Chapter
3

Integration

To remedy deficiencies of Riemann integration that were discussed in Section 1B, in the last chapter we developed measure theory as an extension of the notion of the length of an interval. Having proved the fundamental results about measures, we are now ready to use measures to develop integration with respect to a measure. As we will see, this new method of integration fixes many of the problems with Riemann integration.

Statue in Milan of Italian mathematician Maria Gaetana Agnesi, who in 1748 published one of the first calculus textbooks. A translation of her book into English was published in 1801. In this chapter, we will develop a method of integration more powerful than methods contemplated by the pioneers of calculus.

3A Integration with Respect to a Measure

Integration of Nonnegative Functions

We will first define the integral of a nonnegative function with respect to a measure. Then by writing a real-valued function as the difference of two nonnegative functions, we will define the integral of a real-valued function with respect to a measure. We begin this process with the following definition.

3.1 Definition ***$\mathcal{S}$-partition***

Suppose $\mathcal{S}$ is a σ-algebra on a set X. An *$\mathcal{S}$-partition* of X is a finite collection $A_1, \ldots, A_m$ of disjoint sets in $\mathcal{S}$ such that $A_1 \cup \cdots \cup A_m = X$.

We adopt the convention that $0 \cdot \infty$ and $\infty \cdot 0$ should both be interpreted to be 0.

The next definition should remind you of the definition of the lower Riemann sum (see 1.3). However, now we are working with an arbitrary measure and thus X need not be a subset of $\mathbf{R}$. More importantly, even in the case when X is a closed interval $[a, b]$ in $\mathbf{R}$ and μ is Lebesgue measure on the Borel subsets of $[a, b]$, the sets $A_1, \ldots, A_m$ in the definition below do not need to be subintervals of $[a, b]$ as they do for the lower Riemann sum—they need only be Borel sets.

3.2 Definition ***lower Lebesgue sum***

Suppose that $(X, \mathcal{S}, \mu)$ is a measure space, $f \colon X \to [0, \infty]$ is an $\mathcal{S}$-measurable function, and P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. The *lower Lebesgue sum* $\mathcal{L}(f, P)$ is defined by

$$\mathcal{L}(f, P) = \sum_{j=1}^{m} \mu(A_j) \inf_{x \in A_j} f(x).$$

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. We will denote the integral of an $\mathcal{S}$-measurable function f with respect to μ by $\int f \, d\mu$. Our basic requirements for an integral are that we want $\int \chi_E \, d\mu$ to equal $\mu(E)$ for all $E \in \mathcal{S}$, and we want integration to be a linear function. As we will see, the following definition satisfies both of those requirements (although this is not obvious). Think about why the following definition is reasonable in terms of the integral equaling the area under the graph of the function (in the special case of Lebesgue measure on an interval of $\mathbf{R}$).

3.3 Definition ***integral of a nonnegative function***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f \colon X \to [0, \infty]$ is an $\mathcal{S}$-measurable function. The *integral* of f with respect to μ, denoted $\int f \, d\mu$, is defined by

$$\int f \, d\mu = \sup\{\mathcal{L}(f, P) : P \text{ is an } \mathcal{S}\text{-partition of } X\}.$$

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f\colon X \to [0, \infty]$ is an $\mathcal{S}$-measurable function. Each $\mathcal{S}$-partition $A_1, \dots, A_m$ of X leads to an approximation of f from below by the $\mathcal{S}$-measurable simple function $\sum_{j=1}^{m} \bigl(\inf_{x \in A_j} f(x)\bigr) \chi_{A_j}$. This suggests that

$$\sum_{j=1}^{m} \mu(A_j) \inf_{x \in A_j} f(x)$$

should be an approximation from below of our intuitive notion of $\int f \, d\mu$. Taking the supremum of these approximations leads to our definition of $\int f \, d\mu$.

Let's begin getting more comfortable with the definition of the integral of a nonnegative function with respect to a measure by proving the following result, which gives our first example of evaluating such an integral.

3.4 ***integral of a characteristic function***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E \in \mathcal{S}$. Then

$$\int \chi_E \, d\mu = \mu(E).$$

Proof If P is the $\mathcal{S}$-partition of X consisting of E and its complement $X \setminus E$, then clearly $\mathcal{L}(\chi_E, P) = \mu(E)$. Thus $\int \chi_E \, d\mu \geq \mu(E)$.

To prove the inequality in the other direction, suppose P is an $\mathcal{S}$-partition $A_1, \dots, A_m$ of X. Then $\inf_{x \in A_j} \chi_E(x)$ equals 1 if $A_j \subset E$ and equals 0 otherwise. Thus

$$\begin{aligned} \mathcal{L}(\chi_E, P) &= \sum_{\{j : A_j \subset E\}} \mu(A_j) \\ &= \mu\Bigl(\bigcup_{\{j : A_j \subset E\}} A_j\Bigr) \\ &\leq \mu(E). \end{aligned}$$

Thus $\int \chi_E \, d\mu \leq \mu(E)$, completing the proof. ∎

The symbol d in the expression $\int f \, d\mu$ has no meaning, serving only to separate f from μ. Because the d in $\int f \, d\mu$ does not represent another object, some mathematicians prefer typesetting an upright d *in this situation, producing $\int f \, \mathrm{d}\mu$. However, the upright* d *looks jarring to some readers who are accustomed to italicized symbols. This book takes the compromise position of using slanted d instead of math-mode italicized d in integrals.*

3.5 Example *integrals of $\chi_{\mathbf{Q}}$ and $\chi_{[0,1] \setminus \mathbf{Q}}$*

Suppose λ is Lebesgue measure on $\mathbf{R}$. As a special case of the result above, we have $\int \chi_{\mathbf{Q}} \, d\lambda = 0$ (because $|\mathbf{Q}| = 0$). Recall that $\chi_{\mathbf{Q} \cap [0,1]}$ is not Riemann integrable on $[0,1]$. Thus even at this early stage in our development of integration with respect to a measure, we have fixed one of the deficiencies of Riemann integration.

Note also that 3.4 implies that $\int \chi_{[0,1] \setminus \mathbf{Q}} \, d\lambda = 1$ (because $|[0,1] \setminus \mathbf{Q}| = 1$), which is what we want. In contrast, the lower Riemann integral of $\chi_{[0,1] \setminus \mathbf{Q}}$ on $[0,1]$ equals 0, which is not what we want.

3.6 Example *integration with respect to counting measure is summation*

Suppose μ is counting measure on $\mathbf{Z}^+$ and $b_1, b_2, \ldots$ is a sequence of nonnegative numbers. Think of b as the function from $\mathbf{Z}^+$ to $[0, \infty)$ defined by $b(k) = b_k$. Then

$$\int b\, d\mu = \sum_{k=1}^{\infty} b_k,$$

as you should verify.

Integration with respect to a measure can be called *Lebesgue integration*. The next result shows that Lebesgue integration behaves as expected on simple functions represented as linear combinations of characteristic functions of disjoint sets.

3.7 ***integral of a simple function***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E_1, \ldots, E_n$ are disjoint sets in $\mathcal{S}$, and $c_1, \ldots, c_n \in [0, \infty]$. Then

$$\int \Big(\sum_{k=1}^{n} c_k \chi_{E_k}\Big)\, d\mu = \sum_{k=1}^{n} c_k \mu(E_k).$$

Proof Without loss of generality, we can assume that $E_1, \ldots, E_n$ is an $\mathcal{S}$-partition of X (by replacing n by $n+1$ and setting $E_{n+1} = X \setminus (E_1 \cup \ldots \cup E_n)$ and $c_{n+1} = 0$).

If P is the $\mathcal{S}$-partition $E_1, \ldots, E_n$ of X, then $\mathcal{L}(f, P) = \sum_{k=1}^{n} c_k \mu(E_k)$. Thus $\int \big(\sum_{k=1}^{n} c_k \chi_{E_k}\big)\, d\mu \geq \sum_{k=1}^{n} c_k \mu(E_k)$.

To prove the inequality in the other direction, suppose that P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. Then

$$\begin{aligned}
\mathcal{L}\Big(\sum_{k=1}^{n} c_k \chi_{E_k}, P\Big) &= \sum_{j=1}^{m} \mu(A_j) \min_{\{i : A_j \cap E_i \neq \varnothing\}} c_i \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n} \mu(A_j \cap E_k) \min_{\{i : A_j \cap E_i \neq \varnothing\}} c_i \\
&\leq \sum_{j=1}^{m} \sum_{k=1}^{n} \mu(A_j \cap E_k) c_k \\
&= \sum_{k=1}^{n} c_k \sum_{j=1}^{m} \mu(A_j \cap E_k) \\
&= \sum_{k=1}^{n} c_k \mu(E_k).
\end{aligned}$$

The inequality above implies that $\int \big(\sum_{k=1}^{n} c_k \chi_{E_k}\big)\, d\mu \leq \sum_{k=1}^{n} c_k \mu(E_k)$, completing the proof. ∎

The next easy result gives an unsurprising property of integrals.

