5.3 The Definite Integral

In Section 5.2 we investigated the limit of a finite sum for a function defined over a closed interval $[a, b]$ using n subintervals of equal width (or length), $(b - a)/n$. In this section we consider the limit of more general Riemann sums as the norm of the partitions of $[a, b]$ approaches zero. For general Riemann sums the subintervals of the partitions need not have equal widths. The limiting process then leads to the definition of the *definite integral* of a function over a closed interval $[a, b]$.

Limits of Riemann Sums

The definition of the definite integral is based on the idea that for certain functions, as the norm of the partitions of $[a, b]$ approaches zero, the values of the corresponding Riemann sums approach a limiting value I. What we mean by this converging idea is that a Riemann sum will be close to the number I provided that the norm of its partition is sufficiently small (so that all of its subintervals have thin enough widths). We introduce the symbol ϵ as a small positive number that specifies how close to I the Riemann sum must be, and the symbol δ as a second small positive number that specifies how small the norm of a partition must be in order for that to happen. Here is a precise formulation.

DEFINITION The Definite Integral as a Limit of Riemann Sums

Let $f(x)$ be a function defined on a closed interval $[a, b]$. We say that a number I is the **definite integral of f over $[a, b]$** and that I is the limit of the Riemann sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$ if the following condition is satisfied:

Given any number $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that for every partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ with $\|P\| < \delta$ and any choice of c_k in $[x_{k-1}, x_k]$, we have

$$\left| \sum_{k=1}^{n} f(c_k)\,\Delta x_k - I \right| < \epsilon .$$

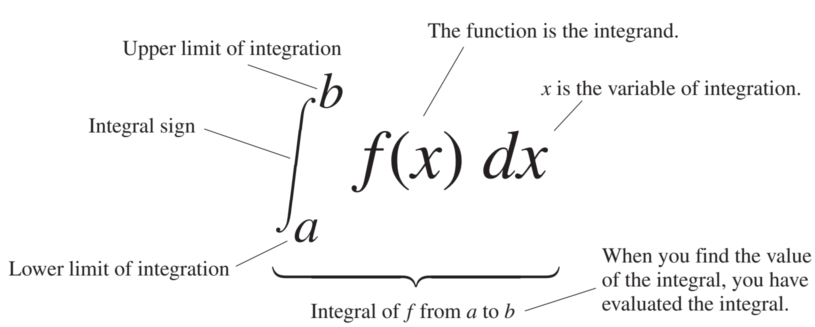
Leibniz introduced a notation for the definite integral that captures its construction as a limit of Riemann sums. He envisioned the finite sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$ becoming an infinite sum of function values $f(x)$ multiplied by "infinitesimal" subinterval widths dx. The sum symbol $\sum$ is replaced in the limit by the integral symbol $\int$, whose origin is in the letter "S." The function values $f(c_k)$ are replaced by a continuous selection of function values $f(x)$. The subinterval widths Δx_k become the differential dx. It is as if we are summing all products of the form $f(x) \cdot dx$ as x goes from a to b. While this notation captures the process of constructing an integral, it is Riemann's definition that gives a precise meaning to the definite integral.

Notation and Existence of the Definite Integral

The symbol for the number I in the definition of the definite integral is

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

which is read as "the integral from a to b of f of x dee x" or sometimes as "the integral from a to b of f of x with respect to x." The component parts in the integral symbol also have names:



When the definition is satisfied, we say the Riemann sums of f on $[a, b]$ **converge** to the definite integral $I = \int_a^b f(x)\,dx$ and that f is **integrable** over $[a, b]$. We have many choices for a partition P with norm going to zero, and many choices of points c_k for each partition. The definite integral exists when we always get the same limit I, no matter what choices are made. When the limit exists we write it as the definite integral

You Try It

$$\lim_{\|P\|\to 0} \sum_{k=1}^{n} f(c_k)\,\Delta x_k = I = \int_a^b f(x)\,dx.$$

When each partition has n equal subintervals, each of width $\Delta x = (b - a)/n$, we will also write

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(c_k)\,\Delta x = I = \int_a^b f(x)\,dx.$$

The limit is always taken as the norm of the partitions approaches zero and the number of subintervals goes to infinity.

