5.2 Sigma Notation and Limits of Finite Sums

In estimating with finite sums in Section 5.1, we often encountered sums with many terms (up to 1000 in Table 5.1, for instance). In this section we introduce a notation to write sums with a large number of terms. After describing the notation and stating several of its properties, we look at what happens to a finite sum approximation as the number of terms approaches infinity.

Finite Sums and Sigma Notation

Sigma notation enables us to write a sum with many terms in the compact form

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n .$$

The Greek letter Σ (capital sigma, corresponding to our letter S), stands for "sum." The **index of summation** k tells us where the sum begins (at the number below the Σ symbol) and where it ends (at the number above Σ). Any letter can be used to denote the index, but the letters i, j, and k are customary.

The index k ends at $k = n$.

The summation symbol (Greek letter sigma) — $\displaystyle\sum_{k=1}^{n} a_k$ — a_k is a formula for the kth term.

The index k starts at $k = 1$.

Thus we can write

$$1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 + 11^2 = \sum_{k=1}^{11} k^2,$$

and

$$f(1) + f(2) + f(3) + \cdots + f(100) = \sum_{i=1}^{100} f(i).$$

The sigma notation used on the right side of these equations is much more compact than the summation expressions on the left side.

You Try It

EXAMPLE 1 Using Sigma Notation

The sum in sigma notation	**The sum written out, one term for each value of k**	**The value of the sum**
$\sum_{k=1}^{5} k$	$1 + 2 + 3 + 4 + 5$	15
$\sum_{k=1}^{3} (-1)^k k$	$(-1)^1(1) + (-1)^2(2) + (-1)^3(3)$	$-1 + 2 - 3 = -2$
$\sum_{k=1}^{2} \frac{k}{k+1}$	$\frac{1}{1+1} + \frac{2}{2+1}$	$\frac{1}{2} + \frac{2}{3} = \frac{7}{6}$
$\sum_{k=4}^{5} \frac{k^2}{k-1}$	$\frac{4^2}{4-1} + \frac{5^2}{5-1}$	$\frac{16}{3} + \frac{25}{4} = \frac{139}{12}$

■

The lower limit of summation does not have to be 1; it can be any integer.

EXAMPLE 2 Using Different Index Starting Values

Express the sum $1 + 3 + 5 + 7 + 9$ in sigma notation.

Solution The formula generating the terms changes with the lower limit of summation, but the terms generated remain the same. It is often simplest to start with $k = 0$ or $k = 1$.

$$\textit{Starting with } k = 0: \qquad 1 + 3 + 5 + 7 + 9 = \sum_{k=0}^{4} (2k + 1)$$

$$\textit{Starting with } k = 1: \qquad 1 + 3 + 5 + 7 + 9 = \sum_{k=1}^{5} (2k - 1)$$

$$\textit{Starting with } k = 2: \qquad 1 + 3 + 5 + 7 + 9 = \sum_{k=2}^{6} (2k - 3)$$

$$\textit{Starting with } k = -3: \qquad 1 + 3 + 5 + 7 + 9 = \sum_{k=-3}^{1} (2k + 7)$$

■

When we have a sum such as

$$\sum_{k=1}^{3} (k + k^2)$$

we can rearrange its terms,

$$\begin{aligned} \sum_{k=1}^{3} (k + k^2) &= (1 + 1^2) + (2 + 2^2) + (3 + 3^2) \\ &= (1 + 2 + 3) + (1^2 + 2^2 + 3^2) \qquad \text{Regroup terms.} \\ &= \sum_{k=1}^{3} k + \sum_{k=1}^{3} k^2 \end{aligned}$$

This illustrates a general rule for finite sums:

$$\sum_{k=1}^{n} (a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$$

Four such rules are given below. A proof that they are valid can be obtained using mathematical induction (see Appendix 1).

Algebra Rules for Finite Sums

1. *Sum Rule*: $\displaystyle\sum_{k=1}^{n} (a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$

2. *Difference Rule*: $\displaystyle\sum_{k=1}^{n} (a_k - b_k) = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} b_k$

3. *Constant Multiple Rule*: $\displaystyle\sum_{k=1}^{n} ca_k = c \cdot \sum_{k=1}^{n} a_k$ (Any number c)

4. *Constant Value Rule*: $\displaystyle\sum_{k=1}^{n} c = n \cdot c$ (c is any constant value.)