3.8 *Integration is order preserving*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g \colon X \to [0, \infty]$ are $\mathcal{S}$-measurable functions such that $f(x) \leq g(x)$ for all $x \in X$. Then $\int f \, d\mu \leq \int g \, d\mu$.

Proof Suppose P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. Then

$$\inf_{x \in A_j} f(x) \leq \inf_{x \in A_j} g(x)$$

for each $j = 1, \ldots, m$. Thus $\mathcal{L}(f, P) \leq \mathcal{L}(g, P)$. Hence $\int f \, d\mu \leq \int g \, d\mu$. ■

Monotone Convergence Theorem

For the proof of the Monotone Convergence Theorem, we will need to use the following mild restatement of the definition of the integral of a nonnegative function.

3.9 *Integrals via simple functions*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f \colon X \to [0, \infty]$ is $\mathcal{S}$-measurable. Then

3.10 $$\int f \, d\mu = \sup\Big\{ \sum_{j=1}^{m} c_j \mu(A_j) : A_1, \ldots, A_m \text{ are disjoint sets in } \mathcal{S},\ c_1, \ldots, c_m \in [0, \infty), \text{ and } f(x) \geq \sum_{j=1}^{m} c_j \chi_{A_j}(x) \text{ for every } x \in X \Big\}.$$

Proof First note that the left side of 3.10 is bigger than or equal to the right side by 3.7 and 3.8.

To prove that the right side of 3.10 is bigger than or equal to the left side, first assume that $\inf_{x \in A} f(x) < \infty$ for every $A \in \mathcal{S}$ with $\mu(A) > 0$. Then for P an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X, take $c_j = \inf_{x \in A_j} f(x)$, which shows that $\mathcal{L}(f, P)$ is in the set on the right side of 3.10. Thus the definition of $\int f \, d\mu$ shows that the right side of 3.10 is bigger than or equal to the left side.

The only remaining case to consider is when there exists a set $A \in \mathcal{S}$ such that $\mu(A) > 0$ and $\inf_{x \in A} f(x) = \infty$ [which implies that $f(x) = \infty$ for all $x \in A$]. In this case, for arbitrary $t \in (0, \infty)$ we can take $m = 1$, $A_1 = A$, and $c_1 = t$. These choices show that the right side of 3.10 is at least $t\mu(A)$. Because t is an arbitrary positive number, this shows that the right side of 3.10 equals ∞, which of course is greater than or equal to the left side, completing the proof. ■

The next result allows us to interchange limits and integrals in certain circumstances. We will see more theorems of this nature in the next section.

3.11 ***Monotone Convergence Theorem***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $0 \leq f_1 \leq f_2 \leq \cdots$ is an increasing sequence of $\mathcal{S}$-measurable functions. Define $f\colon X \to [0, \infty]$ by

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

Proof The function f is $\mathcal{S}$-measurable by 2.52.

Because $f_k(x) \leq f(x)$ for every $x \in X$, we have $\int f_k \, d\mu \leq \int f \, d\mu$ for each $k \in \mathbf{Z}^+$ (by 3.8). Thus $\lim_{k\to\infty} \int f_k \, d\mu \leq \int f \, d\mu$.

To prove the inequality in the other direction, suppose $A_1, \dots, A_m$ are disjoint sets in $\mathcal{S}$ and $c_1, \dots, c_m \in [0, \infty)$ are such that

3.12 $$f(x) \geq \sum_{j=1}^{m} c_j \chi_{A_j}(x) \quad \text{for every } x \in X.$$

Let $t \in (0, 1)$. For $k \in \mathbf{Z}^+$, let

$$E_k = \Big\{ x \in X : f_k(x) \geq t \sum_{j=1}^{m} c_j \chi_{A_j}(x) \Big\}.$$

Then $E_1 \subset E_2 \subset \dots$ is an increasing sequence of sets in $\mathcal{S}$ whose union equals X. Thus $\lim_{k\to\infty} \mu(A_j \cap E_k) = \mu(A_j)$ for each $j \in \{1, \dots, m\}$ (by 2.58).

If $k \in \mathbf{Z}^+$, then

$$f_k(x) \geq \sum_{j=1}^{m} t c_j \chi_{A_j \cap E_k}(x)$$

for every $x \in X$. Thus (by 3.9)

$$\int f_k \, d\mu \geq t \sum_{j=1}^{m} c_j \mu(A_j \cap E_k).$$

Taking the limit as $k \to \infty$ of both sides of the inequality above gives

$$\lim_{k\to\infty} \int f_k \, d\mu \geq t \sum_{j=1}^{m} c_j \mu(A_j).$$

Now taking the limit as t increases to 1 shows that

$$\lim_{k\to\infty} \int f_k \, d\mu \geq \sum_{j=1}^{m} c_j \mu(A_j).$$

Taking the supremum of the inequality above over all $\mathcal{S}$-partitions $A_1, \dots, A_m$ of X and all $c_1, \dots, c_m \in [0, \infty)$ satisfying 3.12 shows (using 3.9) that we have $\lim_{k\to\infty} \int f_k \, d\mu \geq \int f \, d\mu$, completing the proof. ∎

The proof that the integral is additive will use the Monotone Convergence Theorem and our next result. The representation of a simple function $h\colon X \to [0,\infty]$ in the form $\sum_{k=1}^{n} c_k \chi_{E_k}$ is not unique. Requiring the numbers $c_1, \dots, c_n$ to be distinct and $E_1, \dots, E_n$ to be nonempty and disjoint with $E_1 \cup \cdots \cup E_n = X$ does produce what is called the *standard representation* of a simple function [take $E_k = h^{-1}(\{c_k\})$, where $c_1, \dots, c_n$ are the distinct values of h]. The following lemma shows that all representations (including representations with sets that are not disjoint) of a simple measurable function give the same sum that we expect from integration.

3.13 *Integral-type sums for simple functions*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. Suppose $a_1, \dots, a_m, b_1, \dots, b_n \in [0,\infty]$ and $A_1, \dots, A_m, B_1, \dots, B_n \in \mathcal{S}$ are such that $\sum_{j=1}^{m} a_j \chi_{A_j} = \sum_{k=1}^{n} b_k \chi_{B_k}$. Then

$$\sum_{j=1}^{m} a_j \mu(A_j) = \sum_{k=1}^{n} b_k \mu(B_k).$$

Proof We assume $A_1 \cup \cdots \cup A_m = X$ (otherwise add the term $0\chi_{X \setminus (A_1 \cup \cdots \cup A_m)}$).

Suppose A_1 and A_2 are not disjoint. Then we can write

3.14 $$a_1\chi_{A_1} + a_2\chi_{A_2} = a_1\chi_{A_1 \setminus A_2} + a_2\chi_{A_2 \setminus A_1} + (a_1 + a_2)\chi_{A_1 \cap A_2},$$

where the three sets appearing on the right side of the equation above are disjoint.

Now $A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$ and $A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2)$; each of these unions is a disjoint union. Thus $\mu(A_1) = \mu(A_1 \setminus A_2) + \mu(A_1 \cap A_2)$ and $\mu(A_2) = \mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2)$. Hence

$$a_1\mu(A_1) + a_2\mu(A_2) = a_1\mu(A_1 \setminus A_2) + a_2\mu(A_2 \setminus A_1) + (a_1 + a_2)\mu(A_1 \cap A_2).$$

The equation above, in conjunction with 3.14, shows that if we replace the two sets A_1, A_2 by the three disjoint sets $A_1 \setminus A_2, A_2 \setminus A_1, A_1 \cap A_2$ and make the appropriate adjustments to the coefficients $a_1, \dots, a_m$, then the value of the sum $\sum_{j=1}^{m} a_j \mu(A_j)$ is unchanged (although m has increased by 1).

Repeating this process with all pairs of subsets among $A_1, \dots, A_m$ that are not disjoint after each step, in a finite number of steps we can convert the initial list $A_1, \dots, A_m$ into a disjoint list of subsets without changing the value of $\sum_{j=1}^{m} a_j \mu(A_j)$.

The next step is to make the numbers $a_1, \dots, a_m$ distinct. This is done by replacing the sets corresponding to each a_j by the union of those sets, and using finite additivity of the measure μ to show that the value of the sum $\sum_{j=1}^{m} a_j \mu(A_j)$ does not change.

