



CONTENTS

1	Definite Integrals	5
1.1	Definition	5
1.2	Positive and Negative Areas	6
1.3	Properties of Integrals	6
1.3.1	What happens when $a = b$?	7
1.3.2	The integral of a constant	7
1.3.3	The integral of a function multiplied by a constant	8
1.3.4	Integrals of sums and differences of functions	8
1.3.5	Integrals of adjacent areas	10
1.3.6	Estimating the value of an integral	10
1.3.7	Other Properties of Integrals	11
1.4	Applications in Physics	13
1.5	Practice Exercises	14
2	The Fundamental Theorem of Calculus	17
2.1	First Part	17
2.2	Second Part	17
2.3	FTC and Definite Integrals	18
2.4	The Meaning of the FTC	20
2.4.1	FTC Practice	21
2.5	Using Antiderivatives to Evaluate Definite Integrals	21
2.5.1	Definite Integrals Practice	22
2.6	Average Value of a Function	24
2.7	Physics Applications	25

2.7.1	Position, Velocity, and Acceleration	25
2.7.2	Work, Kinetic Energy, and Impulse	26
2.8	Volumes with Integrals	27
2.8.1	Volume of a Sphere	27
2.8.2	Volumes of Solids of Revolution	29
2.8.3	Volumes of Other Solids	32
3	Arc Lengths	35
3.1	Determining the Arc Length of a Curve	35
3.2	Arc Length of Vector-valued Functions	39
3.3	Speed and Distance Along a Parametric Curve	40
3.4	Applications in Physics	41
3.5	Practice	42
4	Continuous Distributions	45
4.1	Continuous Probability Distributions	45
4.1.1	Cumulative Distribution Function	45
4.1.2	Probability Density Function	48
4.1.3	The Continuous Uniform Distribution	49
4.1.4	Continuous Distributions In Python	51
4.2	Probabilities as Areas	55
4.2.1	Exponential PDFs	58
5	The Normal Distribution	61
5.1	The Normal Distribution	61
5.1.1	Defining the Normal Distribution	61
5.1.2	Importance of the Normal Distribution	62
5.2	The Normal Curve as a Density	64
6	Sampling Distributions and the Central Limit Theorem	69
6.1	Statistics Vary from Sample to Sample	69
6.2	A Simulation	69
6.3	The Center and Spread of $\bar{x}$	71
6.4	The Central Limit Theorem	71
6.5	The Sampling Distribution of a Proportion	72
7	Simulation-Based Inference	77
7.1	The Bootstrap	77
7.1.1	Beyond the Mean	79
7.1.2	What the Bootstrap Cannot Do	79

	3
7.2 The Permutation Test	80
7.2.1 Why Shuffling Works	81
A Answers to Exercises	83
Index	99

Definite Integrals

Your car's odometer is doing calculus. Every instant, it takes your speed, multiplies by a tiny slice of time, and adds the result to the running total. That running total of "rate times a little bit" is an *integral*, and in this chapter you will learn what it means and how to use its properties. Integrals measure areas, volumes, distances, energy, and much more. A definite integral calculates the net area between the function and the x-axis over a given interval.

Recall that you can use a Riemann sum to estimate the area under a function, and that as we increase the number of subintervals, the estimated area approaches the actual area. In sigma notation we can express a Riemann sum as

$$\sum_{i=1}^n f(x_i) \Delta x \quad (1.1)$$

1.1 Definition

The definite integral of a function $f(x)$ over an interval $[a, b]$ is defined as the limit of a Riemann sum as n approaches ∞ :

$$\int_a^b f(x) \, dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x \quad (1.2)$$

where x_i^* is a sample point in the i^{th} subinterval of a partition of $[a, b]$, $\Delta x = \frac{b-a}{n}$ is the width of each subinterval, and the limit is taken as the number of subintervals n approaches infinity.

Exercise 1

Express

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n (x_i^3 + x_i \sin x_i) \Delta x$$

as an integral on the interval $[0, \pi]$. Do

Working Space

not solve.

Answer on Page 83

1.2 Positive and Negative Areas

What if the function dips below the x -axis? We consider that area negative. In other words, it represents a *decrease* as opposed to an increase. Consider an oscillating object where $v(t) = \sin \pi t$ (figure 1.1). From $t = 0$ to $t = 1$, the velocity is positive, which means the object is moving *away from* the starting position. This is a positive displacement. From $t = 1$ to $t = 2$, the velocity is negative. What does this tell you about the direction the object is moving and its displacement during this time period? A negative velocity means the object is moving *back towards* the starting position.

In general, areas above the x -axis are positive, while areas below the x -axis are negative.

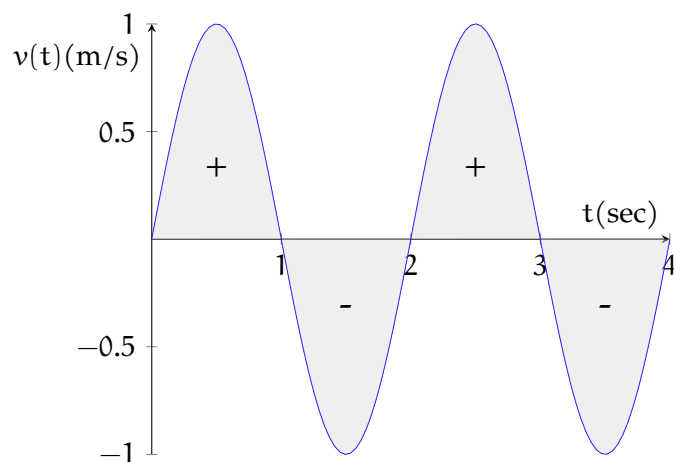


Figure 1.1: velocity of an oscillating object

1.3 Properties of Integrals

There are several important properties of integrals that will help us evaluate more complex integrals in the future. The following examples apply when $f(x)$ is continuous or has a finite number of jump discontinuities on the interval $a \leq x \leq b$.

1.3.1 What happens when $a = b$?

What if the endpoints of the integral are the same? Let us consider $\int_a^b x^2 dx$, and take the limit as $b \rightarrow a$ (shown in figure 1.2). As you can see, as b approaches a , the calculated area decreases. Intuitively, we can guess that when $b = a$, then the width of the area (Δx) is zero, and therefore the area is also zero. Let us prove this formally.

Recall that $\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x$, where $\Delta x = \frac{b-a}{n}$. When $b = a$, every rectangle has width $\Delta x = \frac{a-a}{n} = 0$, so every term in the sum is $f(x_i) \cdot 0 = 0$. The sum is 0 for every n , and so is the limit:

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

This is true for any function.

Note that when $a = b$, the integral just becomes a line. From geometry we also know that a line has no area, as the width or dx is 0.

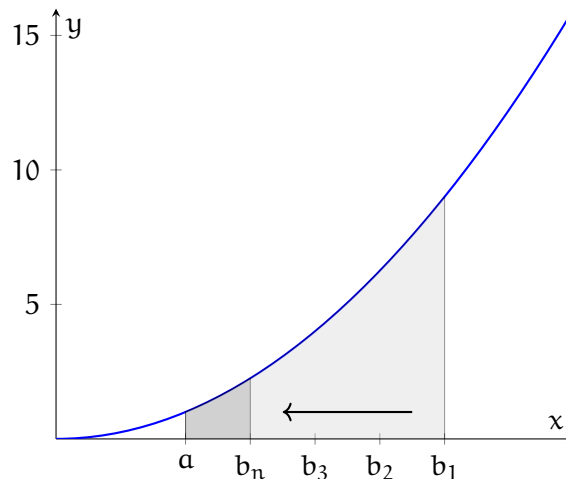
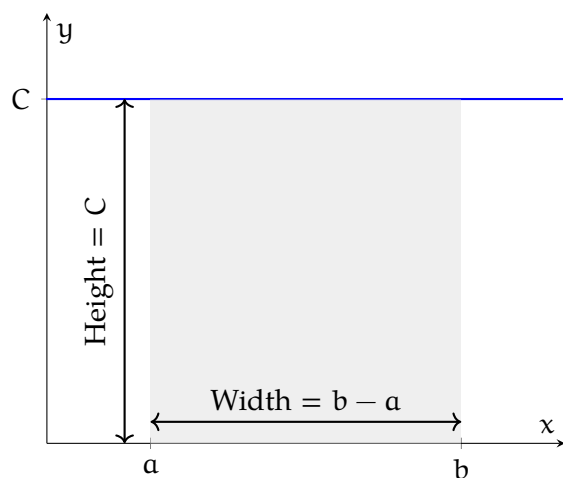


Figure 1.2: As b gets closer to a , the area represented by the integral decreases

1.3.2 The integral of a constant

When the function we are integrating is a constant (that is, it takes the form $f(x) = C$), the area is simply $(b - a) \cdot C$. This is shown graphically in figure 1.3.

Figure 1.3: $\int_a^b f(x) \, dx = (b - a) \cdot C$

Since $f(x) = C$ is a horizontal line, the area under $f(x)$ is simply a rectangle. As you can see in figure 1.3, the width of the rectangle is $b - a$ and the height is C . To find the area of a rectangle, we multiply the width by the height, and therefore $\int_a^b C \, dx = (b - a) \cdot C$.

1.3.3 The integral of a function multiplied by a constant

How is $\int_a^b f(x) \, dx$ related to $\int_a^b C \cdot f(x) \, dx$? Intuitively, we know that multiplying a function by a constant, C , vertically stretches the graph by a factor of C . In turn, the area under the curve increases by a factor of C . Imagine a simple shape, like a triangle. If we keep the base of the triangle the same (analogous to the integral being over the same interval) and make the triangle three times taller (analogous to multiplying the function we are integrating by a factor of $C = 3$), then we would expect the total area of the triangle to be 3 times greater. Therefore, $\int_a^b C \cdot f(x) \, dx = C \int_a^b f(x) \, dx$.