EXAMPLE 3 Using the Finite Sum Algebra Rules

(a) $$\sum_{k=1}^{n}(3k - k^2) = 3\sum_{k=1}^{n}k - \sum_{k=1}^{n}k^2$$ Difference Rule and Constant Multiple Rule

(b) $$\sum_{k=1}^{n}(-a_k) = \sum_{k=1}^{n}(-1)\cdot a_k = -1\cdot\sum_{k=1}^{n}a_k = -\sum_{k=1}^{n}a_k$$ Constant Multiple Rule

(c) $$\sum_{k=1}^{3}(k + 4) = \sum_{k=1}^{3}k + \sum_{k=1}^{3}4$$ Sum Rule

$$= (1 + 2 + 3) + (3\cdot 4)$$ Constant Value Rule

$$= 6 + 12 = 18$$

(d) $$\sum_{k=1}^{n}\frac{1}{n} = n\cdot\frac{1}{n} = 1$$ Constant Value Rule ($1/n$ is constant) ■

HISTORICAL BIOGRAPHY

Carl Friedrich Gauss
(1777–1855)

Over the years people have discovered a variety of formulas for the values of finite sums. The most famous of these are the formula for the sum of the first n integers (Gauss may have discovered it at age 8) and the formulas for the sums of the squares and cubes of the first n integers.

EXAMPLE 4 The Sum of the First n Integers

Show that the sum of the first n integers is

$$\sum_{k=1}^{n}k = \frac{n(n + 1)}{2}.$$

Solution: The formula tells us that the sum of the first 4 integers is

$$\frac{(4)(5)}{2} = 10.$$

Addition verifies this prediction:

$$1 + 2 + 3 + 4 = 10.$$

To prove the formula in general, we write out the terms in the sum twice, once forward and once backward.

$$\begin{array}{ccccccccc} 1 & + & 2 & + & 3 & + & \cdots & + & n \\ n & + & (n - 1) & + & (n - 2) & + & \cdots & + & 1 \end{array}$$

If we add the two terms in the first column we get $1 + n = n + 1$. Similarly, if we add the two terms in the second column we get $2 + (n - 1) = n + 1$. The two terms in any column sum to $n + 1$. When we add the n columns together we get n terms, each equal to $n + 1$, for a total of $n(n + 1)$. Since this is twice the desired quantity, the sum of the first n integers is $(n)(n + 1)/2$. ■

Formulas for the sums of the squares and cubes of the first n integers are proved using mathematical induction (see Appendix 1). We state them here.

The first n squares: $$\sum_{k=1}^{n}k^2 = \frac{n(n + 1)(2n + 1)}{6}$$

The first n cubes: $$\sum_{k=1}^{n}k^3 = \left(\frac{n(n + 1)}{2}\right)^2$$

Limits of Finite Sums

The finite sum approximations we considered in Section 5.1 got more accurate as the number of terms increased and the subinterval widths (lengths) became thinner. The next example shows how to calculate a limiting value as the widths of the subintervals go to zero and their number grows to infinity.

EXAMPLE 5 The Limit of Finite Approximations to an Area

Find the limiting value of lower sum approximations to the area of the region R below the graph of $y = 1 - x^2$ and above the interval [0, 1] on the x-axis using equal width rectangles whose widths approach zero and whose number approaches infinity. (See Figure 5.4a.)