The value of the definite integral of a function over any particular interval depends on the function, not on the letter we choose to represent its independent variable. If we decide to use t or u instead of x, we simply write the integral as

$$\int_a^b f(t)\,dt \qquad \text{or} \qquad \int_a^b f(u)\,du \qquad \text{instead of} \qquad \int_a^b f(x)\,dx.$$

No matter how we write the integral, it is still the same number, defined as a limit of Riemann sums. Since it does not matter what letter we use, the variable of integration is called a **dummy variable**.

Since there are so many choices to be made in taking a limit of Riemann sums, it might seem difficult to show that such a limit exists. It turns out, however, that no matter what choices are made, the Riemann sums associated with a *continuous* function converge to the same limit.

THEOREM 1 The Existence of Definite Integrals

A continuous function is integrable. That is, if a function f is continuous on an interval $[a, b]$, then its definite integral over $[a, b]$ exists.

By the Extreme Value Theorem (Theorem 1, Section 4.1), when f is continuous we can choose c_k so that $f(c_k)$ gives the maximum value of f on $[x_{k-1}, x_k]$, giving an **upper sum**. We can choose c_k to give the minimum value of f on $[x_{k-1}, x_k]$, giving a **lower sum**. We can pick c_k to be the midpoint of $[x_{k-1}, x_k]$, the rightmost point x_k, or a random point. We can take the partitions of equal or varying widths. In each case we get the same limit for $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$ as $\|P\| \to 0$. The idea behind Theorem 1 is that a Riemann sum associated with a partition is no more than the upper sum of that partition and no less than the lower sum. The upper and lower sums converge to the same value when $\|P\| \to 0$. All other Riemann sums lie between the upper and lower sums and have the same limit. A proof of Theorem 1 involves a careful analysis of functions, partitions, and limits along this line of thinking and is left to a more advanced text. An indication of this proof is given in Exercises 80 and 81.

Theorem 1 says nothing about how to *calculate* definite integrals. A method of calculation will be developed in Section 5.4, through a connection to the process of taking antiderivatives.

Integrable and Nonintegrable Functions

Theorem 1 tells us that functions continuous over the interval $[a, b]$ are integrable there. Functions that are not continuous may or may not be integrable. Discontinuous functions that are integrable include those that are increasing on $[a, b]$ (Exercise 77), and the *piecewise-continuous functions* defined in the Additional Exercises at the end of this chapter. (The latter are continuous except at a finite number of points in $[a, b]$.) For integrability to fail, a function needs to be sufficiently discontinuous so that the region between its graph and the x-axis cannot be approximated well by increasingly thin rectangles. Here is an example of a function that is not integrable.

EXAMPLE 1 A Nonintegrable Function on [0, 1]

The function

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational} \end{cases}$$

has no Riemann integral over $[0, 1]$. Underlying this is the fact that between any two numbers there is both a rational number and an irrational number. Thus the function jumps up and down too erratically over $[0, 1]$ to allow the region beneath its graph and above the x-axis to be approximated by rectangles, no matter how thin they are. We show, in fact, that upper sum approximations and lower sum approximations converge to different limiting values.

If we pick a partition P of $[0, 1]$ and choose c_k to be the maximum value for f on $[x_{k-1}, x_k]$ then the corresponding Riemann sum is

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

since each subinterval $[x_{k-1}, x_k]$ contains a rational number where $f(c_k) = 1$. Note that the lengths of the intervals in the partition sum to 1, $\sum_{k=1}^{n} \Delta x_k = 1$. So each such Riemann sum equals 1, and a limit of Riemann sums using these choices equals 1.

On the other hand, if we pick c_k to be the minimum value for f on $[x_{k-1}, x_k]$, then the Riemann sum is

$$L = \sum_{k=1}^{n} f(c_k)\,\Delta x_k = \sum_{k=1}^{n} (0)\,\Delta x_k = 0,$$

since each subinterval $[x_{k-1}, x_k]$ contains an irrational number c_k where $f(c_k) = 0$. The limit of Riemann sums using these choices equals zero. Since the limit depends on the choices of c_k, the function f is not integrable. ■

Properties of Definite Integrals

In defining $\int_a^b f(x)\,dx$ as a limit of sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$, we moved from left to right across the interval $[a, b]$. What would happen if we instead move right to left, starting with $x_0 = b$ and ending at $x_n = a$. Each Δx_k in the Riemann sum would change its sign, with $x_k - x_{k-1}$ now negative instead of positive. With the same choices of c_k in each subinterval, the sign of any Riemann sum would change, as would the sign of the limit, the integral