1.3.4 Integrals of sums and differences of functions

If a function can be described as a sum of two other functions, then the integral of the original function is the same as the sum of the integrals of the two other functions. Concretely, we say $\int_a^b f(x) + g(x) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx$. Figure 1.4 shows $f(x) = x + 2$, $g(x) = 4x^3 - 12x^2 + 10x$, and $f(x) + g(x)$. As you can see, the area under $f(x) + g(x)$ is equal to the area under $f(x)$ (the red area) plus the area under $g(x)$ (the diagonal lined area).

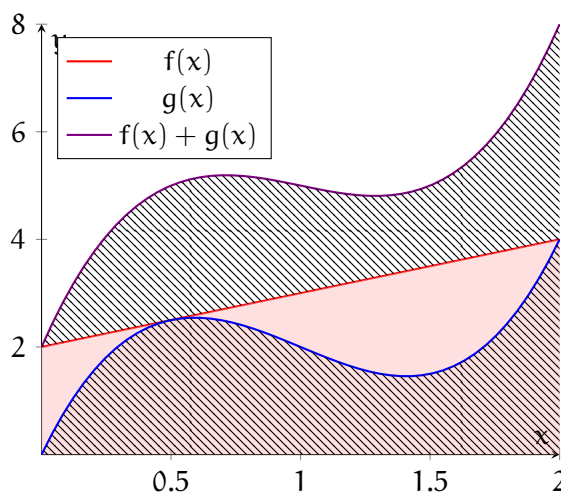


Figure 1.4: The integral of $f(x) + g(x)$ is equal to the integral of $f(x)$ plus the integral of $g(x)$

Mathematically, we can prove this by recalling that the limit of a sum is the sum of the limits:

$$\begin{aligned}
 \int_a^b f(x) + g(x) \, dx &= \lim_{n \rightarrow \infty} \sum_{i=1}^n [f(x_i) + g(x_i)] \Delta x \\
 &= \lim_{n \rightarrow \infty} \left[\sum_{i=1}^n f(x_i) \Delta x + \sum_{i=1}^n g(x_i) \Delta x \right] \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x + \lim_{n \rightarrow \infty} \sum_{i=1}^n g(x_i) \Delta x \\
 &= \int_a^b f(x) \, dx + \int_a^b g(x) \, dx
 \end{aligned}$$

Similar to the addition property, the integral of the difference between two functions is equal to the difference of the integrals of two functions.

$$\int_a^b [f(x) - g(x)] \, dx = \int_a^b f(x) \, dx - \int_a^b g(x) \, dx$$

. This is more difficult to visualize than addition, but we can easily prove it by applying the constant multiple and addition properties. Let us define $f(x) - g(x) = f(x) + (-g(x))$:

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

By the addition property,

$$= \int_a^b f(x) \, dx + \int_a^b -g(x) \, dx = \int_a^b f(x) \, dx + \int_a^b (-1) \cdot g(x) \, dx$$

By the constant multiple property:

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

1.3.5 Integrals of adjacent areas

If c is some x -value between a and b , then $\int_a^c f(x) \, dx + \int_c^b f(x) \, dx = \int_a^b f(x) \, dx$. This is shown graphically in figure 1.5. The total area from $x = a$ to $x = b$ is equal to the red area (the integral from a to c) plus the blue area (the integral from c to b).

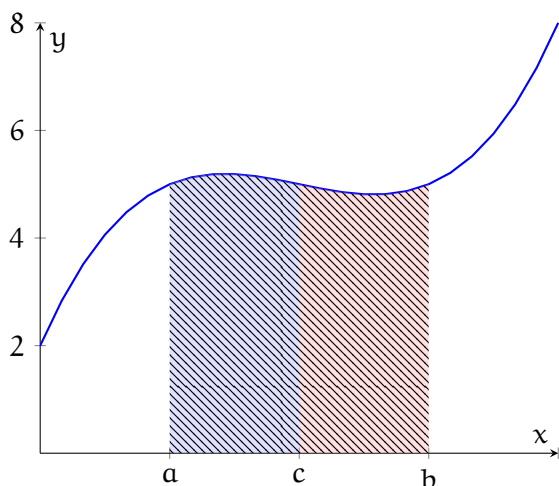


Figure 1.5: $\int_a^c f(x) \, dx + \int_c^b f(x) \, dx = \int_a^b f(x) \, dx$

1.3.6 Estimating the value of an integral

Suppose we need to know the area under a complex function. We can estimate a range for the value of the integral if we can bookend the function over the interval we are interested. Suppose there is some value m such that $f(x) \geq m$, and some other value M such that $f(x) \leq M$ on the interval we are interested in (see figure 1.6). The total area under $f(x)$ is the light blue plus the darker blue. The total area under $y = M$ is the darker blue, plus the light blue, plus the white area. The darker blue area under the curve has total area $m \cdot (b - a)$ and the rectangle under $y = M$ has total area $M \cdot (b - a)$ (since these are both integrals of a constant, which we learned about above). The actual area under our

function is more than just the dark blue area, but less than the total area under $y = M$. Therefore, $m(b - a) \leq \int_a^b f(x) \, dx \leq M(b - a)$ if $m \leq f(x)$ and $M \geq f(x)$ on the interval $x \in [a, b]$.

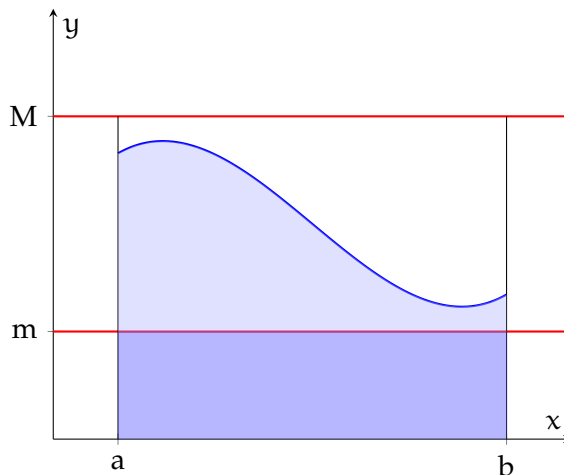
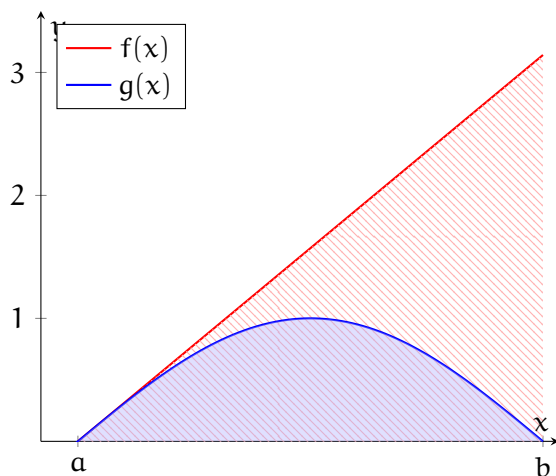


Figure 1.6: $m \leq f(x) \leq M$

1.3.7 Other Properties of Integrals

If $f(x) \geq 0$ over the interval $a \leq x \leq b$, then $\int_a^b f(x) \, dx \geq 0$. We can make an intuitive, geometric argument to support this claim. Recall that areas above the x -axis are considered positive. If $f(x) \geq 0$, then all the area of the integral lies above the x -axis, therefore, the total cannot be negative.

Similarly, if $f(x) \geq g(x)$ on the interval $a \leq x \leq b$, then $\int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx$ (see figure 1.7). The entire area under $g(x)$ is contained in the area under $f(x)$. Therefore, $\int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx$

Figure 1.7: $\int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx$

Lastly, we see what happens when we switch a and b . While it is unusual to integrate from right to left (that is, in a case where $a > b$), this property will be useful. Recall that

$$\int_a^b f(x) \, dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \frac{(b-a)}{n}$$

What is $\int_b^a f(x) \, dx$? Substituting, we see that

$$\int_b^a f(x) \, dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \frac{(a-b)}{n}$$

Noting that $(a-b) = -(b-a)$ we see:

$$\begin{aligned} \int_b^a f(x) \, dx &= \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \frac{-(b-a)}{n} \\ &= (-1) \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \frac{(b-a)}{n} \\ &= (-1) \int_a^b f(x) \, dx \end{aligned}$$

Therefore, it is true that $\int_b^a f(x) \, dx = -\int_a^b f(x) \, dx$.

1.4 Applications in Physics

We have already seen that the area under a velocity function is displacement, and the area under an acceleration function is change in velocity (Riemann Sums). We can use integrals to determine the change in position of an object over a given time frame. If we *also* know the object's starting position, then we can state the object's ending position. Consider the graph of an object's velocity in figure 1.8:

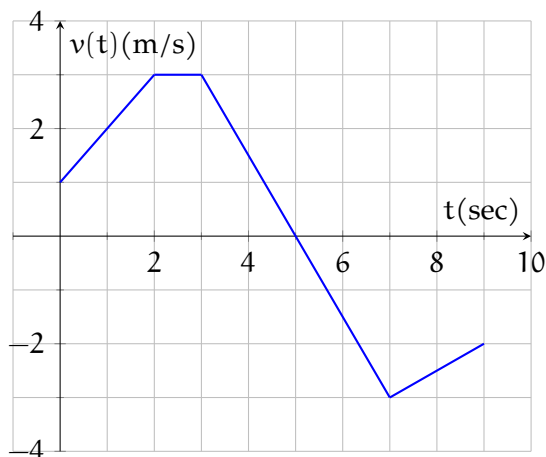


Figure 1.8: Velocity of an object from $t = 0$ to $t = 9$

We can determine the net displacement of the object from $t = 0$ to $t = 9$ by evaluating $\int_0^9 v(t) dt$. Since the definite integral is equal to the area under the curve, we need to find the total area. As the function consists of straight lines, we will leave the explicit calculation of the area as an exercise for the student. You should find that the total positive area (above the x -axis) is 10 meters, and the total negative area (below the x -axis) is 8 meters. Therefore, the object's displacement over the specified time interval is $10 - 8 = 2$ meters.

