^{\infty}$ of simple functions such that*

$$\lim_{k\to\infty} s_k(x) = f(x), \qquad \text{for every } x \in \mathbb{R}^n.$$

Further, if f is measurable, then the s_k's may be chosen to be measurable simple functions. Finally, if $f \geq 0$, then $\{s_k\}_{k=1}^{\infty}$ may be chosen to satisfy $s_1 \leq s_2 \leq \cdots$.

Proof. We first consider the case that $f \geq 0$. In the general case, we use that $f = f_+ - f_-$ and apply the construction below to each of f_+ and f_-.

Fix $f \geq 0$ and a positive integer k; we start by defining the simple function s_k. Define

$$F_k = \{x \,\big|\, f(x) \geq k\},$$

and sets

$$E_j^k = \left\{x \,\Big|\, \frac{j-1}{2^k} \leq f(x) < \frac{j}{2^k}\right\},$$

for each integer j, $1 \leq j \leq k2^k$. Then put

$$s_k(x) = k\chi_{F_k}(x) + \sum_{j=1}^{k2^k} \frac{j-1}{2^k} \chi_{E_j^k}(x).$$

It is left as an exercise to show that the sequence $\{s_k\}_{k=1}^{\infty}$ has all of the desired properties. □

Let $E \in \mathcal{M}$. For a measurable simple function $s(x) = \sum_{k=1}^{N} c_k \chi_{E_k}(x)$, we define the *Lebesgue integral of s over E* by

$$\int_E s\,dm = \sum_{k=1}^{N} c_k m(E \cap E_k).$$

For a measurable, nonnegative function f we define the *Lebesgue integral of f over E* by

$$\int_E f\,dm = \sup\left\{ \int_E s\,dm \,\Big|\, 0 \leq s \leq f,\ s \text{ simple}\right\}.$$

Note that $\int_E f\,dm$ may be infinite. Now let f be an arbitrary (not necessarily nonnegative) measurable function. We say that f is *integrable* if both

$$\int_E f_+ dm \quad \text{and} \quad \int_E f_- dm$$

are finite and define

$$\int_E f\,dm = \int_E f_+ dm - \int_E f_- dm.$$

Integrability is really a statement about *absolute* integrability, as will be seen in the exercises (Exercise 3.3.7). This sometimes causes confusion. The integral

$$\int_E f\,dm$$

is referred to as the *integral of f, with respect to the measure m, over E*. This terminology opens the door for integrating with respect to other measures. We will discuss other measures in the last section of this chapter. We let

$$\mathcal{L}(\mathbb{R}^n, m) \quad \text{or} \quad \mathcal{L}(\mathbb{R}^n)$$

denote the collection of all functions that are integrable with respect to Lebesgue measure m over $\mathbb{R}^n$. This collection forms a real linear space. This fact and other useful properties of the integral are listed in the following theorem.

Theorem 3.11. *The Lebesgue integral enjoys several properties.*

(a) *The integral is linear. That is,*

$$\int_E cf\,dm = c\int_E f\,dm \quad \text{and} \quad \int_E (f+g)dm = \int_E f\,dm + \int_E g\,dm$$

whenever $f, g \in \mathcal{L}(\mathbb{R}^n)$, $c \in \mathbb{R}$, and $E \in \mathcal{M}$.

(b) *The integral is monotone. That is,*

$$\int_E f\,dm \le \int_E g\,dm$$

whenever $f, g \in \mathcal{L}(\mathbb{R}^n)$, $f(x) \le g(x)$ for all $x \in E$.

(c) *For every $f \in \mathcal{L}(\mathbb{R}^n)$, we have $|f| \in \mathcal{L}(\mathbb{R}^n)$ and*

$$\left|\int_E f\,dm\right| \le \int_E |f|dm.$$

(d) *For every $f \in \mathcal{L}(\mathbb{R}^n)$, we have $\int_E f\,dm = 0$ for every measurable set E of measure zero. From this it follows that*

$$\int_A f\,dm = \int_B f\,dm$$

whenever A and B are measurable sets, $B \subseteq A$, and $m(A \setminus B) = 0$.

Proof. A proof of the first part of (a) is straightforward, as is a proof of (b); (c) follows from (b); (d) is straightforward as well. The second part of (a), which certainly should hold if there is any justice in the world, is more subtle than it appears; we will prove it using Lebesgue's monotone convergence theorem (Theorem 3.13). Proofs of the first part of (a), the second part of (a) for simple functions, (b), (c), and (d) are asked for in Exercise 3.3.8. □

In general, if a property P holds on a set A except possibly at each point of some subset of A that can be contained in a measurable set of measure zero, then we say that the property P holds on A *almost everywhere*, or *for almost all* $x \in A$. In light of (d), the phrase "$f(x) \leq g(x)$ for all $x \in E$" in (b) can be replaced by the phrase "$f(x) \leq g(x)$ for almost all $x \in E$." From now on in this chapter expressions such as $f \leq g$, $f = g$, etc., should be interpreted as $f(x) \leq g(x)$ for almost all x, $f(x) = g(x)$ for almost all x, etc.

We end this section with a further property of the integral. We will use this result to prove Lebesgue's monotone convergence theorem.

Theorem 3.12. *Assume that* $f \geq 0$ *is measurable, and that* $A_1, A_2, \ldots \in \mathcal{M}$ *are pairwise disjoint. Then,*

$$\int_{\bigcup_{k=1}^{\infty} A_k} f\,dm = \sum_{k=1}^{\infty} \Big(\int_{A_k} f\,dm \Big).$$

Proof. We first consider the case that $f = \chi_E$ for some $E \in \mathcal{M}$. By the countable additivity of m, we have

$$\int_A f\,dm = m(A \cap E) = m\Big(\bigcup_{k=1}^{\infty} (A_k \cap E) \Big)$$
$$= \sum_{k=1}^{\infty} m(A_k \cap E) = \sum_{k=1}^{\infty} \Big(\int_{A_k} f\,dm \Big).$$

The next case, that f is simple, follows from this first case by the way that we define the integral for simple functions.

Finally, we consider an arbitrary measurable $f \geq 0$. Let $\epsilon > 0$ and choose a simple function s such that $s \leq f$ and

$$\int_A f \leq \Big(\int_A s \Big) + \epsilon.$$

The right-hand side of this inequality is equal to

$$\sum_{k=1}^{\infty} \Big(\int_{A_k} s\,dm \Big) + \epsilon \leq \sum_{k=1}^{\infty} \Big(\int_{A_k} f\,dm \Big) + \epsilon.$$

Thus,

$$\int_A f \leq \sum_{k=1}^{\infty} \Big(\int_{A_k} f\,dm \Big).$$

The proof will be complete when we show that also

$$\int_A f \geq \sum_{k=1}^{\infty} \Big(\int_{A_k} f\,dm \Big).$$

We first consider two disjoint sets $A_1, A_2 \in \mathcal{M}$ and choose two simple functions s_1, s_2 such that $0 \leq s_k \leq f$ and

$$\int_{A_k} s_n dm \geq \Big(\int_{A_k} f\,dm \Big) - \frac{\epsilon}{2}, \quad k = 1, 2.$$

Set $s = \max\{s_1, s_2\}$. Then s is simple and $0 \leq s \leq f$. Also,

$$\int_{A_k} s\,dm \geq \Big(\int_{A_k} f\,dm \Big) - \frac{\epsilon}{2}, \quad k = 1, 2.$$

Therefore,

$$\int_{A_1} s\,dm + \int_{A_2} s\,dm \geq \int_{A_1} f\,dm + \int_{A_2} f\,dm - \epsilon.$$

Put $A = A_1 \cup A_2$. By the first part of the theorem,

$$\int_A s\,dm \geq \int_{A_1} f\,dm + \int_{A_2} f\,dm - \epsilon.$$

By monotonicity,

$$\int_A f\,dm \geq \int_A s\,dm,$$

and so

$$\int_A f\,dm \geq \int_{A_1} f\,dm + \int_{A_2} f\,dm - \epsilon.$$

Since ϵ was arbitrary, we have shown that

$$\int_A f\,dm \geq \int_{A_1} f\,dm + \int_{A_2} f\,dm.$$

We now use induction to show that

$$\int_A f\,dm \geq \sum_{k=1}^{N} \Big(\int_{A_k} f\,dm \Big), \quad N = 1, 2, \ldots.$$

Finally, we return to the general case $A = \bigcup_{k=1}^{\infty} A_k$. For any positive integer N, the preceding inductive argument shows that

$$\int_A f\,dm \geq \int_{A_1 \cup A_2 \cup \cdots \cup A_N} f\,dm \geq \sum_{k=1}^{N} \Big(\int_{A_k} f\,dm \Big).$$

Since this holds for each N, we are done. □

As you are asked to prove in Exercise 3.3.1, continuous functions are always measurable. Also, many continuous functions are integrable; for example, all continuous, bounded functions that vanish outside of some finite interval are integrable. How *discontinuous* can an element of $\mathcal{L}(\mathbb{R}^n)$ be? We know that we can take any continuous integrable function, alter its value on a set M of measure zero, and still have an integrable function. For example, the set M can be taken to be a countable dense subset of $\mathbb{R}^n$. Nonetheless, the continuous functions with "compact support" are dense in $\mathcal{L}(\mathbb{R}^n)$ (see Exercise 3.6.8 to see precisely what is meant by compact support).