Finally, drop any terms for which $A_j = \varnothing$, getting the standard representation for a simple function. We have now shown that the original value of $\sum_{j=1}^{m} a_j \mu(A_j)$ is equal to the value if we use the standard representation of the simple function $\sum_{j=1}^{m} a_j \chi_{A_j}$. The same procedure can be used with the representation $\sum_{k=1}^{n} b_k \chi_{B_k}$ to show that $\sum_{k=1}^{n} b_k \mu(\chi_{B_k})$ equals what we would get with the standard representation. Thus the equality of the functions $\sum_{j=1}^{m} a_j \chi_{A_j}$ and $\sum_{k=1}^{n} b_k \chi_{B_k}$ implies the equality $\sum_{j=1}^{m} a_j \mu(A_j) = \sum_{k=1}^{n} b_k \mu(B_k)$. ∎

If we had already proved that integration is linear, then we could quickly get the conclusion of the previous result by integrating both sides of the equation $\sum_{j=1}^{m} a_j \chi_{A_j} = \sum_{k=1}^{n} b_k \chi_{B_k}$ with respect to μ. However, we need the previous result to prove the next result, which is used in our proof that integration is linear.

Now we can show that our definition of integration does the right thing with simple measurable functions that might not be expressed in the standard representation. The result below differs from 3.7 mainly because the sets $E_1, \dots, E_n$ in the result below are not required to be disjoint. Like the previous result, the next result would follow immediately from the linearity of integration if that property had already been proved.

3.15 *Integral of a linear combination of characteristic functions*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E_1, \dots, E_n \in \mathcal{S}$, and $c_1, \dots, c_n \in [0, \infty]$. Then

$$\int \Big(\sum_{j=1}^{n} c_k \chi_{E_j} \Big) \, d\mu = \sum_{j=1}^{n} c_k \mu(E_j).$$

Proof The desired result follows from writing the simple function $\sum_{k=1}^{n} c_k \chi_{E_k}$ in the standard representation for a simple function and then using 3.7 and 3.13. ∎

Now we can prove that integration is additive on nonnegative functions.

3.16 *Additivity of integration*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g \colon X \to [0, \infty]$ are $\mathcal{S}$-measurable functions. Then

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

Proof The desired result holds for simple nonnegative $\mathcal{S}$-measurable functions (by 3.15). Thus we approximate by such functions.

Specifically, let $f_1, f_2, \dots$ and $g_1, g_2, \dots$ be increasing sequences of simple nonnegative $\mathcal{S}$-measurable functions such that

$$\lim_{k \to \infty} f_k(x) = f(x) \quad \text{and} \quad \lim_{k \to \infty} g_k(x) = g(x)$$

for all $x \in X$ (see 2.82 for the existence of such increasing sequences). Then

$$\begin{aligned} \int (f + g) \, d\mu &= \lim_{k \to \infty} \int (f_k + g_k) \, d\mu \\ &= \lim_{k \to \infty} \int f_k \, d\mu + \lim_{k \to \infty} \int g_k \, d\mu \\ &= \int f \, d\mu + \int g \, d\mu, \end{aligned}$$

where the first and third equalities follow from the Monotone Convergence Theorem and the second equality holds by 3.15. ∎

The lower Riemann integral is not additive, even for bounded nonnegative measurable functions. For example, if $f = \chi_{\mathbf{Q} \cap [0,1]}$ and $g = \chi_{[0,1] \setminus \mathbf{Q}}$, then

$$L(f, [0,1]) = 0 \quad \text{and} \quad L(g, [0,1]) = 0 \quad \text{but} \quad L(f+g, [0,1]) = 1.$$

In contrast, if λ is Lebesgue measure on the Borel subsets of $[0,1]$, then

$$\int f \, d\lambda = 0 \quad \text{and} \quad \int g \, d\lambda = 1 \quad \text{and} \quad \int (f+g) \, d\lambda = 1.$$

More generally, we have just proved that $\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$ for every measure μ and for all nonnegative measurable functions f and g. Recall that integration with respect to a measure is defined via lower Lebesgue sums in a similar fashion to the definition of the lower Riemann integral via lower Riemann sums (with the big exception of allowing measurable sets instead of just intervals in the partitions). However, we have just seen that the integral with respect to a measure (which could have been called the lower Lebesgue integral) has considerably nicer behavior (additivity!) than the lower Riemann integral.

Integration of Real-Valued Functions

The following definition gives us a standard way to write an arbitrary real-valued function as the difference of two nonnegative functions.

3.17 Definition f^+; f^-

Suppose $f \colon X \to [-\infty, \infty]$ is a function. Define functions f^+ and f^- from X to $[0, \infty]$ by

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{if } f(x) < 0 \end{cases} \quad \text{and} \quad f^-(x) = \begin{cases} 0 & \text{if } f(x) \geq 0, \\ -f(x) & \text{if } f(x) < 0. \end{cases}$$

Note that if $f \colon X \to [-\infty, \infty]$ is a function, then

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

The decomposition above allows us to extend our definition of integration to functions that take on negative as well as positive values.

3.18 Definition *integral of a real-valued function*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f \colon X \to [-\infty, \infty]$ is a measurable function such that at least one of $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ is finite. The *integral* of f with respect to μ, denoted $\int f \, d\mu$, is defined by

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

If $f \geq 0$, then $f^+ = f$ and $f^- = 0$; thus this definition is consistent with the previous definition of the integral of a nonnegative function.

The condition $\int |f| \, d\mu < \infty$ is equivalent to the condition $\int f^+ \, d\mu < \infty$ and $\int f^- \, d\mu < \infty$ (because $|f| = f^+ + f^-$).

3.19 Example *a function whose integral is not defined*

Suppose λ is Lebesgue measure on $\mathbf{R}$ and $f \colon \mathbf{R} \to \mathbf{R}$ is the function defined by

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Then $\int f \, d\lambda$ is not defined because $\int f^+ \, d\lambda = \infty$ and $\int f^- \, d\lambda = \infty$.

The next result says that the integral of a number times a function is exactly what we expect.

3.20 *Integration is homogeneous*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f \colon X \to [-\infty, \infty]$ is a function such that $\int f \, d\mu$ is defined. If $c \in \mathbf{R}$, then

$$\int cf \, d\mu = c \int f \, d\mu.$$

Proof First consider the case where f is a nonnegative function and $c \geq 0$. If P is an $\mathcal{S}$-partition of X, then clearly $\mathcal{L}(cf, P) = c\mathcal{L}(f, P)$. Thus $\int cf \, d\mu = c \int f \, d\mu$.

Now consider the general case where f takes values in $[-\infty, \infty]$. Suppose $c \geq 0$. Then

$$\begin{aligned} \int cf \, d\mu &= \int (cf)^+ \, d\mu - \int (cf)^- \, d\mu \\ &= \int cf^+ \, d\mu - \int cf^- \, d\mu \\ &= c\Big(\int f^+ \, d\mu - \int f^- \, d\mu\Big) \\ &= c \int f \, d\mu, \end{aligned}$$

where the third line follows from the first paragraph of this proof.

Finally, now suppose $c < 0$ (still assuming that f takes values in $[-\infty, \infty]$). Then $-c > 0$ and

$$\begin{aligned} \int cf \, d\mu &= \int (cf)^+ \, d\mu - \int (cf)^- \, d\mu \\ &= \int (-c)f^- \, d\mu - \int (-c)f^+ \, d\mu \\ &= (-c)\Big(\int f^- \, d\mu - \int f^+ \, d\mu\Big) \\ &= c \int f \, d\mu, \end{aligned}$$

completing the proof. ∎

Now we prove that integration with respect to a measure has the additive property required for a good theory of integration.

3.21 *Additivity of integration*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g \colon X \to (-\infty, \infty)$ are $\mathcal{S}$-measurable functions such that $\int |f| \, d\mu < \infty$ and $\int |g| \, d\mu < \infty$. Then

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

Proof Clearly

$$\begin{aligned}(f+g)^+ - (f+g)^- &= f+g \\ &= f^+ - f^- + g^+ - g^-.\end{aligned}$$

Thus

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Both sides of the equation above are sums of nonnegative functions. Thus integrating both sides with respect to μ and using 3.16 gives

$$\int (f+g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int (f+g)^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu.$$

Rearranging the equation above gives

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

where the left side is not of the form $\infty - \infty$ because $(f+g)^+ \leq f^+ + g^+$ and $(f+g)^- \leq f^- + g^-$. The equation above can be rewritten as

$$\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu,$$

completing the proof. ∎

The next result resembles 3.8, but now the functions are allowed to be real valued.

3.22 *Integration is order preserving*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g \colon X \to \mathbf{R}$ are $\mathcal{S}$-measurable functions such that $\int f \, d\mu$ and $\int g \, d\mu$ are defined. Suppose also that $f(x) \leq g(x)$ for all $x \in X$. Then $\int f \, d\mu \leq \int g \, d\mu$.