When you push on something to move it, you are applying a force over a distance (assuming you are strong enough to move it!). The integral of force as a function of distance is the *work* done on that object. If it is the only force acting, the work equals the change in the object's kinetic energy (KE). Mathematically, this is

$$\int_a^b F(x) dx = \Delta KE = \frac{1}{2} m(v_f^2 - v_i^2)$$

If you integrate the force as a function of time, that is *impulse*. Impulse is the change in momentum (p) of the object. Mathematically, this is

$$\int_a^b F(t) dt = \Delta p = m(v_f - v_i)$$

Example problem: You push a 3 kg box with force $F(x) = 0.5x$, where x is measured in meters and F is measured in Newtons. If the box was initially at rest, what is its speed when it reaches the 2 meter mark? (Hint: $KE = \frac{1}{2}mv^2$.)

Solution: Change in kinetic energy is the area under a force-distance curve. We can plot the force applied to the box from $d=0$ to $d=2$ (see figure 1.9):

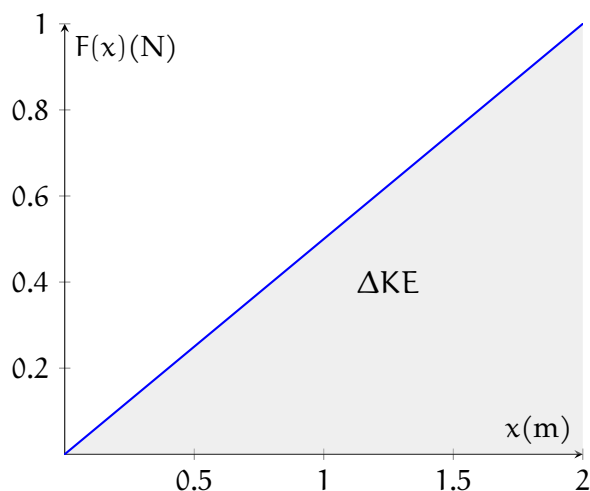


Figure 1.9: Force applied to a box over a distance; the shaded area represents the change in kinetic energy.

Given that the box's initial velocity is $0 \frac{m}{s}$, we know that the initial kinetic energy (KE) is 0J. This implies that $KE_f = \Delta KE$. We can find ΔKE from the shaded area:

$$\Delta KE = \frac{1}{2}(2m)(1N) = 1J = KE_f$$

Solving for the final velocity:

$$KE_f = 1J = \frac{1}{2}(3kg)(v^2)$$

$$2J = (3kg)(v^2)$$

$$\frac{2}{3} \frac{m^2}{s^2} = v^2$$

$$v = \sqrt{\frac{2}{3} \frac{m^2}{s^2}} \approx 0.816 \frac{m}{s}$$

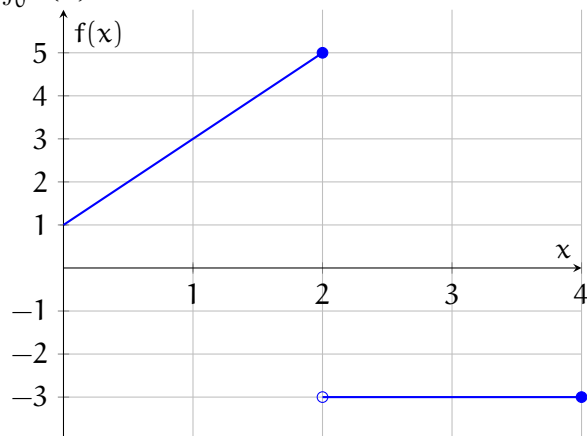
1.5 Practice Exercises

Exercise 2

Given that $\int_0^1 x^2 dx = \frac{1}{3}$, use the properties of integrals to evaluate $\int_0^1 (5 - 6x^2) dx$.

*Working Space**Answer on Page 83***Exercise 3**

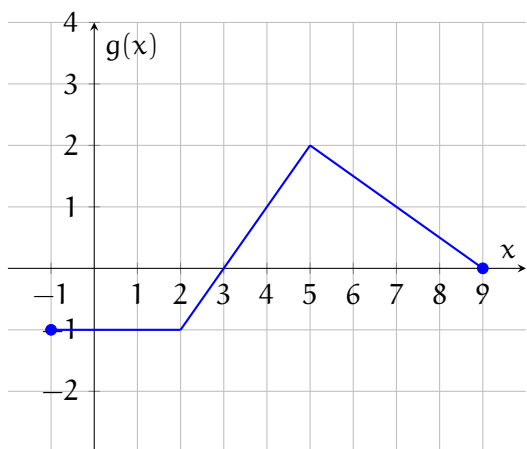
[This question was originally presented as a multiple-choice problem on the 2012 AP Calculus BC exam.] The graph of f is shown. What is the value of $\int_0^4 f(x) dx$?

*Working Space**Answer on Page 83***Exercise 4**

[This question was originally presented as a multiple-choice problem on the 2012 AP Calculus BC exam.] The graph of the piecewise function $g(x)$ is shown. What

Working Space

is the value of $\int_{-1}^9 3g(x) + 2 \, dx$?



Answer on Page 84

Exercise 5

[This question was originally presented as a calculator-allowed, multiple-choice question on the 2012 AP Calculus BC exam.] If $f'(x) > 0$ for all real numbers and $\int_4^7 f(x) \, dx = 0$, which of the following could be a table of values for the function f ?

function f :

(A)

x	$f(x)$
4	-4
5	-3
7	0

(B)

x	$f(x)$
4	-4
5	-2
7	5

(C)

x	$f(x)$
4	-4
5	6
7	3

(D)

x	$f(x)$
4	0
5	0
7	0

(E)

x	$f(x)$
4	0
5	4
7	6

Working Space

Answer on Page 84

The Fundamental Theorem of Calculus

In the last chapter, finding the area under even a simple curve took pages of rectangles and limits. The Fundamental Theorem of Calculus (FTC) is the shortcut that makes integrals easy: to add up a function, find something whose derivative it is. It ties together the two big ideas of calculus, differentiation and integration, and it comes in two parts.

2.1 First Part

The first part of the Fundamental Theorem of Calculus states that if f is a continuous real-valued function defined on a closed interval $[a, b]$ and F is the function defined, for all x in $[a, b]$, by:

$$F(x) = \int_a^x f(t) \, dt \quad (2.1)$$

Then, F is uniformly continuous and differentiable on the open interval (a, b) , and $F'(x) = f(x)$ for all x in (a, b) . (That is $F(x)$ is the antiderivative of $f(x)$.)

2.2 Second Part

The second part of the Fundamental Theorem of Calculus states that if f is a real-valued function defined on a closed interval $[a, b]$ that admits an antiderivative F on $[a, b]$, and f is integrable on $[a, b]$ (it need not be continuous), then

$$\int_a^b f(t) \, dt = F(b) - F(a). \quad (2.2)$$

We will also use shorthand as follows:

$$\int_a^b f(t) \, dt = F(t)|_a^b \quad (2.3)$$

Which means “ $F(t)$ evaluated from $t = a$ to $t = b$ ”.

2.3 FTC and Definite Integrals

Let f be a function that is continuous on the interval $x \in [a, b]$ and $g(x)$ is given by:

$$g(x) = \int_a^x f(t) \, dt$$

So, g is continuous on $[a, b]$ and differentiable on (a, b) . Additionally,

$$g'(x) = f(x)$$

Proof: Let x and $x + h$ be in (a, b) . Then,

$$g(x + h) - g(x) = \int_a^{x+h} f(t) \, dt - \int_a^x f(t) \, dt$$

Recall from the chapter on definite integrals that we can split the first integral, rewriting it as:

$$\begin{aligned} g(x + h) - g(x) &= \left[\int_a^x f(t) \, dt + \int_x^{x+h} f(t) \, dt \right] - \int_a^x f(t) \, dt \\ g(x + h) - g(x) &= \int_x^{x+h} f(t) \, dt \end{aligned}$$

And for $h \neq 0$:

$$\frac{g(x + h) - g(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t) \, dt$$

Since f is continuous, there is some u in (a, b) , such that $f(u) = m$, where m is the minimum value of f on the interval (a, b) . Similarly, there is also some v , such that $f(v) = M$, where M is the maximum value (see figure ??). We can then state the true inequality that:

$$mh \leq \int_x^{x+h} f(t) \, dt \leq Mh$$

Therefore, (assuming $h > 0$):

$$f(u) \leq \frac{1}{h} \int_x^{x+h} f(t) \, dt \leq f(v)$$

Substituting the equation above for the integral, we see that:

$$f(u) \leq \frac{g(x+h) - g(x)}{h} \leq f(v)$$

If we let h approach zero, then the window that u and v are in collapses and u and v both approach x . Therefore,

$$\lim_{h \rightarrow 0} f(u) = \lim_{u \rightarrow x} f(u) = f(x)$$

Recall also that

$$\lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} = g'(x)$$

Taking the limit as $h \rightarrow 0$ of the whole, inequality becomes the Squeeze Theorem:

$$\begin{aligned} \lim_{h \rightarrow 0} f(u) &\leq \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} \leq \lim_{h \rightarrow 0} f(v) \\ f(x) &\leq g'(x) \leq f(x) \end{aligned}$$

Therefore, if $g(x) = \int_a^x f(t) dt$, then $g'(x) = f(x)$. Notice it does not matter what a is!