3.4 The Convergence Theorems

In this section we shall see three theorems about how the Lebesgue integral behaves with respect to limit operations. The properties revealed in these theorems are what distinguish the Lebesgue integral from competitor integrals.

Theorem 3.13 (Lebesgue's Monotone Convergence Theorem)**.** *Suppose that $A \in \mathcal{M}$ and that $\{f_k\}_{k=1}^{\infty}$ is a sequence of measurable functions such that*

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots \quad \text{for almost all } x \in A.$$

Let f be defined to be the pointwise limit, $f(x) = \lim_{k\to\infty} f_k(x)$, of this sequence. Then f is integrable and

$$\lim_{k\to\infty}\left(\int_A f_k dm\right) = \int_A f\,dm.$$

PROOF. We have

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots \leq f(x) = \lim_{k\to\infty} f_k(x) \quad \text{for almost all } x \in A.$$

By monotonicity, we get

$$\int_A f_1 dm \leq \int_A f_2 dm \leq \cdots \leq \int_A f\,dm.$$

Thus $\{\int_A f_k dm\}_{k=1}^{\infty}$ is a bounded and nondecreasing sequence of real numbers, and hence must converge to some real number L. Note that $L \leq \int_A f\,dm$; we aim to show that $L \geq \int_A f\,dm$ also. To do this we choose a number $\delta \in (0, 1)$ and a simple function s satisfying $0 \leq s(x) \leq f(x)$ for almost all $x \in A$. Define

$$A_k = \{x \in A \mid f_k(x) \geq \delta s(x)\}.$$

Then

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

and

$$A = \bigcup_{k=1}^{\infty} A_k.$$

For each positive integer k, we have

$$L = \lim_{k\to\infty}\Big(\int_A f_k dm\Big) \geq \int_A f_k dm \geq \int_{A_k} f_k dm \geq \delta \int_{A_k} s dm.$$

Therefore,

$$L \geq \delta \cdot \lim_{k\to\infty}\Big(\int_{A_k} s dm\Big).$$

We claim that

$$\lim_{k\to\infty}\Big(\int_{A_k} s dm\Big) = \int_A s dm. \qquad (\dagger)$$

Given that this equality holds, we obtain

$$L \geq \delta \cdot \int_A s dm.$$

Since δ was arbitrary,

$$L \geq \int_A s dm.$$

Taking the supremum over all such simple functions now yields

$$L \geq \int_A f dm.$$

This is what we wanted to prove.

To see (†), put $E_1 = A_1$, and $E_k = A_k \setminus A_{k-1}$. Then

$$A = \bigcup_{k=1}^{\infty} E_k, \qquad A_k = \bigcup_{j=1}^{k} E_j,$$

and the E_k's are pairwise disjoint. Theorem 3.12 then implies that

$$\int_A f dm = \sum_{k=1}^{\infty}\Big(\int_{E_k} f dm\Big),$$

which, by definition, is equal to

$$\lim_{k\to\infty}\Big(\sum_{j=1}^{k}\Big(\int_{E_j} f dm\Big)\Big) = \lim_{k\to\infty}\Big(\int_{A_k} f dm\Big). \qquad \square$$

Before moving on to the next "convergence theorem" we fulfill our promise made in the previous section and use Lebesgue's monotone convergence theorem to prove the second part of Theorem 3.11(a). Specifically,

$$\int_E (f+g) dm = \int_E f dm + \int_E g dm$$

whenever $f, g \in \mathcal{L}(\mathbb{R}^n)$ and $E \in \mathcal{M}$. In Exercise 3.3.8(b) you are asked to prove the result in the case that f and g are simple functions. Since $\int_E f dm$ is defined by the difference $\int_E f_+ dm - \int_E f_- dm$, we may assume that f is nonnegative (almost everywhere). Likewise for g. We first appeal to Theorem 3.10 to get sequences of nonnegative, measurable, simple functions $\{s_k\}_{k=1}^{\infty}$ and $\{t_k\}_{k=1}^{\infty}$ satisfying

$$\lim_{k\to\infty} s_k(x) = f(x), \qquad \lim_{k\to\infty} t_k(x) = g(x)$$

almost everywhere. Combining the monotone convergence theorem and the result for simple functions from the exercises, we see that

$$\begin{aligned}\int_E (f+g)dm &= \lim_{k\to\infty} \int_E (s_k + t_k)dm \\ &= \lim_{k\to\infty} \int_E s_k dm + \lim_{k\to\infty} \int_E t_k dm \\ &= \int_E f dm + \int_E g dm.\end{aligned}$$

The next result was proved by Pierre Fatou (1878–1929; France) in his 1906 doctoral dissertation. Fatou was also an astronomer. He studied twin stars and proved a conjecture of Gauss's on planetary orbits.

Theorem 3.14 (Fatou's Lemma). *Assume that $A \in \mathcal{M}$. Let $\{f_k\}_{k=1}^{\infty}$ be a sequence of nonnegative measurable functions and let $f = \liminf_{k\to\infty} f_k$ on A. Then*

$$\int_A f dm \leq \liminf_{k\to\infty} \Big(\int_A f_k dm \Big).$$

PROOF. For each positive integer j, define a function g_j by

$$g_j(x) = \inf_{k\geq j} f_k(x),$$

and a number a_j by

$$a_j = \inf_{k\geq j} \Big(\int_A f_k dm \Big).$$

Theorem 3.9 shows that each g_j is measurable, and clearly $\sup_{j\geq 1} g_j = f$. Since

$$0 \leq g_1(x) \leq g_2(x) \leq \cdots,$$

we have that

$$\lim_{j\to\infty} g_j = \sup_{j\geq 1} g_j = f.$$

Since

$$0 \leq a_1(x) \leq a_2(x) \leq \cdots,$$

we have that

$$\lim_{j\to\infty} a_j = \sup_{j\geq 1} a_j = \liminf_{k\to\infty} \Big(\int_A f_k dm \Big).$$

Observe that $g_j(x) \leq f_k(x)$ for each pair of positive integers j, k with $k \geq j$. Thus,

$$\int_A g_j dm \leq a_j,$$

for each positive integer j. The monotone convergence theorem now implies that

$$\int_A f dm \leq \lim_{j\to\infty} \Big(\int_A g_j dm \Big) \leq \lim_{j\to\infty} a_j = \liminf_{k\to\infty} \Big(\int_A f_k dm \Big). \qquad \square$$

The following is one of the best results for telling us when we may conclude that

$$\lim_{k\to\infty} \Big(\int_A f_k dm \Big) = \int_A \Big(\lim_{k\to\infty} f_k \Big) dm.$$

Theorem 3.15 (Lebesgue's Dominated Convergence Theorem)**.** *Assume that $A \in \mathcal{M}$. Let $\{f_k\}_{k=1}^{\infty}$ be a sequence of measurable functions, and put $f(x) = \lim_{k\to\infty} f_k(x)$. Further, assume that there exists a function $g \in \mathcal{L}(\mathbb{R}^n)$ such that $|f_k(x)| \leq g(x)$ for almost all $x \in A$ and each positive integer k. Then we may conclude that*

$$\lim_{k\to\infty} \Big(\int_A f_k dm \Big) = \int_A f dm.$$

Proof. Begin by noticing that for each k, $(f_k)_+ \leq g$, and $(f_k)_- \leq g$ and thus each f_k is in $\mathcal{L}(\mathbb{R}^n)$.

