CHAPTER 1

HISTORY OF THE SUBJECT

1.1 HISTORY OF THE IDEA

In a broad sense, much of mathematics is devoted to the decomposition, or analysis, of a whole entity into its component parts and the reconstruction, or synthesis, of the whole from its parts. There is in this an expectation that the whole is somehow equal to the sum of its parts, which are simpler individually. We discuss briefly a few examples of this aspect of mathematics.

It was nearly three thousand years ago that Babylonian astronomers successfully predicted the times of lunar and solar eclipses by expressing these complicated events as summations of numerous simpler periodic events. The predictions were fairly accurate, to the extent of predicting eclipses that would be visible at least from some part of the world. That remarkable achievement may be interpreted as the first appearance on Earth of harmonic (or Fourier) analysis.

Measurement has been a special interest for mathematicians and scientists since the early days of civilization. Two thousand years ago, the classical geometers of Greece made profound contributions to the study of measurement. A line segment could be *measured by a shorter segment* if the short one could be laid off end to end in such a way that the long segment would be seen to be an exact positive integer

Measure and Integration: A Concise Introduction to Real Analysis. By Leonard F. Richardson

multiple of the short one. Two segments were called *commensurable* if both could be measured by a *common* shorter segment, or *unit* of length. And the discovery by Pythagoras or his associates of *incommensurable* segments was a momentous event in the history of mathematics. Inherent in the definition Greek geometers used for measurement of segments was the concept that the measure of the *whole segment must be the sum of the measures of its nonoverlapping parts*, as indicated by the lengths marked off according to the shorter segment.

The Greek geometers pushed this technique much farther, successfully studying the circumference and area of a circle and the surface area and volume of a sphere. These challenges were met by the application of what is called the *principle of exhaustion*. The method was to approximate a geometrical measurement from below, and at each successive iteration of the approximation technique, at least of half of what remained to be counted was to be included. (Greek geometers understood that if too little were taken at each stage, even an endless succession of steps might not approach the goal.) For example, the area of a circle is approximated from within by means of an inscribed 2^n-sided regular polygon. As n increases, the area of the resulting 2^n-sided polygon grows in a computable manner. (See Figure 1.1.) We begin with an inscribed square consuming the bulk of the area of the circle. The next polygon is an octagon, which adds the areas of four thin triangles to that of the square. At each stage, one can construct the 2^{n+1}-gon by bisecting the arcs that are

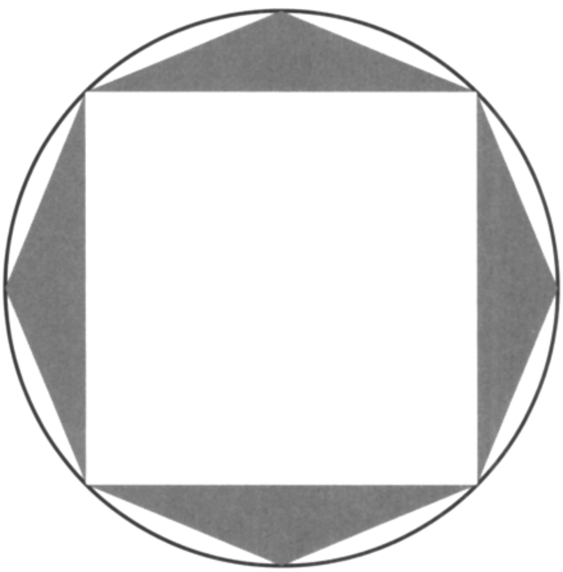

Figure 1.1 Area of a circle by exhaustion.

subtended by the sides of the inscribed 2^n-gon. In effect, the $(n + 1)$th stage of the approximation results from adding a small increment to what was obtained at the nth stage. Such classical achievements in geometry may have been the first instances in which a measure (of area, for example) was sought that would be *countably additive*, meaning that the measure of the whole *should be* the sum of the measures of its *infinite sequence* of nonoverlapping parts.