Proof Case where $\int f \, d\mu = \pm\infty$ or $\int g \, d\mu = \pm\infty$ are left to the reader. Thus we assume that $\int |f| \, d\mu < \infty$ and $\int |g| \, d\mu < \infty$.

The additivity (3.21) and homogeneity (3.20 with $c = -1$) of integration imply that

$$\int g \, d\mu - \int f \, d\mu = \int (g-f) \, d\mu.$$

The last integral is nonnegative because $g(x) - f(x) \geq 0$ for all $x \in X$. ∎

The inequality in the next result receives frequent use.

3.23 *Absolute value of integral $\leq$ integral of absolute value*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f\colon X \to [-\infty, \infty]$ is a function such that $\int f\, d\mu$ is defined. Then

$$\left| \int f\, d\mu \right| \leq \int |f|\, d\mu.$$

Proof Because $\int f\, d\mu$ is defined, f is an $\mathcal{S}$-measurable function and at least one of $\int f^+\, d\mu$ and $\int f^-\, d\mu$ is finite. Thus

$$\begin{aligned} \left| \int f\, d\mu \right| &= \left| \int f^+\, d\mu - \int f^-\, d\mu \right| \\ &\leq \int f^+\, d\mu + \int f^-\, d\mu \\ &= \int (f^+ + f^-)\, d\mu \\ &= \int |f|\, d\mu, \end{aligned}$$

as desired. ■

EXERCISES 3A

1 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f\colon X \to [0, \infty]$ is an $\mathcal{S}$-measurable function such that $\int f\, d\mu < \infty$. Explain why

$$\inf_{x \in E} f(x) = 0$$

for each set $E \in \mathcal{S}$ with $\mu(E) = 0$.

2 Suppose X is a set, $\mathcal{S}$ is a σ-algebra on X, and $c \in X$. Define the Dirac measure δ_c on $(X, \mathcal{S})$ by

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

Prove that if $f\colon X \to [0, \infty]$ is $\mathcal{S}$-measurable, then $\int f\, d\delta_c = f(c)$. [*Careful: $\{c\}$ may not be in $\mathcal{S}$.*]

3 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f\colon X \to [0, \infty]$ is an $\mathcal{S}$-measurable function. Prove that

$$\int f\, d\mu > 0 \text{ if and only if } \mu(\{x \in X : f(x) > 0\}) > 0.$$

4 Give an example of a Borel measurable function $f\colon [0,1] \to (0,\infty)$ such that $L(f,[0,1]) = 0$.
[*Recall that $L(f,[0,1])$ denotes the lower Riemann integral, which was defined in Section 1A. If λ is Lebesgue measure on $[0,1]$, then the previous exercise states that $\int f\,d\lambda > 0$ for this function f, which is what we expect of a positive function. Thus even though both $L(f,[0,1])$ and $\int f\,d\lambda$ are defined by taking the supremum of approximations from below, Lebesgue measure captures the right behavior for this function f and the lower Riemann integral does not.*]

5 Verify the assertion that integration with respect to counting measure is summation (Example 3.6).

6 Suppose $(X,\mathcal{S},\mu)$ is a measure space, $f\colon X \to [0,\infty]$ is $\mathcal{S}$-measurable, and P and P' are $\mathcal{S}$-partitions of X such that every set in P' is contained some set in P. Prove that $\mathcal{L}(f,P) \le \mathcal{L}(f,P')$.

7 Suppose X is a set, $\mathcal{S}$ is the σ-algebra of all subsets of X, and $w\colon X \to [0,\infty]$ is a function. Define a measure μ on $(X,\mathcal{S})$ by

$$\mu(E) = \sum_{x\in E} w(x)$$

for $E \subset X$. Prove that if $f\colon X \to [0,\infty]$ is a function, then

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

where the infinite sums above are defined as the supremum of all sums over finite subsets of X.

8 Suppose λ denotes Lebesgue measure on $\mathbf{R}$. Given an example of a sequence $f_1, f_2, \ldots$ of simple Borel measurable functions from $\mathbf{R}$ to $[0,\infty)$ such that $\lim_{k\to\infty} f_k(x) = 0$ for every $x \in \mathbf{R}$ but $\lim_{k\to\infty} \int f_k\,d\lambda = 1$.

9 Suppose $(X,\mathcal{S},\mu)$ is a measure space and $f\colon X \to [0,\infty]$ is an $\mathcal{S}$-measurable function. Define $\mu_f\colon \mathcal{S} \to [0,\infty]$ by

$$\mu_f(A) = \int \chi_A f\,d\mu$$

for $A \in \mathcal{S}$. Prove that μ_f is a measure on $(X,\mathcal{S})$.

10 Suppose $(X,\mathcal{S},\mu)$ is a measure space and $f_1, f_2, \ldots$ is a sequence of nonnegative $\mathcal{S}$-measurable functions. Define $f\colon X \to [0,\infty]$ by $f(x) = \sum_{k=1}^{\infty} f_k(x)$. Prove that

$$\int f\,d\mu = \sum_{k=1}^{\infty} \int f_k\,d\mu.$$

11 Give an example to show that the Monotone Convergence Theorem (3.11) can fail if the hypothesis that $f_1, f_2, \ldots$ are nonnegative functions is dropped.

12 Give an example to show that the Monotone Convergence Theorem fails if the hypothesis of an increasing sequence of functions is replaced by a hypothesis of a decreasing sequence of functions.
[*This exercise shows that the Monotone Convergence Theorem should be called the Increasing Convergence Theorem. However, see Exercise 16.*]

13 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f_1, f_2, \dots$ are $\mathcal{S}$-measurable functions from X to $\mathbf{R}$ such that $\sum_{k=1}^{\infty} \int |f_k| \, d\mu < \infty$. Prove that $\lim_{k\to\infty} f(x) = 0$ for almost all $x \in X$.

14 Give an example of a sequence $x_1, x_2, \dots$ of real numbers such that

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_k \text{ exists in } \mathbf{R},$$

but $\int x \, d\mu$ is not defined, where μ is counting measure on $\mathbf{Z}^+$ and x is the function from $\mathbf{Z}^+$ to $\mathbf{R}$ defined by $x(k) = x_k$.

For $x_1, x_2, \dots$ a sequence in $[-\infty, \infty]$, define $\liminf_{k\to\infty} x_k$ by

$$\liminf_{k\to\infty} x_k = \lim_{k\to\infty} \inf\{x_k, x_{k+1}, \dots\}.$$

Note that $\inf\{x_k, x_{k+1}, \dots\}$ is an increasing function of k; thus the limit above on the right exists in $[-\infty, \infty]$.

15 Suppose that $(X, \mathcal{S}, \mu)$ is a measure space and $f_1, f_2, \dots$ is a sequence of nonnegative $\mathcal{S}$-measurable functions on X. Define a function $f \colon X \to [0, \infty]$ by $f(x) = \liminf_{k\to\infty} f_k(x)$. Prove that

$$\int f \, d\mu \leq \liminf_{k\to\infty} \int f_k \, d\mu.$$

[*The result above is called Fatou's Lemma. Some textbooks prove Fatou's Lemma and then use it to prove the Monotone Convergence Theorem. Here we are taking the reverse approach—you should be able to use the Monotone Convergence Theorem to give a clean proof of Fatou's Lemma.*]

16 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f_1, f_2, \dots$ is a monotone (meaning either increasing or decreasing) sequence of $\mathcal{S}$-measurable functions. Define $f \colon X \to [-\infty, \infty]$ by

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Prove that if $\int |f_1| \, d\mu < \infty$, then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

17 Show that if $(X, \mathcal{S}, \mu)$ is a measure space and $f \colon X \to [0, \infty)$ is $\mathcal{S}$-measurable, then

$$\mu(X) \inf_{x\in X} f(x) \leq \int f \, d\mu \leq \mu(X) \sup_{x\in X} f(x).$$

18 Henri Lebesgue wrote the following about his method of integration:

> I have to pay a certain sum, which I have collected in my pocket. I take the bills and coins out of my pocket and give them to the creditor in the order I find them until I have reached the total sum. This is the Riemann integral. But I can proceed differently. After I have taken all the money out of my pocket I order the bills and coins according to identical values and then I pay the several heaps one after the other to the creditor. This is my integral.

Use 3.15 to explain what Lebesgue meant and to explain why integration of a function with respect to a measure can be thought of as partitioning the range of the function, while Riemann integration depends upon partitioning the domain of the function.
[*The quote above is taken from page 796 of* The Princeton Companion to Mathematics*, edited by Timothy Gowers.*]

3B Limits of Integrals & Integrals of Limits

This section will focus on interchanging limits and integrals. Those tools will allow us to characterize the Riemann integrable functions in terms of Lebesgue measure. We will also develop some good approximation tools that will be useful in later chapters.