We first want to see that $|f|$ is in $\mathcal{L}(\mathbb{R}^n)$. This follows from Fatou's lemma:

$$\begin{aligned}\int_A |f| dm = \int_A \Big(\lim_{k\to\infty} |f_k| \Big) dm &= \int_A \Big(\liminf_{k\to\infty} |f_k| \Big) dm \\ &\leq \liminf_{k\to\infty} \Big(\int_A |f_k| dm \Big) \leq \int_A g dm.\end{aligned}$$

Since each $f_k + g$ is a nonnegative function, Fatou's lemma shows that

$$\begin{aligned}\int_A f dm + \int_A g dm &= \int_A (f + g) dm \\ &= \int_A \Big(\liminf_{k\to\infty} (f_k + g) \Big) dm \leq \liminf_{k\to\infty} \Big(\int_A (f_k + g) dm \Big).\end{aligned}$$

Because the integral and the processes of taking infima and suprema are additive, the expression on the right becomes

$$\liminf_{k\to\infty} \Big(\int_A f_k dm \Big) + \int_A g dm.$$

Combining these yields

$$\int_A f dm \leq \liminf_{k\to\infty} \Big(\int_A f_k dm \Big).$$

Since each $g - f_k$ is a nonnegative function, we can repeat this argument and get

$$\int_A f\,dm \geq \limsup_{k\to\infty}\Big(\int_A f_k dm\Big).$$

Combining these last two inequalities yields the desired result. □

3.5 Comparison of the Lebesgue Integral with the Riemann Integral

Lebesgue developed his integral in an effort to perfect the integral of Riemann. The main goal of this section is to show that the Riemann integrable functions form a proper subcollection of the Lebesgue integrable functions. In this section we will give several results without including proofs of them. Proofs can be found in any of the books on integration mentioned in the bibliography. *The discussion in this section is limited to integration on* $\mathbb{R}$.

We begin with a brief review of the definition of the Riemann integral. We assume that the reader is familiar with the Riemann integral and its properties and include this material as a reminder and also to establish notation.

We consider a bounded, real-valued function f defined on the closed and bounded interval $[a, b]$. A collection of points $P = \{x_0, x_1, \ldots, x_n\}$ is called a *partition* of $[a, b]$ if

$$a = x_0 < x_1 < \cdots < x_n = b.$$

The length of the longest subinterval $[x_{k-1}, x_k]$ is called the *mesh* of the partition P. Set, for $k = 1, 2, \ldots, n$,

$$m_k = \inf\{f(x)\,\big|\,x \in [x_{k-1}, x_k]\} \qquad \text{and} \qquad M_k = \sup\{f(x)\,\big|\,x \in [x_{k-1}, x_k]\}.$$

The *lower Riemann sum* $L(f, P)$ of f corresponding to the partition P is given by

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1}),$$

and the *upper Riemann sum* $U(f, P)$ of f corresponding to the partition P is given by

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1}).$$

The *lower Riemann integral* of f is defined by

$$L(f) = \sup\{L(f, P)\,\big|\,P \text{ is a partition of } [a, b]\},$$

and the *upper Riemann integral* of f is defined by

$$U(f) = \inf\{U(f, P)\,\big|\,P \text{ is a partition of } [a, b]\}.$$

Finally, a bounded, real-valued function f defined on $[a, b]$ is called *Riemann integrable* if $L(f) = U(f)$. In this case, their common value is denoted by

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

In the preceding paragraph, that the sets used above to define the lower and upper integrals do indeed have upper and lower bounds, respectively, is something one must prove. There is one result about Riemann integrals that we will use in this section: If $\{P_n\}_{n=1}^{\infty}$ is any sequence of partitions of $[a, b]$ such that the meshes of the P_n's converge to zero as $n \to \infty$, then

$$\lim_{n\to\infty} L(f, P_n) = \lim_{n\to\infty} U(f, P_n) = \int_a^b f(x)dx.$$

The Riemann and Lebesgue integrals of a nonnegative real-valued function can be interpreted in terms of area, as you should recall. In a naive way, the difference between the two integrals (on $\mathbb{R}$) can be visualized by noting that the Riemann approach subdivides the domain of the integrand, while the Lebesgue approach subdivides the range.

Recall that Lebesgue was trying, among other things, to increase the number of integrable functions. Our next theorem shows that each Riemann integrable function is Lebesgue integrable (showing that Lebesgue's collection contains Riemann's collection). If we consider the characteristic function of the rational numbers in the interval $[0, 1]$, we get a function with $L(f, P) = 0$ and $U(f, P) = 1$ for each partition of the unit interval. Thus, we have a function that is not Riemann integrable. However, this function is Lebesgue integrable, as you are asked to prove in Exercise 3.3.3. This together with Lemma 3.16 shows that Lebesgue's collection is, in fact, larger than Riemann's. Lebesgue was successful in enlarging the class of integrable functions.

Theorem 3.16. *If f is Riemann integrable on $[a, b]$, then f is Lebesgue integrable on $[a, b]$ and*

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

(Recall that the integral on the left is the Lebesgue integral, and the integral on the right is the Riemann integral.)

Proof. For each positive integer n, partition $[a, b]$ into 2^n subintervals each of length $\frac{b-a}{2^n}$. Let P_n denote this partition and $a = x_0 < x_1 < \cdots < x_n = b$ denote the points of P_n. Define

$$g_n(x) = \sum_{k=1}^{2^n} m_k \chi_{[x_{k-1},x_k)}(x) \qquad \text{and} \qquad h_n(x) = \sum_{k=1}^{2^n} M_k \chi_{[x_{k-1},x_k)}(x).$$

Then $\{g_n\}_{n=1}^{\infty}$ is an increasing sequence, and $\{h_n\}_{n=1}^{\infty}$ is a decreasing sequence. Put

$$g(x) = \lim_{n\to\infty} g_n(x) \qquad \text{and} \qquad h(x) = \lim_{n\to\infty} h_n(x).$$

Then g and h are Lebesgue integrable functions that satisfy

$$g(x) \leq f(x) \leq h(x)$$

for almost all x in $[a, b]$. Also,

$$\int_{[a,b]} g\,dm = L(f, P_n) \qquad \text{and} \qquad \int_{[a,b]} h\,dm = U(f, P_n).$$

Now, $h_n(x) - g_n(x) \geq 0$ for almost all x and

$$\lim_{n\to\infty} \Big(h_n(x) - g_n(x) \Big) = h(x) - g(x).$$

Therefore,

$$\begin{aligned} 0 \leq \int_{[a,b]} (h-g)\,dm &= \lim_{n\to\infty} \Big(\int_{[a,b]} (h_n - g_n)\,dm \Big) \\ &= \lim_{n\to\infty} \int_{[a,b]} h_n\,dm - \lim_{n\to\infty} \int_{[a,b]} g_n\,dm \\ &= \lim_{n\to\infty} U(f, P_n) - \lim_{n\to\infty} L(f, P_n) = 0. \end{aligned}$$

The first equality follows from Lebesgue's monotone convergence theorem, and the last from the result about Riemann integrals referred to in the paragraph preceding this theorem. It now follows that $g = h$ almost everywhere. Thus f is Lebesgue integrable, and

$$\int_{[a,b]} f\,dm = \lim_{n\to\infty} \int_{[a,b]} g_n\,dm = \lim_{n\to\infty} L(f, P_n) = \int_a^b f(x)\,dx. \qquad \square$$

A Riemann integrable function must, by definition, be bounded. Which bounded functions are Riemann integrable? Different characterizations exist, and perhaps most notable is Riemann's own characterization: A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition P of $[a, b]$ such that $U(f, P) - L(f, P) < \epsilon$. Lebesgue gave a characterization in terms of measure. Specifically, a bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is continuous almost everywhere.

It would be remiss not to mention that a version of the fundamental theorem of calculus can be given for the Lebesgue integral. In it, modifications are made to allow for the possibility of bad behavior on a set of measure zero. If f is Lebesgue integrable on $[a, b]$ and we define F by

$$F(x) = \int_{[a,x]} f\,dm,$$

then we cannot conclude that $F'(x) = f(x)$ for every value of x in $[a, b]$ (nor even that F is differentiable everywhere), but we can conclude that F is differentiable almost everywhere and that $F'(x) = f(x)$ for almost every value of x in $[a, b]$.

Finally, let us consider once again the characteristic function of the rationals $\chi_{\mathbb{Q}}$. This function is not Riemann integrable on the interval [0, 1] but is the pointwise limit of the sequence of Riemann integrable functions

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{r_1, \ldots, r_n\}, \\ 0 & \text{otherwise,} \end{cases}$$

where $r_1, r_2, \ldots$ is an enumeration of the rational numbers in the unit interval. In general, we can conclude that the limit of a sequence of Riemann integrable functions is again Riemann integrable if the convergence of the sequence is uniform (though uniform convergence is not necessary). The convergence theorems of Lebesgue (given in the previous section) show that the requirement of uniform convergence may be greatly relaxed to pointwise convergence if other, less restrictive, requirements are imposed on the sequence of functions when the Lebesgue integral is used in place of the Riemann integral. The fact that pointwise limit and the Lebesgue integral may be interchanged is a key property that makes the Lebesgue integral more useful than Riemann's integral.