Greek geometers and philosophers may not have been entirely convinced of the validity of the principle of exhaustion. The expression of some dissatisfaction with this geometric technique may have been the purpose of the famous paradox of Zeno, in which the legendary warrior Achilles was pitted unsuccessfully in a footrace with a persistent tortoise.

The notion that the whole should be the sum of even an infinite sequence of its nonoverlapping parts appears very strongly in modern analysis, both pure and applied. For example, in 1822 Joseph Fourier presented a seminal paper on the heat equation to the French Academy, in which he introduced the use of infinite trigonometric series for the purpose of determining the solution of the heat equation on a finite interval, or rod [6]. Fourier's method required that the hypothetical solution function be *expressible* as the sum of an infinite trigonometric series:

$$f(x) = \sum_{0}^{\infty}(a_n \cos nx + b_n \sin nx).$$

Unfortunately, these so-called Fourier series can diverge for very large sets of numbers x in the domain of even a rather nice function f. Yet, if one ignored the embarrassing reality that $f(x)$ *need not be the sum of its parts*, Fourier's method actually worked. And the search was on for suitable concepts and tools to analyze correctly the right classes of functions for which the Fourier series would converge in a useful sense to the functions being represented.

The efforts took place on a grand scale. Even the theory of sets was invented by Georg Cantor for the purpose of analyzing sets of convergence and divergence of Fourier series. Though Cantor's approach was not sufficient for the needs of Fourier series, it became a cornerstone of modern mathematics.

Early in the twentieth century Henri Lebesgue invented his new and very refined concept of the integral, based on the *measure* of suitable subsets of the line. Lebesgue measure became the foundation not only for Fourier analysis, but also for probability, and for functional analysis which permeates modern analysis.

1.2 DEFICIENCIES OF THE RIEMANN INTEGRAL

The Riemann integral is the integral of elementary calculus. It is the integral developed intuitively by Newton and Leibnitz and put to great use in the classical sciences. Before undertaking the considerable work of developing the Lebesgue integral, the reader and student need to become acquainted with the deficiencies of the Riemann integral. This will motivate the effort that follows.

First, we review the definition of the Riemann integral of a bounded real-valued function f on a closed, finite interval $[a, b]$ of the real line.

Definition 1.2.1 A partition P is an ordered list of finitely many points starting with a and ending with $b > a$. Thus $P = \{x_0, x_1, \ldots, x_n\}$, where

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

These points are regarded as partitioning $[a, b]$ into n contiguous subintervals, $[x_{i-1}, x_i]$, $i = 1, \ldots, n$. The length of the ith subinterval is given by $\Delta x_i = x_i - x_{i-1}$. The *mesh* of the partition is denoted and defined by

$$\|P\| = \max\{\Delta x_i \mid i = 1, 2, \ldots n\}.$$

Definition 1.2.2 Let f be *any* bounded function on $[a, b]$ and let P be any partition of $[a, b]$. Let

$$M_i = \sup\{f(x) \mid x \in [x_{i-1}, x_i]\} \text{ and } m_i = \inf\{f(x) \mid x \in [x_{i-1}, x_i]\}.$$

Define the upper sum,

$$U(f, P) = \sum_{i=1}^{n} M_i \Delta x_i.$$

and the lower sum,

$$L(f, P) = \sum_{i=1}^{n} m_i \Delta x_i.$$

We say that f is Riemann integrable on $[a, b]$ with $\int_a^b f(x)\,dx = L$ if and only if *both* $L(f, P) \to L$ and $U(f, P) \to L$ as $\|P\| \to 0$.

Note that M_i and m_i are real numbers in Definition 1.2.2 because f is bounded.