$\int_b^a f(x)\,dx$. Since we have not previously given a meaning to integrating backward, we are led to define

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

Another extension of the integral is to an interval of zero width, when $a = b$. Since $f(c_k)\,\Delta x_k$ is zero when the interval width $\Delta x_k = 0$, we define

$$\int_a^a f(x)\,dx = 0.$$

Theorem 2 states seven properties of integrals, given as rules that they satisfy, including the two above. These rules become very useful in the process of computing integrals. We will refer to them repeatedly to simplify our calculations.

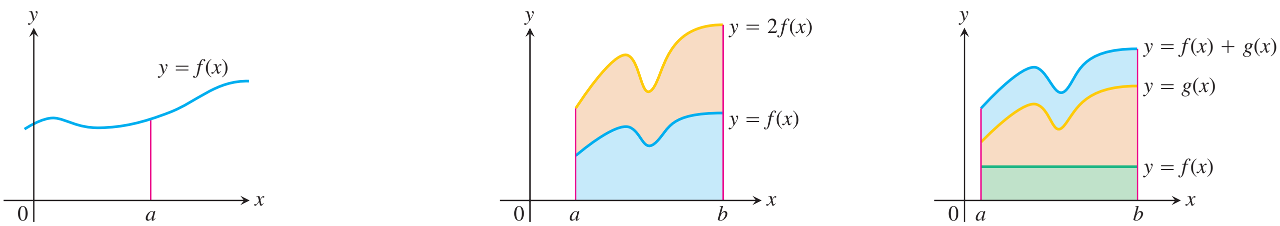
Rules 2 through 7 have geometric interpretations, shown in Figure 5.11. The graphs in these figures are of positive functions, but the rules apply to general integrable functions.

THEOREM 2

When f and g are integrable, the definite integral satisfies Rules 1 to 7 in Table 5.3.

TABLE 5.3 Rules satisfied by definite integrals

1.	*Order of Integration:*	$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$	A Definition
2.	*Zero Width Interval:*	$\int_a^a f(x)\,dx = 0$	Also a Definition
3.	*Constant Multiple:*	$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$	Any Number k
		$\int_a^b -f(x)\,dx = -\int_a^b f(x)\,dx$	$k = -1$
4.	*Sum and Difference:*	$\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$	
5.	*Additivity:*	$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$	
6.	*Max-Min Inequality:*	If f has maximum value max f and minimum value min f on $[a, b]$, then $\min f \cdot (b - a) \le \int_a^b f(x)\,dx \le \max f \cdot (b - a)$.	
7.	*Domination:*	$f(x) \ge g(x)$ on $[a, b] \Rightarrow \int_a^b f(x)\,dx \ge \int_a^b g(x)\,dx$	
		$f(x) \ge 0$ on $[a, b] \Rightarrow \int_a^b f(x)\,dx \ge 0$	(Special Case)



(a) *Zero Width Interval:*

$$\int_a^a f(x)\,dx = 0.$$

(The area over a point is 0.)

(b) *Constant Multiple:*

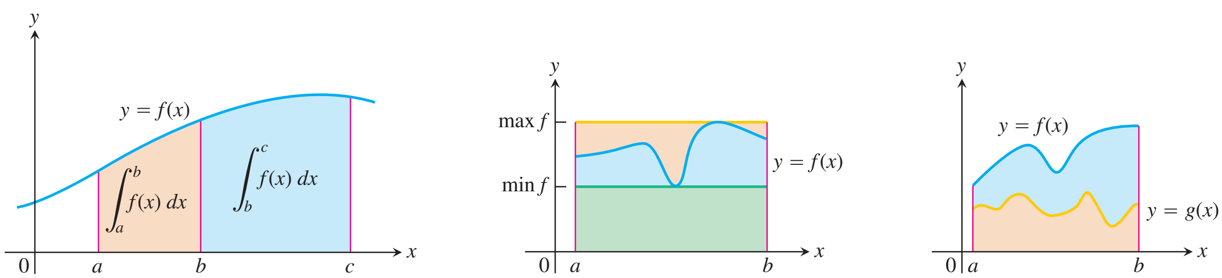
$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx.$$

(Shown for $k = 2$.)