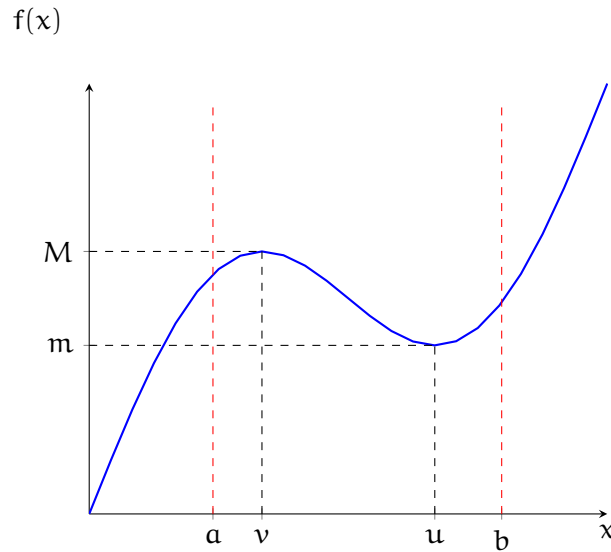


Figure 2.1: $f(v) = M$, the maximum value, and $f(u) = m$, the minimum value on the interval $x \in [a, b]$

Exercise 6

[This question was originally presented as a no-calculator, multiple-choice problem on the 2012 AP Calculus BC Exam.]

Let g be a continuously differentiable function with $g(1) = 6$ and $g'(1) = 3$.

What is the value of $\lim_{x \rightarrow 1} \frac{\int_1^x g(t) dt}{g(x) - 6}$?

- (A) 0
- (B) $\frac{1}{2}$
- (C) 1
- (D) 2
- (E) The limit does not exist

Working Space

Answer on Page 84

2.4 The Meaning of the FTC

What the Fundamental Theorem of Calculus is really saying is that differentiation and integration are opposite processes. Mathematically, we can say

$$\frac{d}{dx} \int_a^x f(t) dt = f(x)$$

This may seem clunky, but many useful functions are defined this way. Consider the Fresnel function: $S(x) = \int_0^x \sin \frac{\pi t^2}{2} dt$. Originally used in optics, this equation is also used by civil engineers to design road and railway curves. According to FTC, then, $S'(x) = \sin \frac{\pi x^2}{2}$.

We can also apply the Chain Rule when taking derivatives of integrals. Let $f(x) = \int_1^{x^4} \sec t dt$. What is $f'(x)$? First, let us define $u = x^4$. By the Chain Rule,

$$\begin{aligned} \frac{d}{dx} \int_1^{x^4} \sec t dt &= \frac{d}{dx} \int_1^u \sec t dt \\ &= \frac{d}{du} \left[\int_1^u \sec t dt \right] \frac{du}{dx} \\ &= \sec u \frac{du}{dx} \end{aligned}$$

Noting that $\frac{du}{dx} = \frac{d}{dx} x^4 = 4x^3$,

$$f'(x) = \sec x^4 (4x^3)$$

2.4.1 FTC Practice

Exercise 7

Use the Fundamental Theorem of Calculus to find the derivative of the function.

1. $g(x) = \int_0^x \sqrt{t + t^3} dt$

2. $F(x) = \int_x^0 \sqrt{1 + \sec t} dt$

3. $h(x) = \int_1^{e^x} \ln t dt$

4. $y = \int_{\sqrt{x}}^{\frac{\pi}{4}} \theta \tan \theta d\theta$

Working Space

Answer on Page 85

2.5 Using Antiderivatives to Evaluate Definite Integrals

In everyday English, the FTC states that the integral from a to b of a function is the antiderivative of that function evaluated from a to b . In the previous chapter, the areas under the curves were triangles, rectangles, and other shapes that could be found with geometry. The FTC connects integrals to antiderivatives, allowing us to evaluate more complex integrals. Consider the following example:

The power consumption of a household can be modeled as $P(t) = \frac{1}{10} t^2 (t - 24)^2$ from $t = 0$ to $t = 24$, where P is measured in watts and t is measured in hours ($t = 0$ is midnight). The total energy the household uses is given by $\int_0^{24} P(t) dt$. As you can see from the graph (see figure 2.2), we cannot simply use our geometry skills to determine the area under the curve.

To determine the total energy use, we need to evaluate $\int_0^{24} \frac{1}{10} t^2 (t - 24)^2 dt$. First, we expand the polynomial:

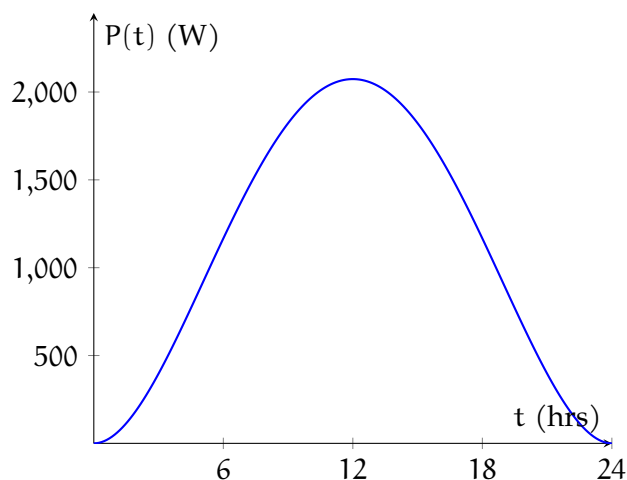


Figure 2.2: Power consumption of a household in a day

$$\begin{aligned} E_{\text{tot}} &= \frac{1}{10} \int_0^{24} t^2(t^2 - 48t + 576) dt = \frac{1}{10} \int_0^{24} t^4 - 48t^3 + 576t^2 dt \\ &= \frac{1}{10} \int_0^{24} t^4 dt - \frac{24}{5} \int_0^{24} t^3 dt + \frac{288}{5} \int_0^{24} t^2 dt \end{aligned}$$

Using the power rule to determine the antiderivatives of t^4 , t^3 , and t^2 , we see:

$$= \frac{1}{10} \left[\frac{1}{5} t^5 \right]_0^{24} - \frac{24}{5} \left[\frac{1}{4} t^4 \right]_0^{24} + \frac{288}{5} \left[\frac{1}{3} t^3 \right]_0^{24} = 26,542 \text{ Wh} \approx 26.5 \text{ kWh}$$

2.5.1 Definite Integrals Practice

Exercise 8

Evaluate the following integral:

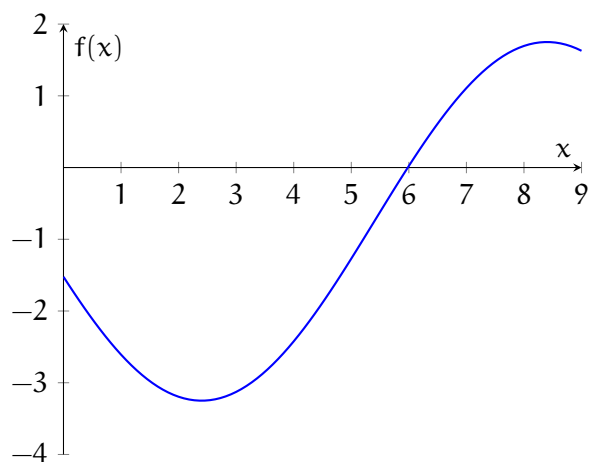
1. $\int_1^4 t^{-3/2} dt$

Working Space

Answer on Page 85

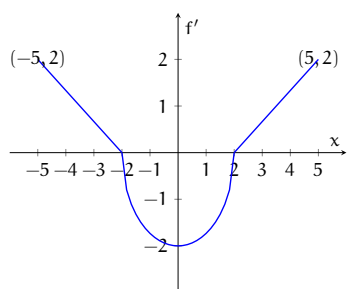
Exercise 9

[This question was originally presented as a multiple-choice, no-calculator problem on the 2012 Calculus BC exam.] The graph of a differentiable function f is shown in the graph. $h(x) = \int_0^x f(t) dt$. Rank the relative values of $h(6)$, $h'(6)$, and $h''(6)$ from lowest to highest.

*Working Space**Answer on Page 85***Exercise 10**

The graph of f' is shown in the graph and consists of a semi-circle and two line segments. If $f(2) = 1$, then what is $f(-5)$?

Working Space



Answer on Page 85

Exercise 11

[This question was originally presented as a calculator-allowed, multiple-choice question on the 2012 AP Calculus BC exam.] A particle moves along a line so that its acceleration for $t \geq 0$ is given by $a(t) = \frac{t+3}{\sqrt{t^3+1}}$. If the particle's velocity at $t = 0$ is 5, what is the velocity of the particle at $t = 3$?

Working Space

Answer on Page 85

2.6 Average Value of a Function

The average value of a function, f , on an interval $[a, b]$ is given by:

Average Value of a Function

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

Example: Find the average value of $f(x) = 3 + x^2$ on the interval $[-2, 1]$.