3.6 General Measures and the Lebesgue L^p-spaces: The Importance of Lebesgue's Ideas in Functional Analysis

At the beginning of the second section of this chapter, we alluded to arbitrary measure spaces $(X, \mathcal{R}, \mu)$. We now give a discussion of these. Recall that a measure space consists of three things:

(i) a nonempty set X;
(ii) a σ-ring $\mathcal{R}$ of subsets of X;
(iii) a function μ defined on $\mathcal{R}$ satisfying

  (a) $0 \leq \mu(A) \leq \infty$ for all $A \in \mathcal{R}$,
  (b) $\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n)$ whenever $A_1, A_2, \ldots \in \mathcal{R}$ satisfy $A_n \cap A_m = \emptyset$, $n \neq m$.

In the preceding sections we have constructed and studied Lebesgue measure on Euclidean space. This is certainly the most important example for our purposes. In this section we meet a few other examples of measure spaces and then introduce, for each measure space $(X, \mathcal{R}, \mu)$, the linear space $L^p(X, \mu)$. As stated at the beginning of Section 3, all the results of that section and of Section 4 that are proved for $(\mathbb{R}^n, \mathcal{M}, m)$ hold for any general measure space $(X, \mathcal{R}, \mu)$. We will use these generalizations freely in this section.

We first give a list of some examples of triples $(X, \mathcal{R}, \mu)$.

Example 1. Let $X = \mathbb{R}^n$, $\mathcal{R} = \mathcal{M}$, $\mu = m$; this is the example we have been considering.

Example 2. Let $X = \mathbb{R}^n$, $\mathcal{R} = \mathcal{B}$, $\mu = m$. Here, $\mathcal{B}$ is defined to be the smallest σ-ring containing all open subsets of $\mathbb{R}^n$. The elements of $\mathcal{B}$ are called the *Borel* *sets* of $\mathbb{R}^n$. Since all open sets are measurable, the Borel sets form a subring of the measurable sets. However, not all measurable sets are Borel sets, and so the two collections are different. To construct a measurable set that is not a Borel set, one can make a "Cantor-like" construction (see, for example, [71] page 110). The following, however, is true: If A is measurable, then A can be written as $(A \setminus B) \cup B$ for some Borel set $B \subseteq A$ satisfying $m(A \setminus B) = 0$. The Borel sets are often favored as the underlying ring because although the ring contains fewer sets, the elements of it can be described more readily than the Lebesgue measurable sets can be.

EXAMPLE 3. Let $X = \mathbb{R}^n$, $\mathcal{R} = \mathcal{M}$. To define the measure, we consider any nondecreasing, continuous function $f : \mathbb{R} \to \mathbb{R}$ and put

$$\mu([a, b]) = f(b) - f(a).$$

Then μ can be extended to all of $\mathcal{M}$ in the same way that Lebesgue measure was. Indeed, this measure μ reduces to Lebesgue measure in the case $f(x) = x$. (The scope of this example can be increased greatly.)

EXAMPLE 4. Let X be any set, $\mathcal{R} = 2^X$, and let μ be *counting measure*:

$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite,} \\ \infty & \text{if } A \text{ is infinite,} \end{cases}$$

where $|A|$ denotes the number of elements in A. This measure might seem a bit simplistic. It is, but it plays an important role in the L^p-theory.

We generate further examples by restricting the space X:

EXAMPLE 5. $X = [0, 1]$ (the unit interval), $\mathcal{R} = \{S \subseteq [0, 1] \mid S \in \mathcal{M}\}$, $\mu = m$.

EXAMPLE 6. Let X be any uncountable set,

$$\mathcal{R} = \{A \in 2^X \mid A \text{ is countable or } X \setminus A \text{ is countable}\},$$

and

$$\mu(A) = \begin{cases} 0 & \text{if } A \text{ is countable,} \\ 1 & \text{if } X \setminus A \text{ is countable.} \end{cases}$$

EXAMPLE 7. Let X be any finite or countable set and $\mathcal{R} = 2^X$. Write $X = \{x_1, x_2, \ldots\}$. Let p_i be a positive number corresponding to each x_i, and assume that $\sum p_i = 1$. Define μ by

$$\mu(A) = \sum_{x_i \in A} p_i.$$

Examples 5, 6, and 7 share the property that the measure of the entire space is 1. Any measure with this property is called a *probability measure*. Our first example (Example 5) of such a measure plays a critical role in abstract probability, as indicated in the first section of this chapter. Our second example (Example 6) is rather silly but, nonetheless, provides another example. Example 7 provides a model for discrete probability theory.

We begin by fixing a measure space $(X, \mathcal{R}, \mu)$ and a real number $1 \leq p < \infty$. We will consider $0 < p < 1$ in Exercise 3.6.2 and the important case $p = \infty$ later in this section. Notice that if $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is a measurable function, then $|f|^p$ is also measurable.

Define, for a measure space $(X, \mathcal{R}, \mu)$ and a real number $1 \leq p < \infty$, the *Lebesgue space* $L^p(X, \mu)$ (or just $L^p(\mu)$, or even just L^p if the measure is clear from context) to be the collection of all μ-measurable functions such that

$$\int_X |f|^p d\mu < \infty.$$

We define the *p-norm* of an element $f \in L^p(\mu)$ to be the number

$$\|f\|_p = \Big(\int_X |f|^p d\mu \Big)^{\frac{1}{p}}.$$

The theory of L^p-spaces was developed by F. Riesz in 1910 [105]. In that article he introduced these spaces for Lebesgue measure on measurable subsets of $\mathbb{R}^n$, proved the Hölder and Minkowski inequalities (see our Theorems 3.18 and 3.19) in this setting, and showed that the step functions are dense in these spaces (our Theorem 3.22). He also showed that these spaces are norm complete (our Theorem 3.21). For $p = 2$ this had already been shown by E. Fischer [42]. Riesz's 1910 paper was a remarkable achievement, and remains one of the most important papers ever published in the field.

We now make two critical remarks regarding these definitions. First, we will want to prove that $\|f\|_p$ defines a norm, and in particular that $\|f\|_p = 0$ if and only if "$f = 0$." The equality in quotation marks where we must be careful. We know that the integral of a function will be zero as long as the function is equal, almost everywhere, to zero. In fact, $L^p(\mu)$ really consists of equivalence classes of functions rather than of functions, where

$$f \sim g \text{ if and only if } f(x) = g(x) \text{ almost everywhere.}$$

We will rarely mention this distinction, but it is important, and you should do your best to understand this point. As our second remark about this definition we point out that we are interested in being able to consider *complex*valued functions. Until now however, integration has been discussed only for real-valued functions. A function $f : X \to \mathbb{C}$ is called measurable if both $\mathrm{re}(f)$ and $\mathrm{im}(f)$ are measurable real-valued functions. In this case, we define the integral of f by

$$\int_X f d\mu = \int_X \mathrm{re}(f) d\mu + i \int_X \mathrm{im}(f) d\mu.$$

We will also use $L^p(\mu)$ to include complex-valued functions.

In summary, $L^p(\mu)$ denotes the set of all equivalence classes of complex-valued functions f defined on X satisfying

$$\int_X |f|^p d\mu < \infty.$$

We write f to stand for the equivalence class of all functions that are equal to f almost everywhere.

Theorem 3.17. *For $1 \leq p < \infty$, $L^p(\mu)$ is a linear space.*

Proof. This is easy to see. It is trivial to show that $\lambda f \in L^p(\mu)$ whenever $f \in L^p(\mu)$ and $\alpha \in \mathbb{C}$. To see that $f + g \in L^p(\mu)$ whenever both f and g are in $L^p(\mu)$, we use the inequality

$$|f + g|^p \leq 2^p(|f|^p + |g|^p).$$ □

Notice that for any $1 < p < \infty$ there is a unique number q such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

If $p = 1$, we define $q = \infty$; if $p = \infty$, we define $q = 1$. This number q is sometimes called the *Hölder conjugate* of p. Note that the Hölder conjugate of 2 is itself, and that this is the only number that is its own Hölder conjugate.