Bounded Convergence Theorem

We begin this section by introducing some useful notation.

3.24 Definition ***integration on a subset***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E \in \mathcal{S}$. If $f\colon X \to [-\infty, \infty]$ is an $\mathcal{S}$-measurable function, then $\int_E f\, d\mu$ is defined by

$$\int_E f\, d\mu = \int \chi_E f\, d\mu$$

if the right side of the equation above is defined; otherwise $\int_E f\, d\mu$ is undefined.

Alternatively, you can think of $\int_E f\, d\mu$ as $\int f|_E\, d\mu_E$, where μ_E is the measure obtained by restricting μ to the elements of $\mathcal{S}$ that are contained in E.

Notice that according to the definition above, the notation $\int_X f\, d\mu$ means the same as $\int f\, d\mu$. The following easy result illustrates the use of this new notation.

3.25 ***Bounding an integral***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E \in \mathcal{S}$, and $f\colon X \to [-\infty, \infty]$ is a function such that $\int_E f\, d\mu$ is defined. Then

$$\left| \int_E f\, d\mu \right| \leq \mu(E) \sup_{x \in E} |f(x)|.$$

Proof Let $c = \sup_{x \in E} |f(x)|$. We have

$$\begin{aligned} \left| \int_E f\, d\mu \right| &= \left| \int \chi_E f\, d\mu \right| \\ &\leq \int \chi_E |f|\, d\mu \\ &\leq \int c\chi_E\, d\mu \\ &= c\mu(E), \end{aligned}$$

where the second line comes from 3.23, the third line comes from 3.8, and the fourth line comes from 3.15. ■

The next result could be proved as a special case of the Dominated Convergence Theorem (3.30), which we will prove later in this section. Thus you could skip the proof here. However, sometimes you get more insight by seeing an easier proof of an important special case. Thus you may want to read the easy proof of the Bounded Convergence Theorem that is presented next.

3.26 *Bounded Convergence Theorem*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space with $\mu(X) < \infty$. Suppose $f \colon X \to \mathbf{R}$ is $\mathcal{S}$-measurable and $f_1, f_2, \ldots$ are $\mathcal{S}$-measurable functions from X to $\mathbf{R}$ such that

$$\lim_{k\to\infty} f_k(x) = f(x)$$

for all $x \in X$. If there exists $c \in (0, \infty)$ such that

$$|f_k(x)| \le c$$

for all $k \in \mathbf{Z}^+$ and all $x \in X$, then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

Note the key role of Egorov's Theorem, which states that pointwise convergence is close to uniform convergence, in proofs involving interchanging limits and integrals.

Proof Suppose c satisfies the hypothesis of this theorem. Let $\varepsilon > 0$. By Egorov's Theorem (2.79), there exists $E \in \mathcal{S}$ such that $\mu(X \setminus E) < \frac{\varepsilon}{4c}$ and $f_1, f_2, \ldots$ converges uniformly to f on E. Now

$$\begin{aligned}\Bigl|\int f_k \, d\mu - \int f \, d\mu\Bigr| &= \Bigl|\int_{X\setminus E} f_k \, d\mu - \int_{X\setminus E} f \, d\mu + \int_E (f_k - f) \, d\mu\Bigr| \\ &\le \int_{X\setminus E} |f_k| \, d\mu + \int_{X\setminus E} |f| \, d\mu + \int_E |f_k - f| \, d\mu \\ &< \frac{\varepsilon}{2} + \mu(E) \sup_{x\in E} |f_k(x) - f(x)|,\end{aligned}$$

where the last inequality follows from 3.25. Because $f_1, f_2, \ldots$ converges uniformly to f on E and $\mu(E) < \infty$, the right side of the inequality above is less than ε for k sufficiently large, which completes the proof. ■

Sets of Measure 0 in Integration Theorems

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. If $f, g \colon X \to [-\infty, \infty]$ are $\mathcal{S}$-measurable functions and

$$\mu(\{x \in X : f(x) \ne g(x)\}) = 0,$$

then the definition of an integral implies that $\int f \, d\mu = \int g \, d\mu$ (or both integrals are undefined). Because what happens on a set of measure 0 often does not matter, the following definition is useful.

3.27 Definition ***almost every***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. A set $E \in \mathcal{S}$ is said to contain *μ-almost every* element of X if $\mu(X \setminus E) = 0$. If the measure μ is clear from the context, then the phrase *almost every* is usually used (abbreviated by some authors to *a.e.*).

For example, almost every real number is irrational (with respect to the usual Lebesgue measure on $\mathbf{R}$) because $|\mathbf{Q}| = 0$.

Theorems about integrals can almost always be relaxed so that the hypotheses apply only almost everywhere instead of everywhere. For example, consider the Bounded Convergence Theorem (3.26), one of whose hypotheses is that

$$\lim_{k\to\infty} f_k(x) = f(x)$$

for all $x \in X$. Suppose that the hypotheses of the Bounded Convergence Theorem hold except that the equation above holds only almost everywhere, meaning there is a set $E \in \mathcal{S}$ such that $\mu(X \setminus E) = 0$ and the equation above holds for all $x \in E$. Define new functions $g_1, g_2, \ldots$ and g by

$$g_k(x) = \begin{cases} f_k(x) & \text{if } x \in E, \\ 0 & \text{if } x \in X \setminus E \end{cases} \quad \text{and} \quad g(x) = \begin{cases} f(x) & \text{if } x \in E, \\ 0 & \text{if } x \in X \setminus E. \end{cases}$$

Then

$$\lim_{k\to\infty} g_k(x) = g(x)$$

for all $x \in X$. Hence the Bounded Convergence Theorem implies that

$$\lim_{k\to\infty} \int g_k \, d\mu = \int g \, d\mu,$$

which immediately implies that

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu$$

because $\int g_k \, d\mu = \int f_k \, d\mu$ and $\int g \, d\mu = \int f \, d\mu$.

Dominated Convergence Theorem

The next result tells us that if a nonnegative function has a finite integral, then its integral over all small sets (in the sense of measure) is small.

3.28 ***Integrals on small sets are small***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $g \colon X \to [0, \infty]$ is $\mathcal{S}$-measurable, and $\int g \, d\mu < \infty$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_B g \, d\mu < \varepsilon$$

for every set $B \in \mathcal{S}$ such that $\mu(B) < \delta$.

Proof Suppose $\varepsilon > 0$. Let $h \colon X \to [0, \infty)$ be a simple $\mathcal{S}$-measurable function such that $0 \le h \le g$ and

$$\int g \, d\mu - \int h \, d\mu < \frac{\varepsilon}{2};$$

the existence of a function h with these properties follows from 3.9. Let

$$H = \max\{h(x) : x \in X\}$$

and let $\delta > 0$ be such that $H\delta < \frac{\varepsilon}{2}$.

Suppose $B \in \mathcal{S}$ and $\mu(B) < \delta$. Then

$$\begin{aligned}
\int_B g \, d\mu &= \int_B (g - h) \, d\mu + \int_B h \, d\mu \\
&\le \int (g - h) \, d\mu + H\mu(B) \\
&< \frac{\varepsilon}{2} + H\delta \\
&< \varepsilon,
\end{aligned}$$

as desired. ■

Some theorems, such as Egorov's Theorem (2.79) have as a hypothesis that the measure of the entire space is finite. The next result sometimes allows us to get around this hypothesis by restricting attention to a key set of finite measure.

3.29 ***Integrable functions live mostly on sets of finite measure***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $g \colon X \to [0, \infty]$ is $\mathcal{S}$-measurable, and $\int g \, d\mu < \infty$. Then for every $\varepsilon > 0$, there exists $E \in \mathcal{S}$ such that $\mu(E) < \infty$ and

$$\int_{X \setminus E} g \, d\mu < \varepsilon.$$

Proof Suppose $\varepsilon > 0$. Let P be an $\mathcal{S}$-partition $A_1, \dots, A_m$ of X such that

$$\int g \, d\mu < \varepsilon + \mathcal{L}(g, P).$$

Let E be the union of those A_j such that $\inf_{x \in A_j} f(x) > 0$. Then $\mu(E) < \infty$ (because otherwise we would have $\mathcal{L}(g, P) = \infty$, which contradicts the hypothesis that $\int g \, d\mu < \infty$). Now

$$\begin{aligned}
\int_{X \setminus E} g \, d\mu &= \int g \, d\mu - \int \chi_E g \, d\mu \\
&< \bigl(\varepsilon + \mathcal{L}(g, P)\bigr) - \mathcal{L}(\chi_E g, P) \\
&= \varepsilon,
\end{aligned}$$

where the last equality holds because $\inf_{x \in A_j} f(x) = 0$ for each A_j not contained in E. ■

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions on X such that $\lim_{k\to\infty} f_k(x) = f(x)$ for all (or almost all) $x \in X$. In general, it is not true that $\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu$; see Exercises 1 and 2.