Solution: Taking $a = -2$ and $b = 1$, we have:

$$\begin{aligned} f_{\text{avg}} &= \frac{1}{1 - (-2)} \int_{-2}^1 [3 + x^2] \, dx \\ f_{\text{avg}} &= \frac{1}{3} \int_{-2}^1 [3 + x^2] \, dx \\ f_{\text{avg}} &= \frac{1}{3} \left[3x + \frac{1}{3}x^3 \right]_{x=-2}^{x=1} \\ f_{\text{avg}} &= \frac{1}{3} \left[\left(3 \cdot 1 + \frac{1}{3} \cdot 1^3 \right) - \left(3 \cdot (-2) + \frac{1}{3} \cdot (-2)^3 \right) \right] \\ f_{\text{avg}} &= \frac{1}{3} \left[\left(3 + \frac{1}{3} \right) - \left(-6 - \frac{8}{3} \right) \right] \\ f_{\text{avg}} &= \frac{1}{3} \left[\frac{10}{3} + 6 + \frac{8}{3} \right] = \frac{1}{3}(12) = 4 \end{aligned}$$

Therefore, the average value of $f(x) = 3 + x^2$ on the interval $[-2, 1]$ is 4.

Exercise 12

[This question was originally presented as a multiple-choice, calculator- allowed problem on the 2012 AP Calculus BC Exam.] What is the average value of $y = \sqrt{\cos x}$ on the interval $0 \leq x \leq \frac{\pi}{2}$?

Working Space

Answer on Page 86

2.7 Physics Applications

2.7.1 Position, Velocity, and Acceleration

When we discussed derivatives, you saw that:

$$v(t) = \frac{d}{dt}x(t)$$

$$a(t) = \frac{d}{dt}v(t)$$

If we integrate both sides of these equations and apply the Fundamental Theorem of Calculus, we see that:

$$x(t) = x(0) + \int_0^t v(s) \, ds$$

$$v(t) = v(0) + \int_0^t a(s) \, ds$$

Taking a derivative loses any constant, so integrating cannot tell us where the object started. We have to supply the initial position $x_0 = x(0)$ and initial velocity $v_0 = v(0)$ ourselves. (The letter s is just a placeholder variable, so that t can be the upper limit.)

Exercise 13

[This question was adapted from a multiple-choice question originally presented in the 2012 AP Physics C public practice exam.] A particle of mass m accelerates from rest starting at time $t = 0$. The acceleration is caused by a single force, $F = bt$, where b is a constant. Derive a formula for the particle's velocity, $v_x(t)$, in terms of b , t , and m .

Working Space

Answer on Page 86

2.7.2 Work, Kinetic Energy, and Impulse

The integral of force over distance is the work done, and work equals the change in kinetic energy. The integral of force over time is the impulse, which equals the change in momentum. Both are worked through in the chapter on definite integrals.

2.8 Volumes with Integrals

Back in the volumes chapter, we found the volume of a pyramid by graphing the area of each slice and finding the area under that curve. We had to take one area on faith. Now you can compute it with an integral: if a solid's slice at position x has area $A(x)$, its volume is

$$V = \int_a^b A(x) dx$$

Each slice is a thin plate of volume $A(x) dx$, and the integral adds them all up.

2.8.1 Volume of a Sphere

Below, we will prove the volume of a sphere is given by $\frac{4}{3}\pi r^3$ using the integral method. Suppose we have a sphere of radius r centered at the origin (see figure 2.3).

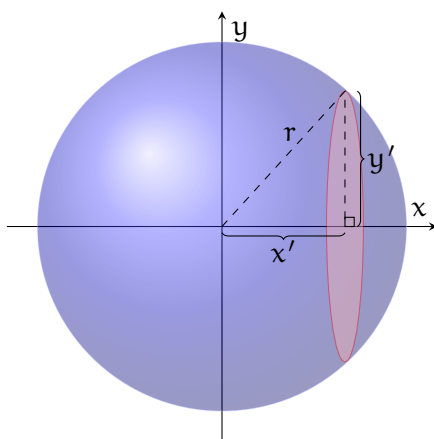


Figure 2.3: A vertical cross-section of a sphere

We begin by taking very thin vertical cross-sections. The radius of the cross-section is the height, y , of the sphere at the horizontal position, x . Since the edges of the cross-section lie on the sphere, we know the edge of the cross-section is distance r from the origin. Applying the Pythagorean theorem, we see that $r^2 = x^2 + y^2$, which implies that $y = \sqrt{r^2 - x^2}$. So, the area of the cross-section is given by $\pi y^2 = \pi(r^2 - x^2)$. If we imagine each cross section as having a width, dx , and taking the sum of all the cross sections from $x = -r$ to $x = r$, we can write an integral equal to the volume of the sphere:

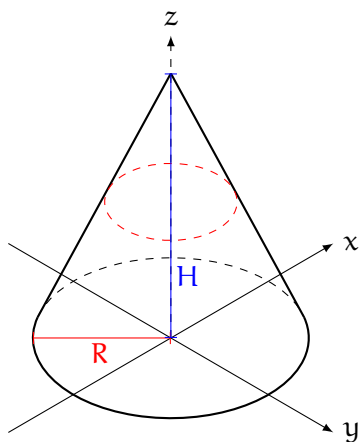
$$V_{\text{sphere}} = \int_{-r}^r \pi(r^2 - x^2) dx$$

We can then evaluate that integral:

$$\begin{aligned}
 V_{\text{sphere}} &= \pi \int_{-r}^r r^2 \, dx - \pi \int_{-r}^r x^2 \, dx \\
 V_{\text{sphere}} &= \pi \left[r^2 x \right]_{x=-r}^{x=r} - \frac{\pi}{3} \left[x^3 \right]_{x=-r}^{x=r} \\
 V_{\text{sphere}} &= \pi \left[r^3 - (-r^3) \right] - \frac{\pi}{3} \left[r^3 - (-r^3) \right] \\
 V_{\text{sphere}} &= 2\pi r^3 - \frac{2\pi}{3} r^3 = \frac{4}{3} \pi r^3
 \end{aligned}$$

Exercise 14

Prove the volume of a regular cone is $\frac{\pi}{3} R^2 H$, where R is the radius of the base and H is the height of the cone. (Hint: A cone is a series of decreasing circles stacked on top of each other; see figure below.)



Working Space

Answer on Page 86

2.8.2 Volumes of Solids of Revolution

We can also find the volume of solids made by revolving a graph about the x or y -axis. Suppose the graph $y = \sin x$ from $x = 0$ to $x = \frac{\pi}{2}$ were rotated vertically about the x -axis to form a solid. How could we find the volume of that solid? Well, we can imagine a rectangle of width dx and height y (see figure 2.4)

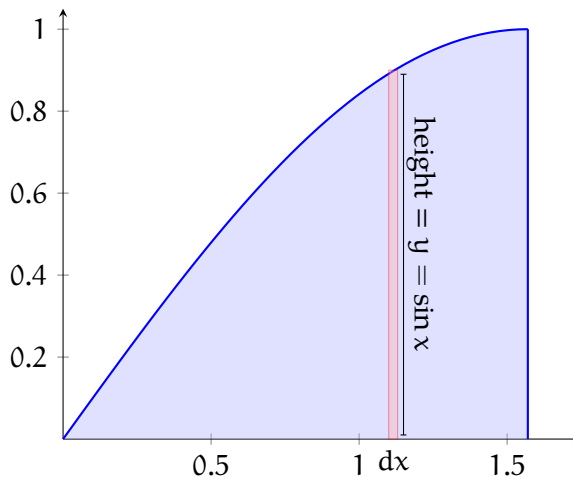


Figure 2.4: A cross section has width dx and height $y = \sin x$

If we rotate the plot vertically about the x -axis, the rectangle becomes a cylinder with radius $y = \sin x$ and height dx (see figure ??). Therefore, the volume of each cylindrical slice is $V_{\text{slice}} = \pi r^2 dx = \pi \cdot \sin^2 x dx$.

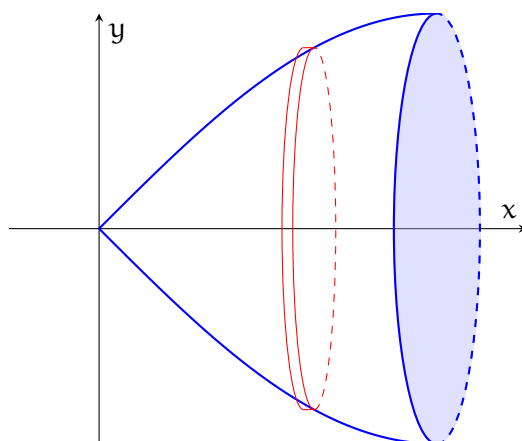


Figure 2.5: When rotated, the cross-section becomes a cylinder with radius $\sin x$ and width dx , which has a total volume of $\pi \sin^2 x dx$

We can find the total volume by integrating from 0 to $\pi/2$:

$$V = \pi \int_0^{\pi/2} \sin^2 x \, dx$$

Recall the half angle formula, $\sin^2 x = \frac{1}{2} (1 - \cos 2x)$. Substituting, we see that:

$$V = \frac{\pi}{2} \int_0^{\pi/2} (1 - \cos 2x) \, dx$$

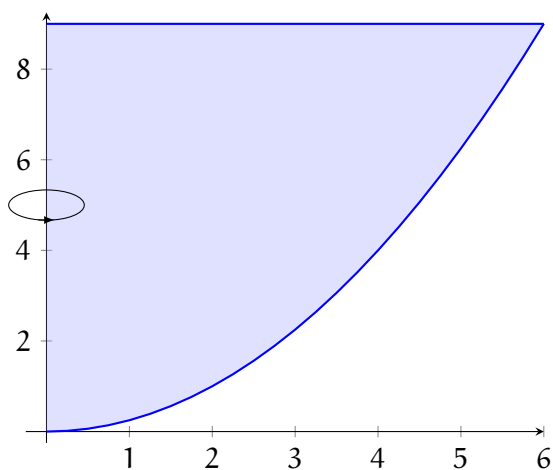
$$V = \frac{\pi}{2} \left(x - \frac{1}{2} \sin 2x \right) \Big|_{x=0}^{x=\pi/2}$$

$$V = \frac{\pi}{2} \left[\left(\pi/2 - \frac{1}{2} \sin \pi \right) - \left(0 - \frac{1}{2} \sin 0 \right) \right]$$

$$V = \frac{\pi}{2} [\pi/2 - 0 - 0 + 0] = \frac{\pi^2}{4}$$

Exercise 15

Find the volume of a solid created by rotating the region bounded by $x = 2\sqrt{y}$, $x = 0$, and $y = 9$ about the y -axis. A graph is shown below.