Let y be a fixed nonnegative number, and $1 < p < \infty$. The maximum of the function $f(x) = xy - \frac{x^p}{p}$ occurs at $x = y^{\frac{1}{p-1}}$ and thus $f(x) \leq f(y^{\frac{1}{p-1}})$ for all nonnegative numbers x. This inequality can be rearranged to yield

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

for all nonnegative numbers x and y (Exercise 3.6.3). We use this to prove our next result.

Theorem 3.18 (Hölder's Inequality)**.** *Assume that $1 < p < \infty$ and $1 < q < \infty$ are Hölder conjugates, and that $f \in L^p$ and $g \in L^q$. Then $fg \in L^1$ and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Proof. If $f = 0$ or $g = 0$ (recall that we mean here that $f = 0$ almost everywhere or $g = 0$ almost everywhere), then the result is trivial. So, we assume that $\|f\|_p > 0$ and $\|g\|_q > 0$. The discussion preceding this theorem shows that

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{p(\|f\|_p)^p} + \frac{|g(x)|^q}{q(\|g\|_q)^q}.$$

Integrating both sides of this yields the desired result. □

The German mathematician Otto Ludwig Hölder (1859–1937) worked mostly in group theory. However, he did work in analysis on the convergence of Fourier series (see Chapter 4). He proved his inequality in 1884.

Theorem 3.19. *For $1 \le p < \infty$, $L^p(\mu)$ is a normed linear space, with norm given by*

$$\|f\|_p = \Big(\int_X |f|^p d\mu \Big)^{\frac{1}{p}}.$$

Proof. It is straightforward to see that $\|f\|_p \ge 0$ for all $f \in L^p(\mu)$, that equality holds if and only if $f = 0$ almost everywhere, and that $\|\lambda f\|_p = |\lambda| \|f\|_p$ for $\lambda \in \mathbb{C}$. We concentrate our efforts on verifying the triangle inequality:

$$\|f + g\|_p \le \|f\|_p + \|g\|_p, \quad \text{for } f, g \in L^p(\mu).$$

If $p = 1$, this follows from Theorem 3.11. In the case $1 < p < \infty$, the result is nontrivial and is called *Minkowski's inequality*. If $\|f + g\|_p = 0$, there is nothing to prove, so we assume that this is not the case. We first note that

$$\||f + g|^{p-1}\|_q = \Big(\|f + g\|_{(p-1)q}\Big)^{p-1} = \Big(\|f + g\|_p\Big)^{p-1}.$$

(This is left as Exercise 3.6.4.) Hölder's inequality then implies

$$\begin{aligned}
1 &= \frac{1}{(\|f + g\|_p)^p} \Big(\int_X |f + g|^p d\mu \Big) \\
&= \frac{1}{(\|f + g\|_p)^p} \Big(\int_X |f + g| \cdot |f + g|^{p-1} d\mu \Big) \\
&\le \frac{1}{(\|f + g\|_p)^p} \Big(\int_X |f| \cdot |f + g|^{p-1} d\mu + \int_X |g| \cdot |f + g|^{p-1} d\mu \Big) \\
&\le \frac{1}{(\|f + g\|_p)^p} \Big(\|f\|_p \cdot \||f + g|^{p-1}\|_q + \|g\|_p \cdot \||f + g|^{p-1}\|_q \Big) \\
&= \frac{1}{\|f + g\|_p} \Big(\|f\|_p + \|g\|_p \Big).
\end{aligned}$$

□

Let $(X, \mathcal{R}, \mu)$ be a measure space. A measurable function f is said to be *essentially bounded* if there exists a nonnegative real number M and a measurable set A of measure zero such that

$$|f(x)| \le M, \quad \text{for all } x \in X \setminus A.$$

Then $L^\infty(X, \mu)$ (or $L^\infty(X)$ or even just L^∞) is defined to be the set of all essentially bounded measurable functions. We define $\|f\|_\infty$ for these functions by

$$\|f\|_\infty = \inf\{M\},$$

where the infimum is taken over all M that provide a bound in the definition of f being essentially bounded.

It is straightforward to verify that L^∞ is a linear space, and that $\|\cdot\|_\infty$ satisfies the properties to make L^∞ into a normed linear space. The space L^∞ (for Lebesgue measure on an interval of $\mathbb{R}$) was introduced by Hugo Steinhaus (1887–1972; Poland) in [116]. We have read a bit about him in Banach's biography. Steinhaus made many contributions to probability, functional analysis, and game theory. In 1923, Steinhaus published the first truly mathematical treatment of coin tossing based on measure theory. He is also known for his very popular books *Mathematical Snapshots* and *One Hundred Problems*.

The next lemma is stated here for its use in proving the theorem that follows it. It is interesting in its own right, since it characterizes completeness in terms of absolute summability in norm. We use f to denote an arbitrary element of a normed linear space mostly because we are now focusing our attention on spaces whose elements are functions. However, it should be noted that the lemma applies to all normed linear spaces.

Lemma 3.20. *A normed linear space* $(X, \|\cdot\|)$ *is complete if and only if* $\sum_{j=1}^{\infty} f_j$ *converges (in norm) whenever* $\sum_{j=1}^{\infty} \|f_j\|$ *converges.*

PROOF. We start by assuming that X is complete, and consider a sequence $\{f_j\}_{j=1}^{\infty}$ in X such that $\sum_{j=1}^{\infty} \|f_j\|$ converges. Let $\epsilon > 0$. Since $\sum_{j=1}^{\infty} \|f_j\|$ converges, there exists N such that

$$\sum_{j=N}^{\infty} \|f_j\| < \epsilon.$$

Let s_n denote the nth partial sum of the series $\sum_{j=1}^{\infty} f_j$; that is $s_n = \sum_{j=1}^{n} f_j$. For $n \geq m \geq N$,

$$\|s_n - s_m\| = \left\| \sum_{j=m+1}^{n} f_j \right\| \leq \sum_{j=m+1}^{n} \|f_j\| \leq \sum_{j=m+1}^{\infty} \|f_j\| < \epsilon.$$

Since X is complete, $\{s_n\}_{n=1}^{\infty}$ converges.

To show the other direction we consider a Cauchy sequence $\{f_j\}_{j=1}^{\infty}$ in X. For each k there exists j_k such that

$$\|f_i - f_j\| < \frac{1}{2^k}, \quad i, j \geq j_k.$$

We may assume that $j_{k+1} > j_k$. This implies that $\{f_{j_k}\}_{k=1}^{\infty}$ is a subsequence of $\{f_j\}_{j=1}^{\infty}$. Set

$$g_1 = f_{j_1}, \text{ and } g_k = f_{j_k} - f_{j_{k-1}} \text{ for } k \geq 2.$$

Observe that

$$\sum_{k=1}^{l} \|g_k\| = \|g_1\| + \sum_{k=2}^{l} \|f_{j_k} - f_{j_{k-1}}\| < \|g_1\| + \sum_{k=2}^{l} \frac{1}{2^{k-1}} \leq \|g_1\| + 1.$$

Therefore $\{\sum_{k=1}^{l} \|g_k\|\}_{l=1}^{\infty}$ is a bounded, increasing sequence which thus converges. By hypothesis, $\sum_{k=1}^{\infty} g_k$ converges. Since $\sum_{k=1}^{n} g_k = f_{j_n}$, it follows that $\{f_{j_n}\}_{n=1}^{\infty}$ converges in X. Let $f \in X$ denote the limit of the subsequence $\{f_{j_n}\}_{n=1}^{\infty}$ of $\{f_j\}_{j=1}^{\infty}$. We will be done when we show that $\{f_j\}_{j=1}^{\infty}$ also converges to f. Let $\epsilon > 0$. Since $\{f_j\}_{j=1}^{\infty}$ is Cauchy, there exists N such that

$$\|f_i - f_j\| < \frac{\epsilon}{2}, \quad i, j \geq N.$$

Also, there is a K such that

$$\|f_{j_k} - f\| < \frac{\epsilon}{2}, \quad k \geq K.$$

Choose k such that $k \geq K$ and $j_k \geq N$. By the triangle inequality we have

$$\|f_j - f\| \leq \|f_j - f_{j_k}\| + \|f_{j_k} - f\| < \epsilon, \quad j \geq N,$$

completing the proof. □

We now present one of F. Riesz's most important results.