Solution We compute a lower sum approximation using n rectangles of equal width $\Delta x = (1 - 0)/n$, and then we see what happens as $n \to \infty$. We start by subdividing [0, 1] into n equal width subintervals

$$\left[0, \frac{1}{n}\right], \left[\frac{1}{n}, \frac{2}{n}\right], \dots, \left[\frac{n-1}{n}, n\right].$$

Each subinterval has width $1/n$. The function $1 - x^2$ is decreasing on [0, 1], and its smallest value in a subinterval occurs at the subinterval's right endpoint. So a lower sum is constructed with rectangles whose height over the subinterval $[(k-1)/n, k/n]$ is $f(k/n) = 1 - (k/n)^2$, giving the sum

$$f\left(\frac{1}{n}\right)\left(\frac{1}{n}\right) + f\left(\frac{2}{n}\right)\left(\frac{1}{n}\right) + \cdots + f\left(\frac{k}{n}\right)\left(\frac{1}{n}\right) + \cdots + f\left(\frac{n}{n}\right)\left(\frac{1}{n}\right).$$

We write this in sigma notation and simplify,

$$\begin{aligned}
\sum_{k=1}^{n} f\left(\frac{k}{n}\right)\left(\frac{1}{n}\right) &= \sum_{k=1}^{n}\left(1 - \left(\frac{k}{n}\right)^2\right)\left(\frac{1}{n}\right) \\
&= \sum_{k=1}^{n}\left(\frac{1}{n} - \frac{k^2}{n^3}\right) \\
&= \sum_{k=1}^{n}\frac{1}{n} - \sum_{k=1}^{n}\frac{k^2}{n^3} && \text{Difference Rule} \\
&= n \cdot \frac{1}{n} - \frac{1}{n^3}\sum_{k=1}^{n} k^2 && \text{Constant Value and Constant Multiple Rules} \\
&= 1 - \left(\frac{1}{n^3}\right)\frac{(n)(n+1)(2n+1)}{6} && \text{Sum of the First } n \text{ Squares} \\
&= 1 - \frac{2n^3 + 3n^2 + n}{6n^3}. && \text{Numerator expanded}
\end{aligned}$$

We have obtained an expression for the lower sum that holds for any n. Taking the limit of this expression as $n \to \infty$, we see that the lower sums converge as the number of subintervals increases and the subinterval widths approach zero:

$$\lim_{n \to \infty}\left(1 - \frac{2n^3 + 3n^2 + n}{6n^3}\right) = 1 - \frac{2}{6} = \frac{2}{3}.$$

The lower sum approximations converge to 2/3. A similar calculation shows that the upper sum approximations also converge to 2/3 (Exercise 35). Any finite sum approximation, in the sense of our summary at the end of Section 5.1, also converges to the same value

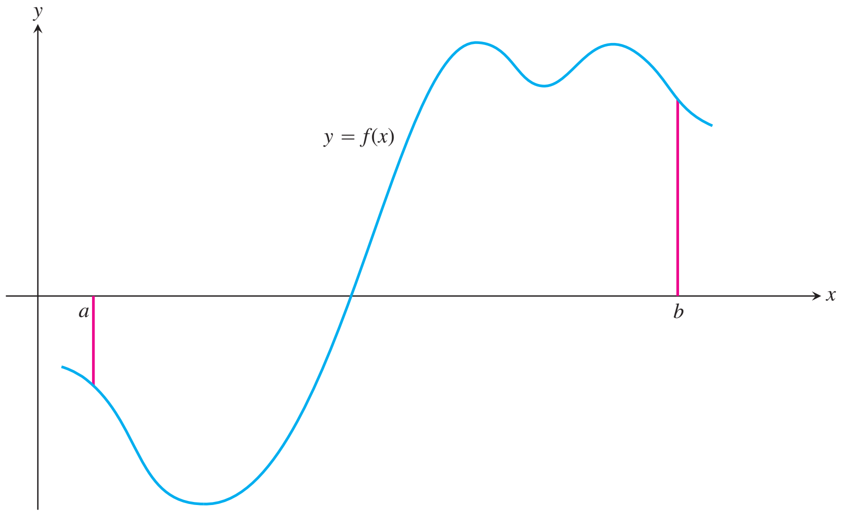


FIGURE 5.8 A typical continuous function $y = f(x)$ over a closed interval $[a, b]$.

2/3. This is because it is possible to show that any finite sum approximation is trapped between the lower and upper sum approximations. For this reason we are led to *define* the area of the region R as this limiting value. In Section 5.3 we study the limits of such finite approximations in their more general setting. ■

Riemann Sums

HISTORICAL BIOGRAPHY

Georg Friedrich
Bernhard Riemann
(1826–1866)

Biography

The theory of limits of finite approximations was made precise by the German mathematician Bernhard Riemann. We now introduce the notion of a *Riemann sum*, which underlies the theory of the definite integral studied in the next section.