(c) *Sum:*

$$\int_a^b (f(x) + g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$

(Areas add)



(d) *Additivity for definite integrals:*

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$$

(e) *Max-Min Inequality*:

$$\min f \cdot (b - a) \le \int_a^b f(x)\,dx \le \max f \cdot (b - a)$$

(f) *Domination:*

$f(x) \ge g(x)$ on $[a, b]$

$$\Rightarrow \int_a^b f(x)\,dx \ge \int_a^b g(x)\,dx$$

FIGURE 5.11

While Rules 1 and 2 are definitions, Rules 3 to 7 of Table 5.3 must be proved. The proofs are based on the definition of the definite integral as a limit of Riemann sums. The following is a proof of one of these rules. Similar proofs can be given to verify the other properties in Table 5.3.

Proof of Rule 6 Rule 6 says that the integral of f over $[a, b]$ is never smaller than the minimum value of f times the length of the interval and never larger than the maximum value of f times the length of the interval. The reason is that for every partition of $[a, b]$ and for every choice of the points c_k,

$$\begin{aligned}
\min f \cdot (b - a) &= \min f \cdot \sum_{k=1}^{n} \Delta x_k && \sum_{k=1}^{n} \Delta x_k = b - a \\
&= \sum_{k=1}^{n} \min f \cdot \Delta x_k && \text{Constant Multiple Rule} \\
&\le \sum_{k=1}^{n} f(c_k)\,\Delta x_k && \min f \le f(c_k) \\
&\le \sum_{k=1}^{n} \max f \cdot \Delta x_k && f(c_k) \le \max f \\
&= \max f \cdot \sum_{k=1}^{n} \Delta x_k && \text{Constant Multiple Rule} \\
&= \max f \cdot (b - a).
\end{aligned}$$

In short, all Riemann sums for f on $[a, b]$ satisfy the inequality

$$\min f \cdot (b - a) \le \sum_{k=1}^{n} f(c_k)\,\Delta x_k \le \max f \cdot (b - a).$$

Hence their limit, the integral, does too. ■

EXAMPLE 2 Using the Rules for Definite Integrals

Suppose that

$$\int_{-1}^{1} f(x)\,dx = 5, \qquad \int_{1}^{4} f(x)\,dx = -2, \qquad \int_{-1}^{1} h(x)\,dx = 7.$$

Then

1. $\displaystyle\int_{4}^{1} f(x)\,dx = -\int_{1}^{4} f(x)\,dx = -(-2) = 2$ — Rule 1

2. $\displaystyle\int_{-1}^{1} [2f(x) + 3h(x)]\,dx = 2\int_{-1}^{1} f(x)\,dx + 3\int_{-1}^{1} h(x)\,dx$ — Rules 3 and 4

$$= 2(5) + 3(7) = 31$$

3. $\displaystyle\int_{-1}^{4} f(x)\,dx = \int_{-1}^{1} f(x)\,dx + \int_{1}^{4} f(x)\,dx = 5 + (-2) = 3$ — Rule 5 ■

EXAMPLE 3 Finding Bounds for an Integral

Show that the value of $\int_0^1 \sqrt{1 + \cos x}\,dx$ is less than 3/2.

Solution The Max-Min Inequality for definite integrals (Rule 6) says that $\min f \cdot (b - a)$ is a *lower bound* for the value of $\int_a^b f(x)\,dx$ and that $\max f \cdot (b - a)$ is an *upper bound*. The maximum value of $\sqrt{1 + \cos x}$ on $[0, 1]$ is $\sqrt{1 + 1} = \sqrt{2}$, so

$$\int_0^1 \sqrt{1 + \cos x}\,dx \le \sqrt{2} \cdot (1 - 0) = \sqrt{2}.$$

Since $\int_0^1 \sqrt{1 + \cos x}\,dx$ is bounded from above by $\sqrt{2}$ (which is 1.414 . . .), the integral is less than $3/2$. ■

Area Under the Graph of a Nonnegative Function

We now make precise the notion of the area of a region with curved boundary, capturing the idea of approximating a region by increasingly many rectangles. The area under the graph of a nonnegative continuous function is defined to be a definite integral.