We already have two good theorems about interchanging limits and integrals. However, both of these theorems have restrictive hypotheses. Specifically, the Monotone Convergence Theorem (3.11) requires all the functions to be nonnegative and it requires the sequence of functions to be increasing. The Bounded Convergence Theorem (3.26) requires the measure of the whole space to be finite and it requires the sequence of functions to be uniformly bounded by a constant.

The next theorem is the grand result in this area. It does not require the sequence of functions to be nonnegative, it does not require the sequence of functions to be increasing, it does not require the measure of the whole space to be finite, and it does not require the sequence of functions to be uniformly bounded. All these hypotheses are replaced only by a requirement that the sequence of functions is pointwise bounded by a function with a finite integral.

Notice that the Bounded Convergence Theorem follows immediately from the result below (take g to be an appropriate constant function and use the hypothesis in the Bounded Convergence Theorem that $\mu(X) < \infty$).

3.30 *Dominated Convergence Theorem*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $f\colon X \to [-\infty, \infty]$ is $\mathcal{S}$-measurable, and $f_1, f_2, \ldots$ are $\mathcal{S}$-measurable functions from X to $[-\infty, \infty]$ such that

$$\lim_{k\to\infty} f_k(x) = f(x)$$

for almost every $x \in X$. If there exists an $\mathcal{S}$-measurable function $g\colon X \to [0, \infty]$ such that

$$\int g \, d\mu < \infty \quad \text{and} \quad |f_k(x)| \le g(x)$$

for every $k \in \mathbf{Z}^+$ and almost every $x \in X$, then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

Proof Suppose $g\colon X \to [0, \infty]$ satisfies the hypotheses of this theorem. If $E \in \mathcal{S}$, then

$$\begin{aligned}
\Bigl| \int f_k \, d\mu - \int f \, d\mu \Bigr| &= \Bigl| \int_{X\setminus E} f_k \, d\mu - \int_{X\setminus E} f \, d\mu + \int_E f_k \, d\mu - \int_E f \, d\mu \Bigr| \\
&\le \Bigl| \int_{X\setminus E} f_k \, d\mu \Bigr| + \Bigl| \int_{X\setminus E} f \, d\mu \Bigr| + \Bigl| \int_E f_k \, d\mu - \int_E f \, d\mu \Bigr| \\
&\le 2\int_{X\setminus E} g \, d\mu + \Bigl| \int_E f_k \, d\mu - \int_E f \, d\mu \Bigr|. \qquad (3.31)
\end{aligned}$$

Case 1: Suppose $\mu(X) < \infty$.

Let $\varepsilon > 0$. By 3.28, there exists $\delta > 0$ such that

3.32
$$\int_B g \, d\mu < \frac{\varepsilon}{4}$$

for every set $B \in \mathcal{S}$ such that $\mu(B) < \delta$. By Egorov's Theorem (2.79), there exists a set $E \in \mathcal{S}$ such that $\mu(X \setminus E) < \delta$ and $f_1, f_2, \dots$ converges uniformly to f on E. Now 3.31 and 3.32 imply that

$$\left| \int f_k \, d\mu - \int f \, d\mu \right| < \frac{\varepsilon}{2} + \left| \int_E (f_k - f) \, d\mu \right|.$$

Because $f_1, f_2, \dots$ converges uniformly to f on E and $\mu(E) < \infty$, the last term on the right is less than $\frac{\varepsilon}{2}$ for all sufficiently large k. Thus $\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu$, completing the proof of case 1.

Case 2: Suppose $\mu(X) = \infty$.

Let $\varepsilon > 0$. By 3.29, there exists $E \in \mathcal{S}$ such that $\mu(E) < \infty$ and

$$\int_{X \setminus E} g \, d\mu < \frac{\varepsilon}{4}.$$

The inequality above and 3.31 imply that

$$\left| \int f_k \, d\mu - \int f \, d\mu \right| < \frac{\varepsilon}{2} + \left| \int_E f_k \, d\mu - \int_E f \, d\mu \right|.$$

By case 1 as applied to the sequence $f_1|_E, f_2|_E, \dots$, the last term on the right is less than $\frac{\varepsilon}{2}$ for all sufficiently large k. Thus $\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu$, completing the proof of case 2. ∎

Riemann Integrals and Lebesgue Integrals

We can now use the tools we have developed to characterize the Riemann integrable functions. In the theorem below, the left side of the last equation denotes the Riemann integral.

3.33 *Riemann integrable $\iff$ continuous almost everywhere*

Suppose $a < b$ and $f \colon [a, b] \to \mathbf{R}$ is a bounded function. Then f is Riemann integrable if and only if

$$\left| \{ x \in [a, b] : f \text{ is not continuous at } x \} \right| = 0.$$

Furthermore, if f is Riemann integrable and λ denotes Lebesgue measure on $\mathbf{R}$, then f is Lebesgue measurable and

$$\int_a^b f = \int_{[a,b]} f \, d\lambda.$$

Proof Suppose $k \in \mathbf{Z}^+$. Consider the partition P_k that divides $[a,b]$ into 2^k subintervals of equal size. Let $I_1, \ldots, I_{2^k}$ be the corresponding closed subintervals, each of length $(b-a)/2^k$. Let

3.34
$$g_k = \sum_{j=1}^{2^k} \bigl(\inf_{x \in I_j} f(x)\bigr)\chi_{I_j} \quad \text{and} \quad h_k = \sum_{j=1}^{2^k} \bigl(\sup_{x \in I_j} f(x)\bigr)\chi_{I_j}.$$

The lower and upper Riemann sums of f for the partition P_k are given by integrals. Specifically,

3.35
$$L(f, P_k, [a,b]) = \int_{[a,b]} g_k \, d\lambda \quad \text{and} \quad U(f, P_k, [a,b]) = \int_{[a,b]} h_k \, d\lambda,$$

where λ is Lebesgue measure on $\mathbf{R}$.

The definitions of g_k and h_k given in 3.34 are actually just a first draft of the definitions. A slight problem arises at each point that is in two of the intervals $I_1, \ldots, I_{2^k}$ (in other words, at endpoints of these intervals other than a and b). At each of these points, change the value of g_k to be the infimum of f over the union of the two intervals that contain the point, and change the value of h_k to be the supremum of f over the union of the two intervals that contain the point. This change modifies g_k and h_k on only a finite number of points. Thus the integrals in 3.35 are not affected. This change is needed in order to make 3.37 true (otherwise the two sets in 3.37 might differ by at most countably many points, which would not really change the proof but which would not be as aesthetically pleasing).

Clearly $g_1 \leq g_2 \leq \cdots$ is an increasing sequence of functions and $h_1 \geq h_2 \geq \cdots$ is a decreasing sequence of functions on $[a,b]$. Define functions $f^{\mathrm{L}} \colon [a,b] \to \mathbf{R}$ and $f^{\mathrm{U}} \colon [a,b] \to \mathbf{R}$ by

$$f^{\mathrm{L}}(x) = \lim_{k \to \infty} g_k(x) \quad \text{and} \quad f^{\mathrm{U}}(x) = \lim_{k \to \infty} h_k(x).$$

Taking the limit as $k \to \infty$ of both equations in 3.35 and using the Bounded Convergence Theorem (3.26) along with Exercise **??** in Section 1A, we see that f_L and f^U are Lebesgue measurable functions and

3.36
$$L(f, [a,b]) = \int_{[a,b]} f^{\mathrm{L}} \, d\lambda \quad \text{and} \quad U(f, [a,b]) = \int_{[a,b]} f^{\mathrm{U}} \, d\lambda.$$

Now 3.36 implies that f is Riemann integrable if and only if

$$\int_{[a,b]} (f^{\mathrm{U}} - f^{\mathrm{L}}) \, d\lambda = 0.$$

Because $f^{\mathrm{L}}(x) \leq f(x) \leq f^{\mathrm{U}}(x)$ for all $x \in [a,b]$, the equation above holds if and only if

$$|\{x \in [a,b] : f^{\mathrm{U}}(x) \neq f^{\mathrm{L}}(x)\}| = 0.$$

The remaining details of the proof can be completed by noting that

3.37
$$\{x \in [a,b] : f^{\mathrm{U}}(x) \neq f^{\mathrm{L}}(x)\} = \{x \in [a,b] : f \text{ is not continuous at } x\}.$$ ■

We previously defined the notation $\int_a^c f$ to mean the Riemann integral of f. Because the Riemann integral and Lebesgue integral agree for Riemann integrable functions (see 3.33), we now redefine $\int_a^c f$ to denote the Lebesgue integral.