Theorem 3.21. *For $1 \leq p \leq \infty$, $L^p(\mu)$ is complete.*

PROOF. We first do the case $p = \infty$; this is the easiest part of the proof. Let $\{f_j\}_{j=1}^{\infty}$ be an arbitrary Cauchy sequence in $L^{\infty}(\mu)$. By Exercise 3.6.7 there exist measure zero sets $A_{m,n}$ and B_j, $j, m, n = 1, 2, \ldots$, such that

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_{\infty}$$

for all $x \notin A_{m,n}$, and

$$|f_j(x)| \leq \|f\|_{\infty}$$

for all $x \notin B_j$. Define A to be the union of these sets for $j, m, n = 1, 2, \ldots$. Then A has measure zero (Exercise 3.2.6). Define

$$f(x) = \begin{cases} 0 & \text{if} \quad x \in A, \\ \lim_{j\to\infty} f_j(x) & \text{if} \quad x \notin A. \end{cases}$$

Then f is measurable. Also, for each $x \notin A$ there exists a positive integer N_x such that

$$|f_n(x) - f(x)| < 1, \qquad n \geq N_x.$$

In particular,

$$|f_{N_x}(x) - f(x)| < 1, \quad x \notin A.$$

From this it follows that

$$|f(x)| < 1 + |f_{N_x}(x)| \leq 1 + \|f_{N_x}\|_{\infty}, \qquad x \notin A,$$

and hence that $f \in L^{\infty}$. Since a Cauchy sequence is bounded, there exists $M > 0$ such that $\|f_j\|_{\infty} \leq M$ for every $j = 1, 2, \ldots$. In particular, $\|f_{N_x}\|_{\infty} \leq M$. The last inequality now shows that $f \in L^{\infty}(\mu)$.

Now we want to show that our given Cauchy sequence actually converges to this element f (in $L^{\infty}(\mu)$; we already know that it converges pointwise almost everywhere, but this is a weaker assertion than we need). To this end, let $\epsilon > 0$. There exists a positive integer K such that $n, m \geq K$ imply

$$\|f_n - f_m\|_{\infty} < \epsilon.$$

Then

$$|f_n(x) - f_m(x)| < \epsilon$$

almost everywhere, and so

$$\lim_{m\to\infty} |f_n(x) - f_m(x)| \leq \epsilon,$$

for $n \geq K$ and $x \notin A$. This shows that

$$\|f - f_n\|_\infty \leq \epsilon,$$

for $n \geq K$, as desired.

We next tackle the other cases, for $1 \leq p < \infty$. The proof is not trivial, and the artillery required is somewhat substantial; we will use the lemma preceding this theorem, Fatou's lemma, Lebesgue's dominated convergence theorem, and Minkowski's inequality. Consider a Cauchy sequence $\{f_k\}_{k=1}^\infty$ in $L^p(\mu)$, $1 \leq p < \infty$, such that

$$\sum_{k=1}^\infty \|f_k\|_p = M < \infty.$$

By Lemma 3.20, it suffices to show that $\sum f_k$ converges (in norm), that is, that there exists a function $s \in L^p$ such that

$$\left\|\left(\sum_{k=1}^n f_k\right) - s\right\|_p \to 0, \quad \text{as} \quad n \to \infty.$$

We work on determining this s. Define, for each positive integer n,

$$g_n(x) = \sum_{k=1}^n |f_k(x)|.$$

Minkowski's inequality implies that

$$\|g_n\|_p \leq \sum_{k=1}^n \|f_k\|_p \leq M.$$

Therefore,

$$\int_X (g_n)^p d\mu \leq M^p.$$

For each $x \in X$, $\{g_n(x)\}_{n=1}^\infty$ is an increasing sequence of numbers in $\mathbb{R} \cup \{\infty\}$, and so there exists a number $g(x) \in \mathbb{R} \cup \{\infty\}$ to which the sequence $\{g_n(x)\}_{n=1}^\infty$ converges. The function g on X defined in this way is measurable, and Fatou's lemma asserts that

$$\int_X g^p d\mu \leq \liminf_{n\to\infty} \left(\int_X (g_n)^p d\mu\right) \leq M^p.$$

In particular, this shows that $g(x) < \infty$ almost everywhere. For each x such that $g(x)$ is finite, the series $\sum_{k=1}^\infty f_k(x)$ is an absolutely convergent series. Let

$$s(x) = \begin{cases} 0 & \text{if} \quad g(x) \text{ is infinite,} \\ \sum_{k=1}^\infty f_k(x) & \text{if} \quad g(x) \text{ is finite.} \end{cases}$$

This function is equal, almost everywhere, to the limit of the partial sums $s_n(x) = \sum_{k=1}^{n} f_k(x)$, and hence is itself measurable. Since

$$|s_n(x)| \leq g(x),$$

we have that

$$|s(x)| \leq g(x).$$

Thus, $s \in L^p$ and

$$|s_n(x) - s(x)|^p \leq \Big(|s_n(x)| + |s(x)|\Big)^p \leq 2^p\Big(g(x)\Big)^p.$$

We can now apply Lebesgue's dominated convergence theorem to get

$$\lim_{n\to\infty}\Big(\int_X (s_n - s)^p d\mu\Big) = 0.$$

In other words,

$$\lim_{n\to\infty}\Big(\|s_n - s\|_p\Big)^p = 0,$$

and hence

$$\lim_{n\to\infty}\|s_n - s\|_p = 0,$$

which is precisely what we wanted to prove. □

The most important L^p-spaces for us will be $L^p(X, \mu)$ where μ is either Lebesgue measure on some (not necessarily proper) subset of $\mathbb{R}^n$, or μ is counting measure on $X = \mathbb{N}$. In the case of counting measure, the L^p-space is denoted by $\ell^p(\mathbb{N})$, or ℓ^p (read "little ell p"), and *is* (the reader should come to grips with this assertion) the space of all sequences $\{x_n\}_{n=1}^{\infty}$ satisfying

$$\sum_{n=1}^{\infty} |x_n|^p < \infty,$$

with norm given by

$$\|\{x_n\}_{n=1}^{\infty}\|_p = \Big(\sum_{n=1}^{\infty} |x_n|^p\Big)^{\frac{1}{p}}.$$

Note that ℓ^∞ is the set of all bounded sequences, with

$$\|\{x_n\}_{n=1}^{\infty}\|_\infty = \sup\{|x_n| \,\big|\, n = 1, 2, \ldots\}.$$

(Recall the material of Section 1.2.)

Theorem 3.21 shows that all L^p-spaces, $1 \leq p \leq \infty$, are complete. The cases $1 \leq p < \infty$ are deeper than the case $p = \infty$, and it is harder to supply a proof that applies for all measures. There are, however, some specific measures for which there are easier proofs. It is instructive to see some of these as well, and for this reason an alternative proof of the completeness of ℓ^p, $1 \leq p < \infty$, is now given.

We consider a Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ (a sequence of sequences!) in ℓ^p, with the sequence x_n given by

$$x_n = \left(a_1^{(n)}, a_2^{(n)}, \dots\right).$$

For a fixed k, observe that

$$\left|a_k^{(n)} - a_k^{(m)}\right| \le \left(\sum_{i=1}^{\infty} \left|a_i^{(n)} - a_i^{(m)}\right|^p\right)^{\frac{1}{p}}.$$

This shows that for each fixed k, the sequence $\{a_k^{(n)}\}_{n=1}^{\infty}$ is a Cauchy sequence of real numbers. Therefore, $\{a_k^{(n)}\}_{n=1}^{\infty}$ converges; let

$$a_k = \lim_{n\to\infty} a_k^{(n)}.$$

We now show two things:

(i) $a = \{a_k\}_{k=1}^{\infty}$ is in ℓ^p;
(ii) $\lim_{n\to\infty} \|x_n - a\|_p = 0$.

Exercise 2.3.1 shows that there exists M such that

$$\|x_n\|_p \le M,$$

for all $n = 1, 2, \dots$. For any k, we thus have,

$$\left(\sum_{i=1}^{k} |a_i^{(n)}|^p\right)^{\frac{1}{p}} \le \|x_n\|_p \le M.$$

Letting $n \to \infty$ yields

$$\left(\sum_{i=1}^{k} |a_i|^p\right)^{\frac{1}{p}} \le M.$$

Since k is arbitrary, this shows that $a = \{a_k\}_{k=1}^{\infty}$ is in ℓ^p, and also that

$$\|a\|_p \le M.$$

To show (ii), let $\epsilon > 0$. Then there exists a positive integer N such that

$$\|x_n - x_m\|_p < \epsilon, \qquad n, m \ge N.$$

For any k, we thus have,

$$\left(\sum_{i=1}^{k} |a_i^{(n)} - a_i^{(m)}|^p\right)^{\frac{1}{p}} \le \|x_n - x_m\|_p < \epsilon, \qquad n, m \ge N.$$

If we now keep both n and k fixed, and let $m \to \infty$, we get

$$\left(\sum_{i=1}^{k} |a_i^{(n)} - a_i|^p\right)^{\frac{1}{p}} < \epsilon, \qquad n \ge N.$$

Since k is arbitrary, this shows that $\|x_n - a\|_p \leq \epsilon$ for $n \geq N$, which is equivalent to (ii).