We begin with an arbitrary function f defined on a closed interval $[a, b]$. Like the function pictured in Figure 5.8, f may have negative as well as positive values. We subdivide the interval $[a, b]$ into subintervals, not necessarily of equal widths (or lengths), and form sums in the same way as for the finite approximations in Section 5.1. To do so, we choose $n - 1$ points $\{x_1, x_2, x_3, \ldots, x_{n-1}\}$ between a and b and satisfying

$$a < x_1 < x_2 < \cdots < x_{n-1} < b.$$

To make the notation consistent, we denote a by x_0 and b by x_n, so that

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

The set

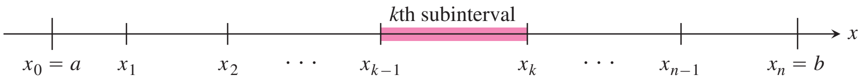
$$P = \{x_0, x_1, x_2, \ldots, x_{n-1}, x_n\}$$

is called a **partition** of $[a, b]$.

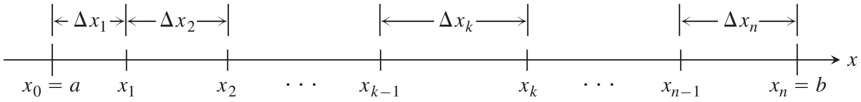
The partition P divides $[a, b]$ into n closed subintervals

$$[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n].$$

The first of these subintervals is $[x_0, x_1]$, the second is $[x_1, x_2]$, and the ***k*th subinterval of** P is $[x_{k-1}, x_k]$, for k an integer between 1 and n.



The width of the first subinterval $[x_0, x_1]$ is denoted Δx_1, the width of the second $[x_1, x_2]$ is denoted Δx_2, and the width of the kth subinterval is $\Delta x_k = x_k - x_{k-1}$. If all n subintervals have equal width, then the common width Δx is equal to $(b - a)/n$.



In each subinterval we select some point. The point chosen in the kth subinterval $[x_{k-1}, x_k]$ is called c_k. Then on each subinterval we stand a vertical rectangle that stretches from the x-axis to touch the curve at $(c_k, f(c_k))$. These rectangles can be above or below the x-axis, depending on whether $f(c_k)$ is positive or negative, or on it if $f(c_k) = 0$ (Figure 5.9).

On each subinterval we form the product $f(c_k) \cdot \Delta x_k$. This product is positive, negative or zero, depending on the sign of $f(c_k)$. When $f(c_k) > 0$, the product $f(c_k) \cdot \Delta x_k$ is the area of a rectangle with height $f(c_k)$ and width Δx_k. When $f(c_k) < 0$, the product $f(c_k) \cdot \Delta x_k$ is a negative number, the negative of the area of a rectangle of width Δx_k that drops from the x-axis to the negative number $f(c_k)$.