Working Space

Answer on Page 87

Exercise 16

Let $f(x) = (ax^3 + bx^2 + cx + d)\sqrt{1 - x^2}$. Bird's eggs of various sizes can be modeled by rotating $f(x)$ about the x -axis, with different values of a , b , c , and d defining different sizes and shapes of eggs. For a domestic chicken, $a = -0.02$, $b = 0.03$, $c = 0.12$, and $d = 0.454$. For a mallard duck, $a = -0.06$, $b = 0.04$, $c = 0.1$, and $d = 0.54$. Use a calculator, such as a TI-89 or Wolfram Alpha, to determine which species lays a bigger egg.

Working Space

Answer on Page 88

Using donuts for solids of revolution

Sometimes there is space between the region we are rotating and the line we are rotating it about. Consider the region bounded between $y = 2x$ and $y = x^2$ (see figure 2.6):

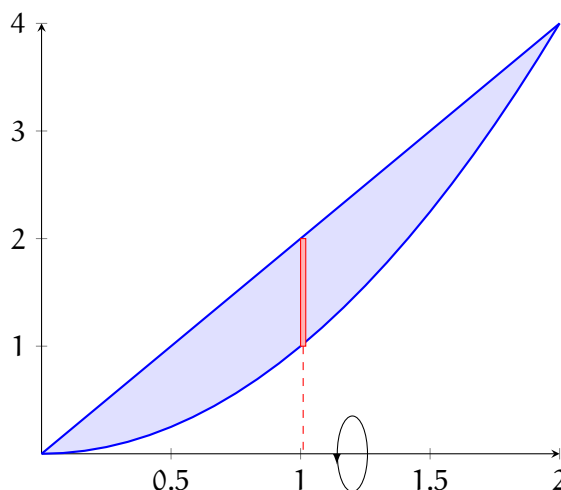


Figure 2.6: When rotated, the slices will become donuts with outer radius $2x$ and inner radius x^2

When rotated, the slices will take the form of donuts (or washers), the volume of which is $\pi(R^2 - r^2) dx$, where R is the outer radius and r is the inner radius. Therefore, in this case, the total volume of the rotated region is given by:

$$\begin{aligned}
 V &= \int_0^2 \pi \left[(2x)^2 - (x^2)^2 \right] dx \\
 V &= \pi \int_0^2 4x^2 - x^4 dx = \pi \left[\frac{4}{3}x^3 - \frac{1}{5}x^5 \right]_{x=0}^{x=2} \\
 V &= \pi \left[\frac{4}{3}2^3 - \frac{1}{5}2^5 \right] = \pi \left[\frac{32}{3} - \frac{32}{5} \right] \\
 V &= \frac{64\pi}{15}
 \end{aligned}$$

Exercise 17

What is the volume of the region bounded by $y = x^2$ and $y = 2\sqrt{x}$ when rotated about the y -axis?

Working Space

Answer on Page 88

2.8.3 Volumes of Other Solids

You can also model a solid as a base defined by a function with cross-sections of specific shapes. Consider the function $y = x^2$ from $x = 0$ to $x = 2$ (see figure 2.7). Suppose the area between the curve, the y -axis, and the line $y = 4$ defines a base and each vertical cross-section is a square. So, the width of the each cross section is dx , the length is $4 - x^2$, and (because they are squares) the height in the z -plane is also $4 - x^2$. The volume of each cross-section is $V_{\text{slice}} = (4 - x^2)^2 dx$ and the total volume of the solid is:

$$\begin{aligned}
 V &= \int_0^2 (4 - x^2)^2 dx \\
 V &= \int_0^2 (16 - 8x^2 + x^4) dx \\
 V &= \left[16x - \frac{8}{3}x^3 + \frac{1}{5}x^5 \right]_{x=0}^{x=2}
 \end{aligned}$$

$$V = 16(2) - \frac{8}{3}(2)^3 + \frac{1}{5}(2)^5 = \frac{256}{15} \approx 17.067$$

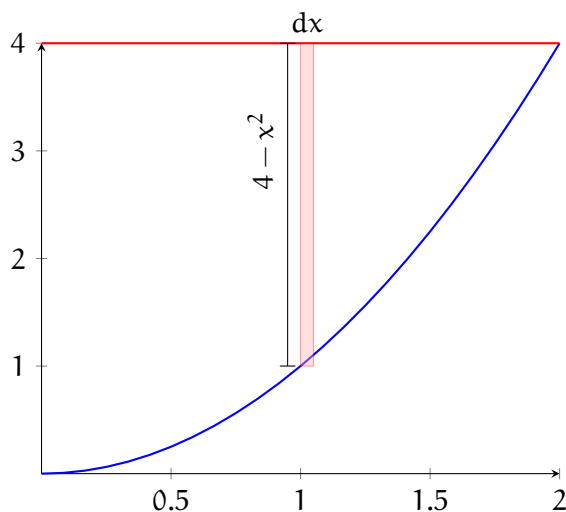


Figure 2.7: $y = x^2$ with a vertical cross-section

You can use a similar method for triangular, semi-circular, or any other shape cross-section. The trick is writing everything in terms of x (when your cross sections are vertical and have width dx) or y (when your cross section are horizontal and have length dy).

Exercise 18

[This question was originally presented as a multiple-choice, calculator-allowed question on the 2012 AP Calculus BC exam.] Let R be the region in the first quadrant bounded above by the graph $y = \ln(3 - x)$, for $0 \leq x \leq 2$. R is the base of a solid for which each cross section perpendicular to the x -axis is square. What is the volume of the solid? Give your answer to 3 decimal places.

Working Space

Answer on Page 90

Exercise 19

Find the volume of a solid whose base is defined by the ellipse $9x^2 + 16y^2 = 25$ and is made up of right isosceles triangle cross-sections perpendicular to the x -axis (with the hypotenuse in the base of the solid).

Working Space

Answer on Page 90

Arc Lengths

3.1 Determining the Arc Length of a Curve

How long is the cable hanging between two telephone poles? How far does a roller coaster car actually travel along its twisting track? These are questions about the length of a curve, and integrals answer them. In real life, we could do this by laying a piece of string up against the curve, then straightening out the string and measuring its length with a ruler (you may have done this in elementary school when you were first learning about the relationship between the radius and circumference of a circle).

Archimedes estimated the circumference of a circle by fitting polygons inside it (and around it) with more and more sides. (Archimedes' proof that π is between $3\frac{10}{71}$ and $3\frac{1}{7}$ is more complicated, but we will not dive into that here.) As we increase the number of sides of the inscribed polygon, the perimeter of the polygon (the sum of the lengths of the sides) gets closer to the circumference of the circle (see figure 3.1).

Now, it is easy to find the length of a polygon: just add up the length of the line segments! Using this, we can find the length of a curve by approximating it as many short lines and adding up the lengths of those lines.

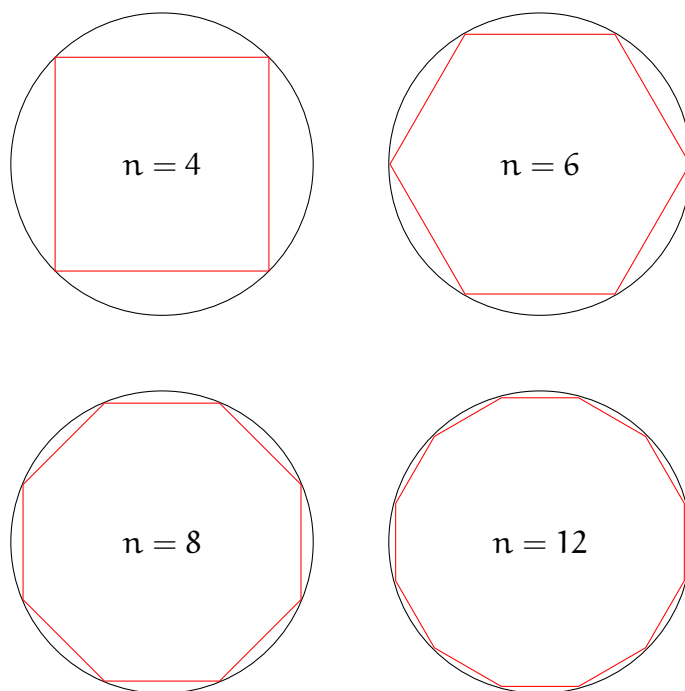


Figure 3.1: As n increases, the perimeter of the inscribed polygon approaches the circumference of the circle