3.38 Definition $\int_a^c f$

Suppose $-\infty \leq a < c \leq \infty$ and $f\colon [a,c] \to \mathbf{R}$ is Lebesgue measurable. Then

- $\int_a^c f$ is defined to equal $\int_{(a,c)} f \, d\lambda$, where λ is Lebesgue measure on $\mathbf{R}$;
- $\int_c^a f$ is defined to equal $-\int_a^c f$.

The definition in the second bullet point above is made so that equations such as

$$\int_a^c f = \int_a^b f + \int_b^c f$$

remain valid even if, for example, $a < c < b$.

Approximation by Nice Functions

In the next definition, the notation $\|f\|_1$ should be $\|f\|_{1,\mu}$ because it depends upon the measure μ as well as upon f. However, μ is usually clear from the context. In some books, you may see the notation $\mathcal{L}^1(X, \mathcal{S}, \mu)$ instead of $\mathcal{L}^1(\mu)$.

3.39 Definition $\|f\|_1$; $\mathcal{L}^1(\mu)$

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. If $f\colon X \to [-\infty, \infty]$ is $\mathcal{S}$-measurable, then the *$\mathcal{L}^1$-norm* of f is denoted by $\|f\|_1$ and is defined by

$$\|f\|_1 = \int |f| \, d\mu.$$

The *Lebesgue space* $\mathcal{L}^1(\mu)$ is defined by

$$\mathcal{L}^1(\mu) = \{f : f \text{ is an } \mathcal{S}\text{-measurable function from } X \text{ to } \mathbf{R} \text{ and } \|f\|_1 < \infty.\}$$

The terminology and notation used above are convenient even though $\|\cdot\|_1$ might not be a genuine norm (to be defined in Chapter 6).

3.40 Example $\mathcal{L}^1(\mu)$ *functions that take on only finitely many values*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E_1, \dots, E_n$ are disjoint subsets of X. Suppose $a_1, \dots, a_n$ are distinct nonzero real numbers. Then

$$a_1 \chi_{E_1} + \cdots + a_n \chi_{E_n} \in \mathcal{L}^1(\mu)$$

if and only if $E_k \in \mathcal{S}$ and $\mu(E_k) < \infty$ for all $k \in \{1, \dots, n\}$. Furthermore,

$$\|a_1 \chi_{E_1} + \cdots + a_n \chi_{E_n}\|_1 = |a_1|\mu(E_1) + \cdots + |a_n|\mu(E_n).$$

3.41 Example *ℓ^1 equals $\mathcal{L}^1$(counting measure on $\mathbf{Z}^+$)*

If μ equals counting measure on $\mathbf{Z}^+$ and $x = x_1, x_2, \ldots$ is a sequence of real numbers (thought of as a function on $\mathbf{Z}^+$), then $\|x\|_1 = \sum_{k=1}^{\infty} |x_k|$. Is this case, $\mathcal{L}^1(\mu)$ is often denoted by ℓ^1 (pronounced *little-el-one*). In other words, ℓ^1 is the set of all sequences $x_1, x_2, \ldots$ of real numbers such that $\sum_{k=1}^{\infty} |x_k| < \infty$.

The easy proof of the following result is left to the reader.

3.42 ***Properties of the $\mathcal{L}^1$-norm***

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g \in \mathcal{L}^1(\mu)$. Then

- $\|f\|_1 \geq 0$;
- $\|f\|_1 = 0$ if and only $f(x) = 0$ for almost every $x \in X$;
- $\|cf\|_1 = |c|\|f\|_1$ for all $c \in \mathbf{R}$;
- $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.

The next result states every function in $\mathcal{L}^1(\mu)$ can be approximated in $\mathcal{L}^1$-norm by measurable functions that take on only finitely many values.

3.43 ***Approximation by simple functions***

Suppose μ is a measure and $f \in \mathcal{L}^1(\mu)$. Then for every $\varepsilon > 0$, there exists a simple function $g \in \mathcal{L}^1(\mu)$ such that

$$\|f - g\|_1 < \varepsilon.$$

Proof Suppose $\varepsilon > 0$. Then there exist simple functions $g_1, g_2 \in \mathcal{L}^1(\mu)$ such that $0 \leq g_1 \leq f^+$ and $0 \leq g_2 \leq f^-$ and

$$\int (f^+ - g_1)\, d\mu < \frac{\varepsilon}{2} \quad \text{and} \quad \int (f^- - g_2)\, d\mu < \frac{\varepsilon}{2},$$

where we have used 3.9 to provide the existence of g_1, g_2 with these properties.

Let $g = g_1 - g_2$. Then g is a simple function in $\mathcal{L}^1(\mu)$ and

$$\begin{aligned} \|f - g\|_1 &= \|(f^+ - g_1) - (f^- - g_2)\|_1 \\ &= \int (f^+ - g_1)\, d\mu + \int (f^- - g_2)\, d\mu \\ &< \varepsilon, \end{aligned}$$

as desired. ∎

3.44 Definition $\mathcal{L}^1(\mathbf{R})$; $\|f\|_1$

- The notation $\mathcal{L}^1(\mathbf{R})$ denotes $\mathcal{L}^1(\lambda)$, where λ is Lebesgue measure on either the Borel subsets of $\mathbf{R}$ or the Lebesgue measurable subsets of $\mathbf{R}$.
- When working with $\mathcal{L}^1(\mathbf{R})$, the notation $\|f\|_1$ denotes the integral of the absolute value of f with respect to Lebesgue measure on $\mathbf{R}$.

3.45 Definition *step function*

A *step function* is a function $g\colon \mathbf{R} \to \mathbf{R}$ of the form

$$g = a_1\chi_{I_1} + \cdots + a_n\chi_{I_n},$$

where $I_1, \dots, I_n$ are intervals of $\mathbf{R}$ and $a_1, \dots, a_n$ are nonzero real numbers.

Suppose g is a step function of the form above and the intervals $I_1, \dots, I_n$ are disjoint. Then

$$\|g\|_1 = |a_1|\,|I_1| + \cdots + |a_n|\,|I_n|.$$

In particular, $g \in \mathcal{L}^1(\mathbf{R})$ if and only if all the intervals $I_1, \dots, I_n$ are bounded.

The intervals in the definition of a step function can be open intervals, closed intervals, or half-open intervals. We will be using step functions in integrals, where the inclusion or exclusion of the endpoints of the intervals does not matter.

Even though the coefficients $a_1, \dots, a_n$ in the definition of a step function are required to be nonzero, the function 0 that is identically 0 on $\mathbf{R}$ is a step function. To see this, take $n = 1$, $a_1 = 1$, and $I_1 = \emptyset$.

3.46 ***Approximation by step functions***

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Then for every $\varepsilon > 0$, there exists a step function $g \in \mathcal{L}^1(\mathbf{R})$ such that

$$\|f - g\|_1 < \varepsilon.$$

Proof Suppose $\varepsilon > 0$. By 3.43, there exist Borel (or Lebesgue) measurable subsets $A_1, \dots, A_n$ of $\mathbf{R}$ and nonzero numbers $a_1, \dots, a_n$ such that $|A_k| < \infty$ for all $k \in \{1, \dots, n\}$ and

$$\Bigl\| f - \sum_{k=1}^{n} a_k\chi_{A_k} \Bigr\|_1 < \frac{\varepsilon}{2}.$$

For each $k \in \{1, \dots, n\}$, there is an open subset B_k of $\mathbf{R}$ that contains A_k and whose Lebesgue measure is as close as we want to $|A_k|$ [by part (e) of 2.70]. Each B_k is a countable union of disjoint open intervals (by 0.59 in the Appendix). Thus for each k, there is a set E_k that is a finite union of bounded open intervals contained in B_k whose Lebesgue measure is as close as we want to $|B_k|$. Hence for each k, there is a set E_k that is a finite union of bounded intervals such that

$$
\begin{aligned}
|E_k \setminus A_k| + |A_k \setminus E_k| &\le |B_k \setminus A_k| + |B_k \setminus E_k| \\
&< \frac{\varepsilon}{2|a_k|n};
\end{aligned}
$$

in other words,

$$\|\chi_{A_k} - \chi_{E_k}\|_1 < \frac{\varepsilon}{2|a_k|n}.$$

Now

$$
\begin{aligned}
\Big\| f - \sum_{k=1}^{n} a_k \chi_{E_k} \Big\|_1 &\le \Big\| f - \sum_{k=1}^{n} a_k \chi_{A_k} \Big\|_1 + \Big\| \sum_{k=1}^{n} a_k \chi_{A_k} - \sum_{k=1}^{n} a_k \chi_{E_k} \Big\|_1 \\
&< \frac{\varepsilon}{2} + \sum_{k=1}^{n} |a_k| \|\chi_{A_k} - \chi_{E_k}\|_1 \\
&< \varepsilon.
\end{aligned}
$$

Each E_k is a finite union of bounded intervals. Thus the inequality above completes the proof because $\sum_{k=1}^{n} a_k \chi_{E_k}$ is a step function. ■

Luzin's Theorem (2.83 and 2.85) gives a spectacular way to approximate a Borel measurable function by a continuous function. However, the following approximation theorem is usually more useful than Luzin's Theorem. For example, the next result plays a major role in the proof of the Lebesgue Differentiation Theorem (4.10).