Finally we sum all these products to get

$$S_P = \sum_{k=1}^{n} f(c_k)\, \Delta x_k .$$

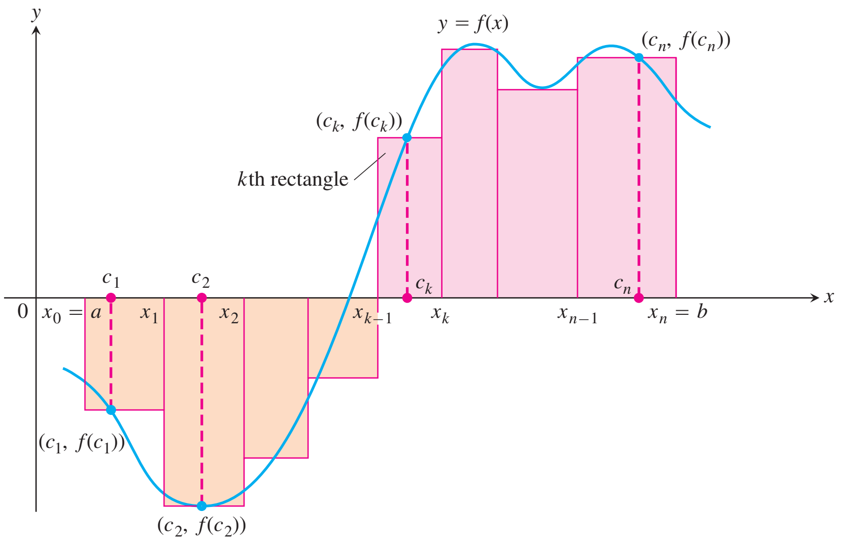


FIGURE 5.9 The rectangles approximate the region between the graph of the function $y = f(x)$ and the x-axis.

The sum S_P is called a **Riemann sum for *f* on the interval [*a*, *b*]**. There are many such sums, depending on the partition P we choose, and the choices of the points c_k in the subintervals.

In Example 5, where the subintervals all had equal widths $\Delta x = 1/n$, we could make them thinner by simply increasing their number n. When a partition has subintervals of varying widths, we can ensure they are all thin by controlling the width of a widest (longest) subinterval. We define the **norm** of a partition P, written $\|P\|$, to be the largest of all the subinterval widths. If $\|P\|$ is a small number, then all of the subintervals in the partition P have a small width. Let's look at an example of these ideas.

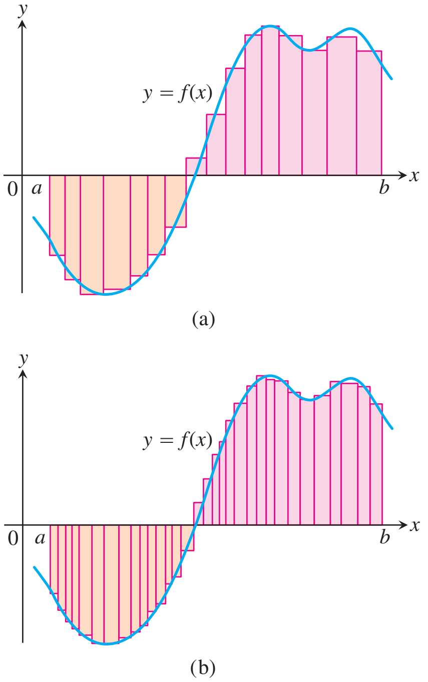
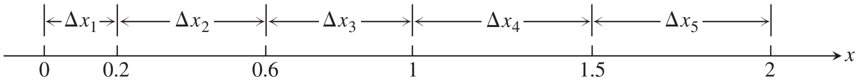


FIGURE 5.10 The curve of Figure 5.9 with rectangles from finer partitions of [*a*, *b*]. Finer partitions create collections of rectangles with thinner bases that approximate the region between the graph of *f* and the *x*-axis with increasing accuracy.

EXAMPLE 6 Partitioning a Closed Interval

The set $P = \{0, 0.2, 0.6, 1, 1.5, 2\}$ is a partition of $[0, 2]$. There are five subintervals of P: $[0, 0.2]$, $[0.2, 0.6]$, $[0.6, 1]$, $[1, 1.5]$, and $[1.5, 2]$:



The lengths of the subintervals are $\Delta x_1 = 0.2$, $\Delta x_2 = 0.4$, $\Delta x_3 = 0.4$, $\Delta x_4 = 0.5$, and $\Delta x_5 = 0.5$. The longest subinterval length is 0.5, so the norm of the partition is $\|P\| = 0.5$. In this example, there are two subintervals of this length. ■

Any Riemann sum associated with a partition of a closed interval $[a, b]$ defines rectangles that approximate the region between the graph of a continuous function f and the x-axis. Partitions with norm approaching zero lead to collections of rectangles that approximate this region with increasing accuracy, as suggested by Figure 5.10. We will see in the next section that if the function f is continuous over the closed interval $[a, b]$, then no matter how we choose the partition P and the points c_k in its subintervals to construct a Riemann sum, a single limiting value is approached as the subinterval widths, controlled by the norm of the partition, approach zero.