DEFINITION **Area Under a Curve as a Definite Integral**

If $y = f(x)$ is nonnegative and integrable over a closed interval $[a, b]$, then the **area under the curve $y = f(x)$ over $[a, b]$** is the integral of f from a to b,

$$A = \int_a^b f(x)\,dx.$$

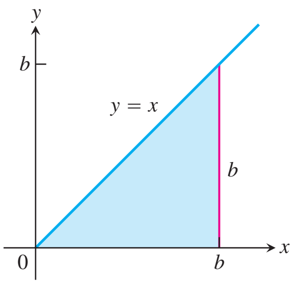


FIGURE 5.12 The region in Example 4 is a triangle.

For the first time we have a rigorous definition for the area of a region whose boundary is the graph of any continuous function. We now apply this to a simple example, the area under a straight line, where we can verify that our new definition agrees with our previous notion of area.

You Try It

Video

EXAMPLE 4 Area Under the Line $y = x$

Compute $\int_0^b x\,dx$ and find the area A under $y = x$ over the interval $[0, b]$, $b > 0$.

Solution The region of interest is a triangle (Figure 5.12). We compute the area in two ways.

(a) To compute the definite integral as the limit of Riemann sums, we calculate $\lim_{\|P\|\to 0} \sum_{k=1}^{n} f(c_k)\,\Delta x_k$ for partitions whose norms go to zero. Theorem 1 tells us that it does not matter how we choose the partitions or the points c_k as long as the norms approach zero. All choices give the exact same limit. So we consider the partition P that subdivides the interval $[0, b]$ into n subintervals of equal width $\Delta x = (b - 0)/n = b/n$, and we choose c_k to be the right endpoint in each subinterval. The partition is $P = \left\{0, \frac{b}{n}, \frac{2b}{n}, \frac{3b}{n}, \cdots, \frac{nb}{n}\right\}$ and $c_k = \frac{kb}{n}$. So

$$\sum_{k=1}^{n} f(c_k)\,\Delta x = \sum_{k=1}^{n} \frac{kb}{n}\cdot\frac{b}{n} \qquad f(c_k) = c_k$$

$$= \sum_{k=1}^{n} \frac{kb^2}{n^2}$$

$$= \frac{b^2}{n^2}\sum_{k=1}^{n} k \qquad \text{Constant Multiple Rule}$$

$$= \frac{b^2}{n^2}\cdot\frac{n(n+1)}{2} \qquad \text{Sum of First } n \text{ Integers}$$

$$= \frac{b^2}{2}\left(1 + \frac{1}{n}\right)$$

As $n \to \infty$ and $\|P\| \to 0$, this last expression on the right has the limit $b^2/2$. Therefore,

$$\int_0^b x\,dx = \frac{b^2}{2}.$$

(b) Since the area equals the definite integral for a nonnegative function, we can quickly derive the definite integral by using the formula for the area of a triangle having base length b and height $y = b$. The area is $A = (1/2)\,b \cdot b = b^2/2$. Again we have that $\int_0^b x\,dx = b^2/2$. ■

Example 4 can be generalized to integrate $f(x) = x$ over any closed interval $[a, b]$, $0 < a < b$.

$$\int_a^b x\,dx = \int_a^0 x\,dx + \int_0^b x\,dx \qquad \text{Rule 5}$$

$$= -\int_0^a x\,dx + \int_0^b x\,dx \qquad \text{Rule 1}$$

$$= -\frac{a^2}{2} + \frac{b^2}{2}. \qquad \text{Example 4}$$

In conclusion, we have the following rule for integrating $f(x) = x$:

$$\int_a^b x\,dx = \frac{b^2}{2} - \frac{a^2}{2}, \qquad a < b \tag{1}$$

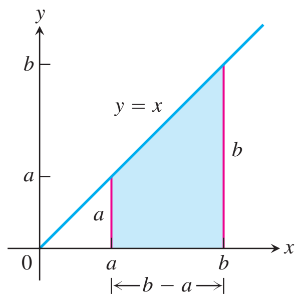


FIGURE 5.13 The area of this trapezoidal region is $A = (b^2 - a^2)/2$.

This computation gives the area of a trapezoid (Figure 5.13). Equation (1) remains valid when a and b are negative. When $a < b < 0$, the definite integral value $(b^2 - a^2)/2$ is a negative number, the negative of the area of a trapezoid dropping down to the line $y = x$ below the x-axis. When $a < 0$ and $b > 0$, Equation (1) is still valid and the definite integral gives the difference between two areas, the area under the graph and above $[0, b]$ minus the area below $[a, 0]$ and over the graph.