We can choose n points along the graph of $f(x)$ and connect each point with a straight line (this is shown in figure 3.2). If we add up the length of the lines, we get an estimate of the length of the curve. We represent the length of the line between the i^{th} point, P_i and the previous point, P_{i-1} as $|P_{i-1}P_i|$ (recall that the absolute value sign can be used to signify the length or **magnitude** of something). Therefore, the sum of the lengths of the lines approximating the curve is:

$$\sum_{i=1}^n |P_{i-1}P_i|$$

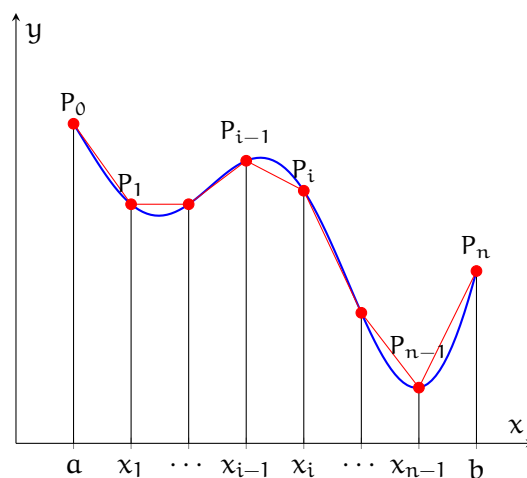
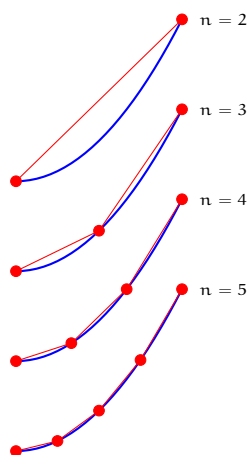
Figure 3.2: Polygon approximation of $f(x)$ 

Figure 3.3: As the number of points increases, the total length of the line segments approaches the true length of the curve

The more points we choose, the closer the lines lay to the actual curve (see figure 3.3), and the closer our estimate is to the true length. So, to find the true length, we will want to take n to ∞ . Therefore, the actual curve length is the limit as $n \rightarrow \infty$ of that sum:

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n |P_{i-1} P_i|$$

The length of each segment can be found using the Pythagorean theorem. Recall that the distance between two points on the xy -plane is $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. (For a reminder of why this is, see figure 3.4.) The coordinates of P_{i-1} are $(x_{i-1}, f(x_{i-1}))$ and the coordinates of P_i are $(x_i, f(x_i))$. Substituting this into the above sum, we see that the total length of

the segments is

$$\sum_{i=1}^n \sqrt{(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2}$$

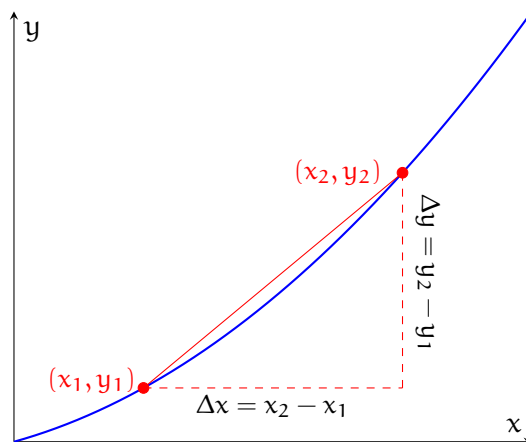


Figure 3.4: The distance between two points on the xy -plane

Recall the Mean Value Theorem, which states that there is some x_i^* such that $f(x_i) - f(x_{i-1}) = f'(x_i^*)(x_i - x_{i-1})$. Substituting this into the above sum, we get:

$$\sum_{i=1}^n \sqrt{(x_i - x_{i-1})^2 + (f'(x_i^*)(x_i - x_{i-1}))^2}$$

Recall from the chapter on Riemann sums and the integral that we defined $\Delta x = x_i - x_{i-1}$ and we can further rewrite the sum as

$$\sum_{i=1}^n \sqrt{(\Delta x)^2 + [f'(x_i^*)\Delta x]^2} = \sum_{i=1}^n \sqrt{1 + [f'(x_i^*)]^2} \sqrt{(\Delta x)^2} = \sum_{i=1}^n \sqrt{1 + [f'(x_i^*)]^2} \Delta x$$

Putting this all together, we see that the actual length of the curve is defined as

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{1 + [f'(x_i^*)]^2} \Delta x$$

This is the definition of the integral of $\sqrt{1 + [f'(x)]^2}$ and therefore the length of some function $f(x)$ on the interval $a < x < b$ is $\int_a^b \sqrt{1 + [f'(x)]^2} dx$. In another notation, this is equivalent to $\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$.

3.2 Arc Length of Vector-valued Functions

Suppose you have a vector-valued function, $f(t) = [x(t), y(t)]$. A common example might be an artillery shell shot at an angle. For a shell shot with an initial velocity v_0 at angle θ from the ground, its position can be described with the vector-valued function $f(t) = [v_0 \cos \theta t, v_0 \sin \theta t - 4.9t^2]$. (A concrete example where $v_0 = 12 \frac{m}{s}$ and $\theta = 30^\circ$ is shown in figure 3.5.)

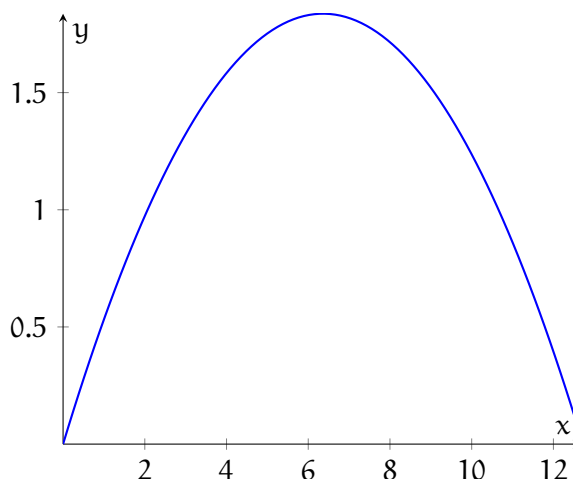


Figure 3.5: The path of an artillery shell with shot with initial velocity v_0 at angle θ

How can we find the length of the flight path of the artillery shell? We can reinterpret the length integral for a vector-valued function, $f(t) = [x(t), y(t)]$.

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_a^b \sqrt{1 + \frac{(dy/dt)^2}{(dx/dt)^2}} \frac{dx}{dt} dt$$

Moving the $\frac{dx}{dt}$ under the square root, we see that

$$L = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

which is equivalent to

$$L = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2} dt$$

1

¹Note that this is similar to distance formula, just applied at every dx along the curve. Zooming in on a small piece of the curve, it looks like a straight line, and seems as if the distance formula is being applied.

3.3 Speed and Distance Along a Parametric Curve

In the parametric functions chapter, a Python animation showed a point moving along a curve as t increased. Parametric equations describe not just a shape but *motion* along it, and the point does not necessarily move at a constant rate.

The point's velocity is the vector $[x'(t), y'(t)]$, and its *speed* is the length of that vector:

$$\text{Speed} = \sqrt{(x'(t))^2 + (y'(t))^2}$$

The arc length formula above is just the integral of the speed: adding up speed $\times$ time over $a \leq t \leq b$ gives the total distance traveled, which (if the curve is traced once) is its length.

Example: Finding Speed

Given

$$x'(t) = 8t - t^2, \quad y'(t) = -t + \sqrt{t^{1.2} + 20},$$

the speed of the particle at time t is

$$\text{Speed} = \sqrt{(8t - t^2)^2 + \left(-t + \sqrt{t^{1.2} + 20}\right)^2}.$$

Evaluating at $t = 2$ gives a speed of approximately 12.3 cm/s.

Example: Length of a Curve

Consider the parametric equations

$$x = \cos t, \quad y = \sin t,$$

for $0 \leq t \leq 2\pi$.

First compute the derivatives:

$$x'(t) = -\sin t, \quad y'(t) = \cos t.$$

Substitute into the arc length formula:

$$L = \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2} dt.$$

Using the identity $\sin^2 t + \cos^2 t = 1$, this simplifies to

$$L = \int_0^{2\pi} 1 \, dt = 2\pi.$$

Thus, the length of the curve traced is 2π , which matches the circumference of the unit circle.

3.4 Applications in Physics

When we take the integral of a velocity function, we get the *displacement*. For example, if you drove to school and home again, your displacement would be zero. However, the *distance* you traveled is not zero! Distance counts every movement, so it is the integral of the speed, $\int_a^b |v(t)| \, dt$. (Remember, that if $x(t)$ is the object's position, then its velocity is given by $x'(t)$.)

Suppose a block is attached to a spring on a frictionless horizontal surface. You pull on the block, initiating harmonic motion described by $v(t) = (-0.16) \sin 9t$, where v is in $\frac{\text{m}}{\text{s}}$ and t is in sec. (Note: We are working in radians, not degrees.) What is the block's displacement from $t = 0$ to $t = 3$? What is the total distance the block moves from $t = 0$ to $t = 3$?

To find the displacement, we integrate the velocity function over the specified interval:

$$\begin{aligned} \int_0^3 (-0.16) \sin 9t \, dt &= \frac{-0.16}{9} (-\cos 9t) \Big|_0^3 \\ &= \frac{0.16}{9} [\cos 27 - \cos 0] = \frac{(0.16)(-0.29 - 1)}{9} = -0.0229\text{m} \end{aligned}$$

The position function and displacement are shown in figure 3.6. To find the total distance traveled, we need to add up the speed over time. Before we do so, take a minute to mentally predict: Will the distance be more or less than the displacement?