Note that the outline for the proof in the special case of counting measure is exactly the same as for the general measure proof given in Theorem 3.21: Take a specially designed function (sequence), which is a "pointwise" limit of sorts of the given Cauchy sequence, and then

(i) prove that this specially designed function (sequence) is in fact an element of the space;
(ii) prove that the convergence is in fact in norm (and not just "pointwise").

The difficulty encountered in the proof in the general case comes in having to prove these two properties for *arbitrary* measures.

We end this chapter with a big theorem, proved (for Lebesgue measure on Euclidean space) by F. Riesz in the 1910 paper [105]. Recall the definition of a simple function. A simple function is called a *step function* if each of the sets E_k has finite measure.

Theorem 3.22. *The step functions are dense in* $L^p(\mu)$*, for each* $1 \leq p < \infty$.

PROOF. Let $0 \leq f \in L^p$. By Theorem 3.10 we can construct a sequence of simple functions, $\{s_n\}_{n=1}^{\infty}$, such that

$$0 \leq s_1 \leq s_2 \leq \cdots \leq f, \quad \lim_{n\to\infty} s_n(x) = f(x) \text{ almost everywhere.}$$

Each of these simple functions is, in fact, a step function. Furthermore,

$$(f - s_1)^p \geq (f - s_2)^p \geq \cdots \geq 0, \text{ and } \lim_{n\to\infty}((f - s_n)(x))^p = 0 \text{ almost everywhere.}$$

Lebesgue's dominated convergence theorem now tells us that

$$\|f - s_n\|_p = \lim_{n\to\infty}\left(\int_X |f - s_n|^p\right)^{\frac{1}{p}} = 0.$$

Since every element of L^p can be written as the difference of two nonnegative functions in L^p, the proof is done. □

We end this section by remarking that L^2 is a Hilbert space. What remains to be seen in this is that the norm comes from an inner product. This is easily seen, by defining

$$\langle f, g\rangle = \int_X f\overline{g}\,d\mu$$

for complex-valued functions. The Hilbert space L^2 will be discussed in great detail in the next chapter. One can also show that the norm on L^p, for $p \neq 2$, does not come from an inner product. A proof of this is outlined in Exercise 3.6.9.

Frigyes Riesz was born in Gyor, Austria–Hungary (in what is now Hungary) on January 22, 1880 (Figure 3.2). His father was a physician. His younger brother Marcel was also a distinguished mathematician.

Frigyes Riesz studied in Budapest and then went to Göttingen and Zurich before returning to Budapest, where he received his doctorate in 1902. His dissertation built on ideas of Fréchet, and made connections between Lebesgue's work on measure-theoretic notions and the work of Hilbert and his student Erhard Schmidt (1876–1959; Russia, now Estonia) on integral equations. Hilbert and Schmidt had been working with integral equations in which the functions were assumed continuous. Riesz, in this context, introduced the Lebesgue square integrable (L^2-) functions. He was also interested in knowing which sequences of real numbers could arise as the Fourier coefficients of some function. He answered this question, as did Fischer, and the result is known as the Riesz–Fischer theorem (see Section 4.2).

Over the next few years, and in an attempt to generalize the Riesz–Fischer theorem, Riesz introduced the L^p-spaces for $p > 1$, and the general theory of normed linear spaces. One of the most important results about the L^p-spaces is his Riesz representation theorem. This theorem completely describes all the continuous linear functionals (see Section 6.3) from L^p to $\mathbb{C}$. Riesz is often considered to be the "father" of abstract operator theory. Hilbert's eigenvalue problem for integral equations was dealt with quite effectively by Riesz in this more abstract setting. Riesz was able to obtain many results about the spectra of the integral operators associated with the integral equations of Hilbert.

FIGURE 3.2. Frigyes Riesz.

As alluded to in the preceding paragraph, Riesz also introduced the notion of a norm, but this idea did not come to fruition until Banach wrote down his axioms for a normed linear space in [10].

Frigyes Riesz made many important contributions to functional analysis, as well as to the mathematics profession as a whole. His ideas show great originality of thought, and aesthetic judgment in mathematical taste. He is one of the founders of the general theory of normed linear spaces and the operators acting on them. His theory of compact operators, which generalizes work of Fredholm, set the stage for future work on classes of operators. While he did so much on this abstract theory, Riesz was originally motivated by very concrete problems, and often returned to them in his work. Most of Riesz's work on operator theory in general, and spectral theory in particular, lies beyond the scope of this book. For a detailed historical account of Riesz's contributions, see [34] or [94].

Riesz was able to communicate about mathematics superbly. He wrote several books and many articles, and served as editor of the journal ***Acta Scientiarum Mathematicarum***. His book [107] is a classic that continues to serve as an excellent introduction to the subject.

Riesz died on February 28, 1956, in Budapest.

Exercises for Chapter 3

Section 3.1

3.1.1 Write out the details of the proof, using diagonalization, that $\mathcal{B}$ is uncountable.

3.1.2 Prove that $\mathcal{B}_T$ is countable, and use this result to give an alternative proof (as outlined in the text) that $\mathcal{B}$ is uncountable.

3.1.3 In this exercise you are asked to determine the set B_E for given events E.

- **(a)** Determine the set B_E if E is the event that in the first three tosses, exactly two heads are seen. What is the probability of this event occurring? Does your answer to the last question agree with what you think the Lebesgue measure of B_E should be?
- **(b)** Determine the set B_E if E is the event that in the first n tosses, exactly k heads are seen. What is the probability of this event occurring? Does your answer to the last question agree with what you think the Lebesgue measure of B_E should be? Explain.

3.1.4 Prove that $(0, 1] \setminus S$ is uncountable, where S is the set referred to in the strong law of large numbers. Hint: Consider the map from $(0, 1]$ to itself that maps the binary expression $\omega = .a_1a_2\ldots$ to $.a_1 11 a_2 11 a_3 11 a_4 \ldots$. Prove that this map is one-to-one and its image is contained in $(0, 1] \setminus S$.

Section 3.2

3.2.1 Let $\mathcal{R}$ be a σ-ring. Prove that $\bigcap_{n=1}^{\infty} A_n \in \mathcal{R}$ whenever $A_n \in \mathcal{R}$, $n = 1, 2, \ldots$. Hint: Verify, and use, that $\bigcap_{n=1}^{\infty} A_n = A_1 \setminus \bigcup_{n=1}^{\infty}(A_1 \setminus A_n)$.

3.2.2 Assume that μ is a nonnegative, additive function defined on a ring $\mathcal{R}$.

- **(a)** Prove that μ is monotone; that is, show that $\mu(A) \leq \mu(B)$ whenever $A, B \in \mathcal{R}$ and $A \subseteq B$.
- **(b)** Prove that μ is *finitely subadditive*; that is, show that

$$m\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \sum_{k=1}^{n} m(A_k)$$

whenever $A_1, A_2, \ldots \in \mathcal{R}$.

3.2.3 Assume that μ is a countably additive function defined on a ring $\mathcal{R}$, that $A_n \in \mathcal{R}$, $A \in \mathcal{R}$, that

$$A_1 \subseteq A_2 \subseteq \cdots,$$

and that

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

Prove that

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

(Hint: Put $B_1 = A_1$, and $B_n = A_n \setminus A_{n-1}$ for $n = 2, 3, \ldots$.)

3.2.4 Prove that $\mathcal{E}$ is a ring, but is not a σ-ring.

3.2.5 Prove Lemma 3.1. (Hint: Consider the case that A is an interval first, and then consider finite unions of disjoint intervals.)

3.2.6 Prove that m^* is countably subadditive.

3.2.7 This exercise is about the symmetric difference and distance functions S and D defined on 2^X.

(a) Let $A = [0, 4] \times (1, 10]$ and $B = (0, 1] \times [0, 2]$ in $\mathbb{R}^2$. Draw a picture of the set $S(A, B)$, and compute $D(A, B)$.

(b) Consider arbitrary subsets A and B of an arbitrary set X. Prove that $D(A, B) = D(B, A)$.

(c) Consider arbitrary subsets A and B of an arbitrary set X. Does $D(A, B) = 0$ necessarily imply that $A = B$? Either prove that it does, or give a counterexample to show that it does not.