The following results can also be established using a Riemann sum calculation similar to that in Example 4 (Exercises 75 and 76).

$$\int_a^b c\,dx = c(b - a), \qquad c \text{ any constant} \tag{2}$$

$$\int_a^b x^2\,dx = \frac{b^3}{3} - \frac{a^3}{3}, \qquad a < b \tag{3}$$

Average Value of a Continuous Function Revisited

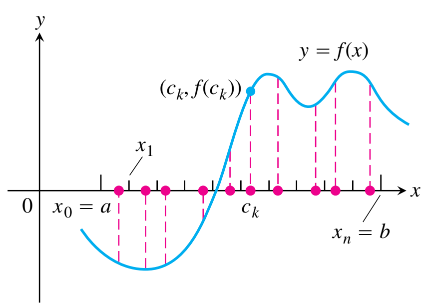


FIGURE 5.14 A sample of values of a function on an interval $[a, b]$.

In Section 5.1 we introduced informally the average value of a nonnegative continuous function f over an interval $[a, b]$, leading us to define this average as the area under the graph of $y = f(x)$ divided by $b - a$. In integral notation we write this as

$$\text{Average} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

We can use this formula to give a precise definition of the average value of any continuous (or integrable) function, whether positive, negative or both.

Alternately, we can use the following reasoning. We start with the idea from arithmetic that the average of n numbers is their sum divided by n. A continuous function f on $[a, b]$ may have infinitely many values, but we can still sample them in an orderly way. We divide $[a, b]$ into n subintervals of equal width $\Delta x = (b - a)/n$ and evaluate f at a point c_k in each (Figure 5.14). The average of the n sampled values is

$$\begin{aligned}\frac{f(c_1) + f(c_2) + \cdots + f(c_n)}{n} &= \frac{1}{n}\sum_{k=1}^{n} f(c_k)\\ &= \frac{\Delta x}{b-a}\sum_{k=1}^{n} f(c_k) \qquad \Delta x = \frac{b-a}{n},\ \text{so}\ \frac{1}{n} = \frac{\Delta x}{b-a}\\ &= \frac{1}{b-a}\sum_{k=1}^{n} f(c_k)\,\Delta x\end{aligned}$$

The average is obtained by dividing a Riemann sum for f on $[a, b]$ by $(b - a)$. As we increase the size of the sample and let the norm of the partition approach zero, the average approaches $(1/(b - a))\int_a^b f(x)\,dx$. Both points of view lead us to the following definition.

DEFINITION The Average or Mean Value of a Function

If f is integrable on $[a, b]$, then its **average value on** $[a, b]$, also called its **mean value**, is

$$\text{av}(f) = \frac{1}{b - a}\int_a^b f(x)\,dx.$$

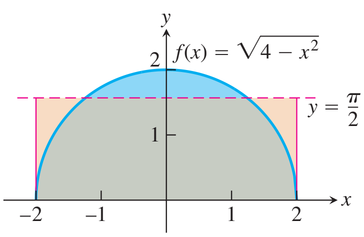


FIGURE 5.15 The average value of $f(x) = \sqrt{4 - x^2}$ on $[-2, 2]$ is $\pi/2$ (Example 5).

EXAMPLE 5 Finding an Average Value

Find the average value of $f(x) = \sqrt{4 - x^2}$ on $[-2, 2]$.

Solution We recognize $f(x) = \sqrt{4 - x^2}$ as a function whose graph is the upper semicircle of radius 2 centered at the origin (Figure 5.15).

The area between the semicircle and the x-axis from -2 to 2 can be computed using the geometry formula

$$\text{Area} = \frac{1}{2} \cdot \pi r^2 = \frac{1}{2} \cdot \pi(2)^2 = 2\pi.$$

Because f is nonnegative, the area is also the value of the integral of f from -2 to 2,

$$\int_{-2}^{2} \sqrt{4 - x^2}\,dx = 2\pi.$$

Therefore, the average value of f is

$$\text{av}(f) = \frac{1}{2 - (-2)}\int_{-2}^{2} \sqrt{4 - x^2}\,dx = \frac{1}{4}(2\pi) = \frac{\pi}{2}.$$

■

You Try It

You Try It

Video