■ **EXAMPLE 1.1**

Since the set $\mathbb{Q}$ of rational numbers is countably infinite, the same is true of the set S of all rational numbers in $[a, b]$ for any $a < b$. So, write $S = \{q_n \mid n \in \mathbb{N}\}$. Now define the functions

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{q_1, \ldots, q_n\}, \\ 0 & \text{if } x \in [a, b] \setminus \{q_1, \ldots, q_n\}. \end{cases}$$

It is known from advanced calculus[1] that each function f_n lies in $\mathfrak{R}[a, b]$, the set of Riemann integrable functions on $[a, b]$. In the following exercises, the reader will prove that the pointwise limit of the sequence f_n is not Riemann integrable.

[1]See [20], for example.

EXERCISES

The exercises below refer to the functions f_n in Example 1.1.

1.1 Prove that each function f_n lies in $\mathfrak{R}[a,b]$, the set of Riemann integrable functions on $[a,b]$.

1.2 Prove that for each x in $[a,b]$, $f_n(x) \to \mathbf{1}_S(x)$, where

$$\mathbf{1}_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{if } x \in [a,b]\setminus S, \end{cases}$$

the *indicator function* of the set S of rational numbers in $[a,b]$.

1.3 Prove that the function $\mathbf{1}_S$ is not Riemann integrable. That is,

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx \neq \int_a^b \lim_{n\to\infty} f_n(x)\,dx,$$

because the latter integral does not exist.

The failure of the pointwise limit of a sequence of Riemann integrable functions to be Riemann integrable is considered a serious shortcoming of the Riemann integral. The following example will illustrate a deficiency that is shared by the Riemann integral and the Lebesgue integral that we will define.

■ EXAMPLE 1.2

Let

$$f_n(x) = \begin{cases} n & \text{if } 0 < x \leqslant \frac{1}{n}, \\ 0 & \text{if } \frac{1}{n} < x \leqslant 1, \\ 0 & \text{if } x = 0 \end{cases}$$

for all $n \in \mathbb{N}$. The reader should do the following exercise.

EXERCISE

1.4 Let f_n be as in Example 1.2. Prove that $f_n(x) \to f(x) \equiv 0$ pointwise on $[0,1]$.

Also, it is clear that $f_n \in \mathfrak{R}[0,1]$ for all n, and $f \in \mathfrak{R}[0,1]$ as well. Yet

$$\int_0^1 f_n(x)\,dx \equiv 1 \to 1 \neq 0 = \int_0^1 f(x)\,dx.$$

Thus it occurs for some convergent sequences of functions that

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx \neq \int_a^b \lim_{n\to\infty} f_n(x)\,dx \tag{1.1}$$

even when all the integrals exist. For the Lebesgue integral, however, Theorem 5.3.1 will identify useful conditions under which equality would be guaranteed in Equation (1.1).

EXAMPLE 1.3

Let

$$f_n(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{if } \frac{1}{n} \leqslant x \leqslant 1, \\ 0 & \text{if } 0 \leqslant x < \frac{1}{n} \end{cases}$$

for each $n \in \mathbb{N}$. The reader should check that each f_n is Riemann integrable but that if $f(x) = \lim_{n\to\infty} f_n(x)$, then $f \notin \mathcal{R}[0,1]$ because f is not bounded. The reader should recall from elementary calculus that f is, however, *improperly* Riemann integrable. In Exercise 5.44 the reader will see a generalization of this example that satisfies Lebesgue convergence theorems but that cannot be corrected with improper Riemann integration.

1.3 MOTIVATION FOR THE LEBESGUE INTEGRAL

The Lebesgue integral begins with a seemingly simple reversal of the intuitively appealing process of Definition 1.2.2. Instead of partitioning the interval $[a,b]$ on the x-axis into subintervals and considering the range of values of a bounded function f on each small subinterval, Lebesgue began with the interval $[m,M]$ on the y-axis, where

$$M = \sup\{f(x) \mid x \in [a,b]\}, \text{ and } m = \inf\{f(x) \mid x \in [a,b]\}.$$