3.47 *Approximation by continuous functions*

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Then for every $\varepsilon > 0$, there exists a continuous function $g \colon \mathbf{R} \to \mathbf{R}$ such that

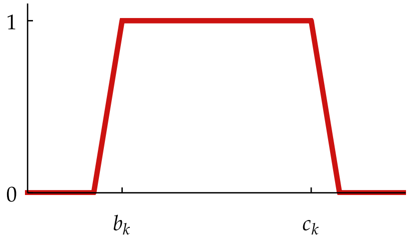
$$\|f - g\|_1 < \varepsilon$$

and $\{x \in \mathbf{R} : g(x) \neq 0\}$ is a bounded set.

Proof For every $a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n \in \mathbf{R}$ and $g_1, \dots, g_n \in \mathcal{L}^1(\mathbf{R})$, we have

$$
\begin{aligned}
\Big\| f - \sum_{k=1}^{n} a_k g_k \Big\|_1 &\le \Big\| f - \sum_{k=1}^{n} a_k \chi_{[b_k, c_k]} \Big\|_1 + \Big\| \sum_{k=1}^{n} a_k (\chi_{[b_k, c_k]} - g_k) \Big\|_1 \\
&\le \Big\| f - \sum_{k=1}^{n} a_k \chi_{[b_k, c_k]} \Big\|_1 + \sum_{k=1}^{n} |a_k| \, \|\chi_{[b_k, c_k]} - g_k\|_1,
\end{aligned}
$$

where the inequalities above follow from 3.42. By 3.46, we can choose $a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n \in \mathbf{R}$ to make $\|f - \sum_{k=1}^{n} a_k \chi_{[b_k, c_k]}\|_1$ as small as we wish. The figure here then shows that there exist continuous functions $g_1, \dots, g_n \in \mathcal{L}^1(\mathbf{R})$ that make $\sum_{k=1}^{n} |a_k| \|\chi_{[b_k, c_k]} - g_k\|_1$ as small as we wish. Now take $g = \sum_{k=1}^{n} a_k g_k$. ■



The graph of a continuous function g_k such that $\|\chi_{[b_k, c_k]} - g_k\|_1$ is small.

EXERCISES 3B

1 Give an example of a sequence $f_1, f_2, \ldots$ of functions from $\mathbf{Z}^+$ to $[0, \infty)$ such that
$$\lim_{k\to\infty} f_k(m) = 0$$
for every $m \in \mathbf{Z}^+$ but $\displaystyle\lim_{k\to\infty} \int f_k \, d\mu = 1$, where μ is counting measure on $\mathbf{Z}^+$.

2 Give an example of a sequence $f_1, f_2, \ldots$ of continuous functions from $\mathbf{R}$ to $[0, 1]$ such that
$$\lim_{k\to\infty} f_k(x) = 0$$
for every $x \in \mathbf{R}$ but $\displaystyle\lim_{k\to\infty} \int f_k \, d\lambda = \infty$, where λ is Lebesgue measure on $\mathbf{R}$.

3 Suppose λ is Lebesgue measure on $\mathbf{R}$ and $f\colon \mathbf{R} \to \mathbf{R}$ is a Borel measurable function such that $\int |f| \, d\lambda < \infty$. Define $g\colon \mathbf{R} \to \mathbf{R}$ by
$$g(x) = \int_{(-\infty, x)} f \, d\lambda.$$
Prove that g is uniformly continuous on $\mathbf{R}$.

4 (a) Suppose $(X, \mathcal{S}, \mu)$ is a measure space with $\mu(X) < \infty$. Suppose that $f\colon X \to \mathbf{R}$ is a bounded $\mathcal{S}$-measurable function. Prove that
$$\int f \, d\mu = \inf\Bigl\{ \int h \, d\mu : h \text{ is a simple } \mathcal{S}\text{-measurable function and } f \le h \Bigr\}.$$

(b) Show that the conclusion of part (a) can fail if the hypothesis that f is bounded is replaced by the hypothesis that $\int |f| \, d\mu < \infty$.

(c) Show that the conclusion of part (a) can fail if the condition that $\mu(X) < \infty$ is deleted.

[*Part (a) of this exercise shows that if we had defined an upper Lebesgue sum, then we could have used it to define the integral. However, parts (b) and (c) show that the hypotheses that f is bounded and that $\mu(X) < \infty$ would be needed if defining the integral via the equation above. The definition of the integral via the lower Lebesgue sum does not require these hypotheses, showing that the approach via the lower Lebesgue sum is the right definition.*]

5 Let λ denote Lebesgue measure on $\mathbf{R}$. Suppose $f\colon \mathbf{R} \to \mathbf{R}$ is a Borel measurable function such that $\int |f| \, d\lambda < \infty$. Prove that
$$\lim_{k\to\infty} \int_{[-k,k]} f \, d\lambda = \int f \, d\lambda.$$

6 Let λ denote Lebesgue measure on $\mathbf{R}$. Give an example of a continuous function $f\colon [0, \infty) \to \mathbf{R}$ such that $\lim_{t\to\infty} \int_{[0,t]} f \, d\lambda$ exists (in $\mathbf{R}$) but $\int_{[0,\infty)} f \, d\lambda$ is not defined.

7 Let λ denote Lebesgue measure on $\mathbf{R}$. Give an example of a continuous function $f\colon (0,1) \to \mathbf{R}$ such that $\lim_{n\to\infty} \int_{(\frac{1}{n},1)} f\,d\lambda$ exists (in $\mathbf{R}$) but $\int_{(0,1)} f\,d\lambda$ is not defined.

8 Verify the assertion in 3.37.

9 Verify the assertion in Example 3.40.

10 Suppose $(X, \mathcal{S}, \mu)$ is a measure space such that $\mu(X) < \infty$. Suppose p, r are positive numbers with $p < r$. Prove that if $f\colon X \to [0,\infty)$ is an $\mathcal{S}$-measurable function such that $\int f^r\,d\mu < \infty$, then $\int f^p\,d\mu < \infty$.

11 Give an example to show that the result in the previous exercise can be false without the hypothesis that $\mu(X) < \infty$.

12 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f \in \mathcal{L}^1(\mu)$. Prove that

$$\{x \in X : f(x) \neq 0\}$$

is the countable union of sets with finite μ-measure.

13 Suppose

$$f_k(x) = \frac{(1-x)^k \cos\frac{k}{x}}{\sqrt{x}}.$$

Prove that $\displaystyle\lim_{k\to\infty} \int_0^1 f_k = 0$.

14 Give an example of a sequence of nonnegative Borel measurable functions $f_1, f_2, \dots$ on $[0,1]$ such that both the following conditions hold:

- $\displaystyle\lim_{k\to\infty} \int_0^1 f_k = 0$;
- $\displaystyle\sup_{k\geq m} f_k(x) = \infty$ for every $m \in \mathbf{Z}^+$ and every $x \in [0,1]$.

15 Let λ denote Lebesgue measure on $\mathbf{R}$.

(a) Let $f(x) = 1/\sqrt{x}$. Prove that $\int_{[0,1]} f\,d\lambda = 2$.

(b) Let $f(x) = 1/(1+x^2)$. Prove that $\int_{\mathbf{R}} f\,d\lambda = \pi$.

(c) Let $f(x) = (\sin x)/x$. Show that $\int_{[0,\infty)} f\,d\lambda$ is not defined but

$$\lim_{t\to\infty} \int_{[0,t]} f\,d\lambda$$

exists in $\mathbf{R}$.

16 For $f\colon \mathbf{R} \to \mathbf{R}$ and $t \in \mathbf{R}$, define $f_t\colon \mathbf{R} \to \mathbf{R}$ by $f_t(x) = f(x-t)$ [thus if $t > 0$, then the graph of f_t is obtained by shifting the graph of f to the right by t units]. Prove that if $f \in \mathcal{L}^1(\mathbf{R})$, then

$$\lim_{t\to 0} \|f_t - f\|_1 = 0.$$