(d) Consider arbitrary subsets A, B, C of an arbitrary set X. Prove that

$$S(A, C) \subseteq S(A, B) \cup S(B, C),$$

and deduce that

$$D(A, C) \le D(A, B) + D(B, C).$$

(e) Consider arbitrary subsets A_1, A_2, B_1, B_2 of an arbitrary set X. Prove

$$S(A_1 \cup A_2, B_1 \cup B_2) \subseteq S(A_1, B_1) \cup S(A_2, B_2),$$

and deduce that

$$D(A_1 \cup A_2, B_1 \cup B_2) \le D(A_1, B_1) + D(A_2, B_2).$$

3.2.8 If you have studied some abstract algebra, you may know a different use of the term “ring” (the definition is given, incidentally, at the beginning of Section 6.6). In this exercise, the term “ring” refers to the algebraic notion. Prove that $2^{(\mathbb{R}^n)}$ becomes a commutative ring with “multiplication” of two sets taken to be their intersection, and with “addition” of two sets taken to be their symmetric difference.

3.2.9 **(a)** Prove that all open subsets of $\mathbb{R}^n$ are in $\mathcal{M}$.

(b) Prove that all closed subsets of $\mathbb{R}^n$ are in $\mathcal{M}$.
(c) Prove that all countable unions and intersections of open and closed subsets of $\mathbb{R}^n$ are in $\mathcal{M}$.

3.2.10 Prove that the Cantor set is in $\mathcal{M}$ and that it has Lebesgue measure zero.

Section 3.3

3.3.1 Prove that every continuous function is measurable.

3.3.2 Give an example of a function f such that f is not measurable but $|f|$ is measurable.

3.3.3 Which characteristic functions are (Lebesgue) integrable on $\mathbb{R}$? Is the characteristic function of the rational numbers integrable on the unit interval? If so, what is value of this integral?

3.3.4 Prove Theorem 3.8. (Hint: One approach is to notice that

$$\{x \,\big|\, |f(x)| > a\} = \{x \,\big|\, f(x) > a\} \cup \{x \,\big|\, f(x) < -a\}.)$$

3.3.5 Fill in the details of the proof of Theorem 3.9.

3.3.6 Complete the proof of Theorem 3.10.

3.3.7 Prove that $f \in \mathcal{L}(\mathbb{R}^n)$ if and only if $\int_{\mathbb{R}^n} |f| dm < \infty$.

3.3.8 Supply proofs for the missing parts of Theorem 3.11.

(a) Prove the first part of part (a) of Theorem 3.11.
(b) Prove the second part of part (a) of Theorem 3.11, for simple functions. (The general case appears after the proof of the monotone convergence theorem.)
(c) Prove part (b) of Theorem 3.11.
(d) Prove part (c) of Theorem 3.11.
(e) Prove part (d) of Theorem 3.11.

3.3.9 You have read about the phrase "almost everywhere" in the text. In particular, we say that two measurable functions are "equal almost everywhere" if the set of points where they differ has measure zero.

(a) Prove that this relation is an equivalence relation on the set of integrable functions.
(b) Prove that f and g are equal almost everywhere if and only if

$$\int_E f dm = \int_E g dm$$

for every measurable set E.

(c) Prove that $f = 0$ almost everywhere if $\int_E f dm = 0$ for every $E \in \mathcal{M}$.

Section 3.4

3.4.1 Give an example to show that strict inequality can hold in Fatou's lemma.

3.4.2 Give an example to show that without the existence of the function g in the dominated convergence theorem, the conclusion may fail.

Section 3.6

3.6.1 Prove that the triples $(X, \mathcal{R}, \mu)$ given in Examples 4 and 6 at the beginning of Section 6 are, in fact, measure spaces. For the measure in Example 3, show that

$$\mu\Big(\bigcup_{n=1}^{\infty}[a_n, b_n]\Big) = \sum_{n=1}^{\infty}\mu([a_n, b_n])$$

whenever $n \neq m$ implies $[a_n, b_n] \cap [a_m, b_m] = \emptyset$.

3.6.2 For $0 < p < 1$, we can define the L^p-space and $\|\cdot\|_p$ in the same way that we did for $1 \leq p < \infty$. Prove, by giving a suitable example, that $\|\cdot\|_p$ does not satisfy the triangle inequality, and hence is not a norm.

3.6.3 Prove, as outlined in the text, that

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q},$$

for $x, y \geq 0$ and Hölder conjugates p and q with $1 < p < \infty$.

3.6.4 Prove that

$$\||f+g|^{p-1}\|_q = \Big(\|f+g\|_{(p-1)q}\Big)^{p-1} = \Big(\|f+g\|_p\Big)^{p-1},$$

for $f, g \in L^p(\mu)$ and $1 < p < \infty$. (This equality is used in the proof of Theorem 3.19.)

3.6.5 Prove that L^∞, with norm $\|\cdot\|_\infty$, is a normed linear space.

3.6.6 In this exercise you will investigate relations between the various L^p-spaces.

(a) Let $1 \leq p < q \leq \infty$. Consider Lebesgue measure on $\mathbb{R}^n$. Construct examples to show that neither $L^p \subseteq L^q$ nor $L^q \subseteq L^p$ holds.

(b) Next, suppose that $1 \leq p < r < q < \infty$. Show that $L^p \cap L^q \subseteq L^r$. (This is for any measure space $(X, \mathcal{R}, \mu)$.)

(c) Now assume that X is a *finite* measure space, i.e., that X is measurable and that $\mu(X) < \infty$. Prove that $L^q \subseteq L^p$ for $1 \leq p < q < \infty$, and give an example to show that this is a proper inclusion. Now prove that

$$\|f\|_1 \leq \|f\|_p \leq \|f\|_q \leq \|f\|_\infty, \quad 1 \leq p < q < \infty,$$

whenever μ is a probability measure.

(d) Prove that $\ell^p \subseteq \ell^q$ for $1 \leq p < q \leq \infty$, and give an example to show that this is a proper inclusion. Now prove that

$$\|f\|_\infty \leq \|f\|_q \leq \|f\|_p \leq \|f\|_1, \quad 1 \leq p < q < \infty.$$

3.6.7 Show that for each $f \in L^\infty$, $|f(x)| \leq \|f\|_\infty$ almost everywhere.

3.6.8 Let I be an interval in $\mathbb{R}$.

(a) Assume that $1 \leq p < \infty$. Prove that $C(I)$ is dense in $L^p(m)$ whenever I is closed and bounded.

(b) Now drop the assumption that I is closed and bounded. We certainly cannot expect $C(I)$ to be dense in $L^p(m)$, since it is not even contained in it. We define the "continuous functions with compact support" on $X \subseteq \mathbb{R}$ to be

$$C_c(X) = \{f \mid f \in C(X) \text{ and } \overline{\{x \in X \mid f(x) \neq 0\}} \text{ is compact}\}.$$

The set $\overline{\{x \in X \mid f(x) \neq 0\}}$ is often called the "support" of f. Again, assume that $1 \leq p < \infty$. Prove that $C_c(I)$ is dense in $L^p(m)$. (Note that this result subsumes the result of (a).)

(c) Prove that $C_c(I)$ is not dense in $L^\infty(m)$.

(The results of Exercise 8 give an alternative proof of Theorem 4.7, and give a way to define L^p that is independent of measure theory. See the paragraph following the proof of Theorem 4.7.)

3.6.9 The point of this exercise is to show that the norm on L^p, for $p \neq 2$, does not come from an inner product. We start by considering $L^p([-1, 1])$, with Lebesgue measure. Set $f(x) = 1 + x$ and $g(x) = 1 - x$.

(a) Show that

$$\|f\|_p^p = \frac{2^{p+1}}{p+1} = \|g\|_p^p,$$
$$\|f + g\|_p^p = 2^{p+1},$$
$$\|f - g\|_p^p = \frac{2^{p+1}}{p+1}.$$

(b) Using part (a), show that the parallelogram equality asserts that

$$(p+1)^{\frac{2}{p}} = 3.$$

Verify that this equality holds for $p = 2$.

(c) Prove that the parallelogram equality does not hold for values $p \neq 2$. Hint: Show that the function $(p + 1)^{\frac{2}{p}} - 3$ is a strictly decreasing function of $p \geq 1$ and thus takes on the value zero for at most one value of p.

(d) Modify the functions given in part (a) to prove the result for $L^p(I)$ where I is any interval, bounded or unbounded, in $\mathbb{R}$.

Springer

http://www.springer.com/978-0-387-95224-6

Beginning Functional Analysis
Saxe, K.
2002, XI, 197 p., Hardcover
ISBN: 978-0-387-95224-6