Thus $P = \{y_0, y_1, \ldots, y_n\}$, where $m = y_0 < y_1 < \cdots < y_n = M$. Next, instead of forming a sum of the lengths Δx_i of the x-intervals weighted by the heights M_i or m_i, Lebesgue sought to form a sum of the heights, y_i, each weighted by some suitable concept of the length, or *measure* μ, of the set $f^{-1}([y_{i-1}, y_i])$, the set of points x for which $f(x) \in [y_{i-1}, y_i]$. The difficulty is that the set $f^{-1}([y_{i-1}, y_i])$ does not need to be an interval. Indeed, $f^{-1}([y_{i-1}, y_i])$ can be a very complicated subset of the x-axis.[2]

EXERCISE

1.5 Give an example of a real-valued function $f : \mathbb{R} \to \mathbb{R}$ for which

$$f^{-1}\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) = \mathbb{R}\backslash\mathbb{Q},$$

the set of *irrational* numbers.

It turns out that the definition on the real line of the Lebesgue integral—a wonderful improvement upon the Riemann integral—is very simple once one has defined a

[2]The comparison of Riemann with Lebesgue integration has been likened to a story about a smart merchant who sorts money into denominations before counting the day's receipts. Riemann adds the figures as they come in, but Lebesgue sorts first according to values. Lebesgue integration is subtler, however, than this analogy suggests, because the sets $f^{-1}([y_{i-1}, y_i])$ can be very intricate indeed.

suitable concept of the positive real-valued measure of a subset of the line. The desired measure should agree with the concept of *length* when applied to a subset that is an interval. The key property that one needs for a concept of the measure of a set is that if one takes any infinite sequence of *mutually disjoint* sets E_i, one needs to have

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

That is, one needs a *countably additive measure* on subsets of the line which generalizes the concept of length of an interval. Unfortunately, no measure exists that agrees with the concept of the length of an interval and that can be defined on *all* the subsets of the line. Thus it turns out that defining the family of *Lebesgue measurable sets* is a very serious undertaking in the construction of the Lebesgue integral. And that is why we will begin our task in the next chapter with the definition of Lebesgue measurable sets and the definition of Lebesgue measure on those sets. This will turn out to be a *lengthy* task. (Confession: the pun *is* intended.)

We can see in advance how the Lebesgue integral will resolve some of the deficiencies of the Riemann integral. Suppose that we have defined already a countably additive Lebesgue measure that generalizes the concept of the length of an interval in the real line. The set $S = \mathbb{Q} \cap [a, b]$ of Exercise 1.3 is a countably infinite set. That is, the points of S can be arranged into a single infinite sequence: $S = \{s_n \mid n \in \mathbb{N}\}$. Each point is an interval of length zero. Thus it will need to be the case that the Lebesgue measure

$$l(S) = \sum_{n=1}^{\infty} l\{s_n\} = 0.$$

If the reader finds it believable in advance that the Lebesgue integral of a constant function on a measurable set will be that constant times the Lebesgue measure of the set, then

$$\int_a^b 1_S(x)\, dx = 1 \cdot 0 = 0$$

in the sense of Lebesgue integration.

The reader can understand at this point why Lebesgue measure is required to be *only* countably additive. If Lebesgue measure were to be *uncountably* additive,[3] then every set would have measure zero because every set is a disjoint union of singleton sets. Thus the theory of Lebesgue measure would collapse.

It will be seen in the coming chapters that the Lebesgue measure of each interval on the line will be its Euclidean length, that each Riemann integrable function will still be Lebesgue integrable, and that the value of that integral will be unchanged. Thus the reader is advised *not* to forget everything that he or she has learned before!

[3]One could define a concept of the sum of an uncountable family $\{x_a \mid a \in A\}$ of nonnegative real numbers indexed by an uncountable set A. For example, the sum could be taken to mean the supremum of the sums over all countable subsets of A. It is a simple exercise to show that, with this definition, a sum must be infinite unless $x_a = 0$ for all a outside some countable subset of A.

■ **EXAMPLE 1.4**

The fact that the Lebesgue measure of a countable set, such as the set of rational numbers, is zero will resolve another shortcoming of the Riemann integral. Let $S = \mathbb{Q} \cap [0,1] = \{q_n \mid n \in \mathbb{N}\}$, as before. Define the functions

$$f_n(x) = \begin{cases} n & \text{if } x \in \{q_1, \ldots, q_n\}, \\ 0 & \text{if } x \in [0,1] \setminus \{q_1, \ldots, q_n\}. \end{cases}$$

It is easy to see that $f_n(x)$ diverges to ∞ on the dense set S, whereas $f_n(x) \to 0$ at every other value of $x \in [0,1]$. We can define a function

$$f(x) = \begin{cases} \infty & \text{if } x \in S, \\ 0 & \text{if } x \in [0,1] \setminus S. \end{cases}$$

This function f is not real-valued at the points of S—we say that it is *extended* real-valued. But because the set S has Lebesgue measure zero, it will turn out that f is Lebesgue integrable and that $\int_0^1 f(x)\, dx = 0$, in the sense of Lebesgue. Thus we do have

$$\int_0^1 \lim_n f_n(x)\, dx = \lim_n \int_0^1 f_n(x)\, dx,$$

despite the fact that the pointwise limit of f_n exists only in the *extended* real-valued sense. Here we have benefited from the fact that the functions f_n are *uniformly bounded*, except on a set of measure zero. The reader should note that the function f is not even *improperly* Riemann integrable in any plausible sense.

Before proceeding to the task at hand, we explain why it is necessary to develop both Lebesgue measure and the Lebesgue integral for functions mapping a domain X that is an abstract set into the set $\mathbb{R}$ of real numbers. One reason for working at this level of generality, in which X is simply a set (not necessarily a set of real numbers) and $f : X \to \mathbb{R}$, is that it is important to define the Lebesgue integral for functions of several variables. That is, we wish also to be able to integrate $f : \mathbb{R}^n \to \mathbb{R}$. An element of $\mathbb{R}^n$ is not a real number, but rather an n-tuple of real numbers. Moreover, in higher analysis, both pure and applied, it is necessary to work with functions defined on groups, such as the important classical groups of matrices, and this requires knowledge of Lebesgue integration on abstract sets. Moreover, the study of Fourier inversion for functions defined on groups requires the introduction of a measure on what is called the *dual object*[4] of the group, and the Fourier transform must be integrated on that object.

[4]The dual object of a group is the set of all unitary equivalence classes of irreducible unitary representations of the underlying group. When endowed with a usable topology, such objects can be quite complex topologically.

Finally, there is a very important motivation for the abstract study of measure theory from probability. In a *probability model*, the outcomes of an experiment are pictured as points in a so-called *sample space* X. An *event* is conceptualized as a *subset* $E \subseteq X$. The idea behind this is that E denotes the event that the experiment yields a result that is an element of E. For example, X could be the real line. The experiment could be measuring the temperature of the mathematics classroom at 3 P.M. on a certain day. The interval $E = [80, 90]$ would represent the event that the temperature turns out to be between 80° and 90°F.

In probability theory, one wishes very strongly to have a concept of the probability $\mu(E)$ that has the following properties:

- The probability $\mu(E) \in [0, 1]$ for each event E.
- The probability measure μ is additive on all countably infinite sequences of mutually disjoint events.

As discussed above, this cannot be done for all subsets of $\mathbb{R}$—at least not with a measure that generalizes reasonably the length of an interval. Thus for probability also, we must study the concept of measurable sets and the measure of such a set.

The sample space for a probability model need not be a subset of the real line. For example, Brownian motion is the type of motion exhibited by a particle suspended in a fluid. In order to study Brownian motion by means of probability theory, one must place a measure on the set of all possible *paths* that a Brownian motion may follow.[5]

Since the sample space of an experiment could be a set quite different from $\mathbb{R}$, we must develop the theory of measure and integration on abstract sets.

[5]It turns out that these paths are continuous, but they are nowhere differentiable with probability 1!