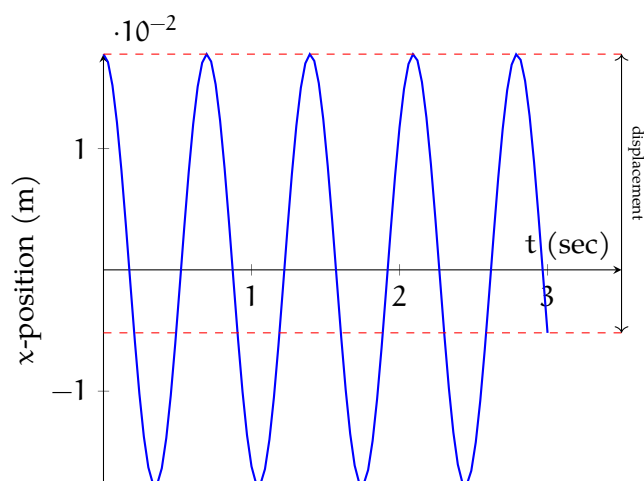


Figure 3.6: The position of the block with displacement shown. The total distance traveled counts every back-and-forth movement.

The distance traveled is $\int_a^b |v(t)| dt$, so for the block we need

$$\int_0^3 |-0.16 \sin 9t| dt$$

The absolute value makes this integral piecewise, so we split it at each zero of $\sin 9t$. Each piece is a simple sine integral. Adding them up gives a total distance of about 0.307 meters. A calculator with Riemann sums gives the same value. Did you predict that the distance would be greater than the displacement?

3.5 Practice

Exercise 20

Write an integral that gives the length of the requested curve.

Working Space

1. $y = \ln x$ from $x = 1$ to $x = 3$
2. $y = \sin x$ from $x = 0$ to $x = \pi$
3. $y = \frac{x^3}{3} + \frac{1}{4x}$ from $x = 1$ to $x = 4$

4. $y = \ln(\cos x)$ from $x = 0$ to $x = \frac{\pi}{3}$

Answer on Page 90

Exercise 21

The arc length function for a curve $f(x)$, where f is an increasing function, is given by $s(x) = \int_0^x \sqrt{3t+5} \, dt$. If f has y -intercept 2, find an equation for f . What point on f is three units from the y -intercept? Give your answer to the thousandths place.

Working Space

Answer on Page 91

Exercise 22

[This problem was originally presented as a no-calculator, multiple-choice question on the 2012 AP Calculus BC exam.] Write an integral that gives the length of the curve $y = \ln x$ from $x = 1$ to $x = 2$.

Working Space

Answer on Page 91

Exercise 23

An out-of-control rocket ship is spiraling out of control through space. Its velocity can be described with the vector-valued function $\mathbf{v}(t) = [-1412 \sin t, 1412 \cos t, t]$

Working Space

where v is in $\frac{\text{m}}{\text{s}}$ and t is in sec. How far does the ship travel in the first 60 seconds? In the second 60 seconds? [Hint: in three dimensions, the length of a vector-valued function is $\int_a^b \sqrt{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2 + (\frac{dz}{dt})^2} dt$]

Answer on Page 92

Continuous Distributions

4.1 Continuous Probability Distributions

What is the probability that a newborn weighs exactly 3.4 kg? Not 3.4001, not 3.3999, but exactly 3.4000000...? The answer is zero, and yet babies are born near 3.4 kg all the time. Continuous quantities need a new way of thinking about probability.

When we talked about the probability distribution of the sum of two dice, we assigned a probability to each of the 11 possibilities:

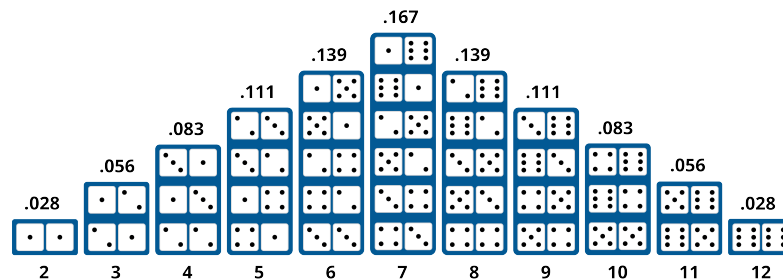


Figure 4.1: Rolling two dice probability distribution.

The probabilities all added up to be 1.0. That is a way of saying “100% of the times you throw the dice, the sum will be an integer between 2 and 12.”

Now we need to talk about probabilities of properties that are continuous, not discrete. For example, we might want to ask the question, “If I randomly pick a cow from all the cows in the world, what is the probability that it will weigh less than 597.34 kg?” What does a probability distribution for a continuous variable look like?

4.1.1 Cumulative Distribution Function

Imagine that you live in ancient times. You buy, sell, and ship cows. A lot of people come into your office and brag about their cows: “Bessie is heavier than 99% of the cows in world!” So, you need to develop some statistics on cow weights.

You have ten cows. You weigh them:

Cow	Mass in kg
Cow 1	580.22
Cow 2	540.07
Cow 3	538.20
Cow 4	512.39
Cow 5	589.75
Cow 6	456.91
Cow 7	583.09
Cow 8	493.56
Cow 9	489.97
Cow 10	704.15

If someone comes into your shop and says, “My Bessie is an astonishing 530 kg!” it would be cool to have a list on the wall that would let you yell back, “Most of my cows are heavier than that, Silly!” So, you sort the cows by weight. For each weight, you say how many of your cows are lighter than that:

Proportion of cows lighter	Mass in kg
0.00	456.91
0.10	489.97
0.20	493.56
0.30	512.39
0.40	538.20
0.50	540.07
0.60	580.22
0.70	583.09
0.80	589.75
0.90	704.15

In fact, for easy reference, you make a plot:

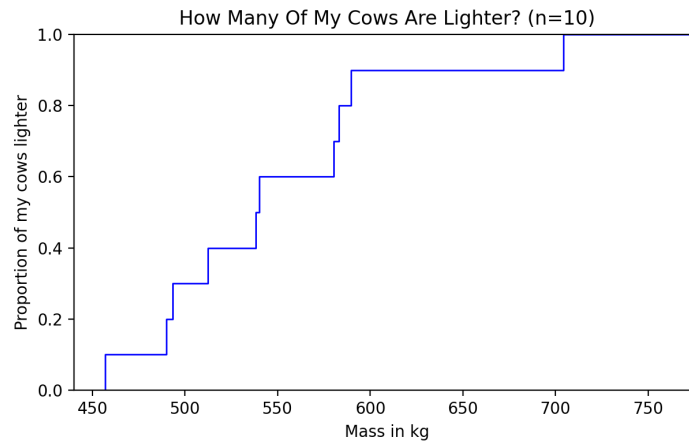


Figure 4.2: 10 cow cumulative distribution function.

Now, for any weight, you can quickly look up what proportion of your cows are lighter (and, if you subtract that from 1, what proportion of your cows are heavier).

See how jagged that graph is? That is because you only have the data for 10 cows. However, as the years pass and you weigh thousands of cows, the plot will become smoother. Because it always accumulates more cows as you move from left to right, this is known as a *cumulative distribution function*, or CDF:

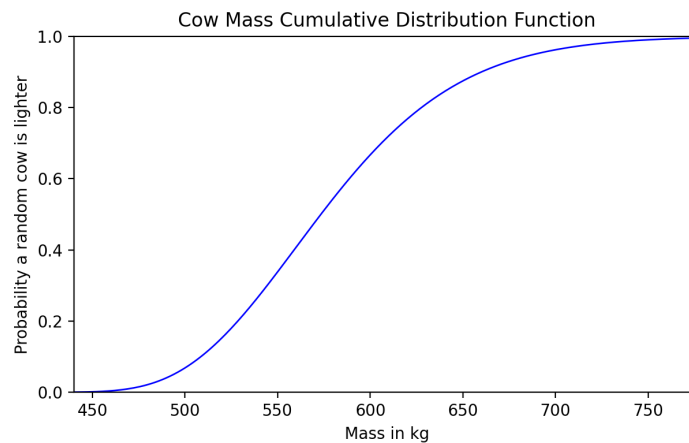


Figure 4.3: Cumulative distribution function for cow weights.

A cumulative distribution function always starts at 0 and ends at 1. On that journey, it never decreases. Reading the graph tells you what proportion of the population is *less than* any given value. For bell-shaped data, the CDF is an S-shaped curve: flat at the ends and steepest in the middle, where most of the values are.

Let us say you want to know what proportion of cows weigh between 620 kg and 635 kg. Using the CDF, you could figure out that 77% of all cows weigh less than 620 kg and 83% of all cows weigh less than 635 kg. Thus 6% of all cows must weigh more than 620 kg and less than 635 kg.

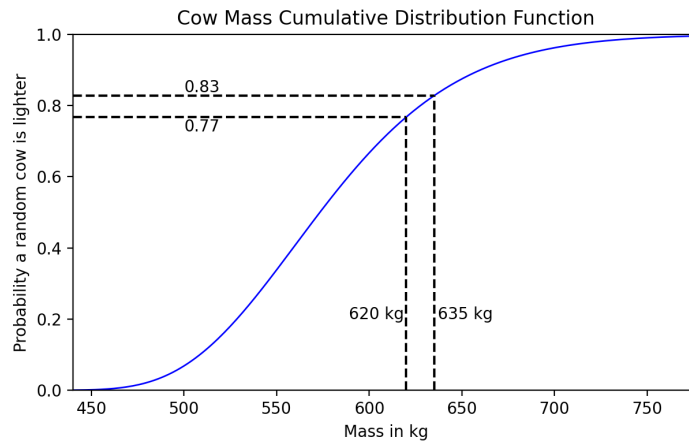


Figure 4.4: Cumulative distribution function with bounds from 620 kg to 635 kg.

4.1.2 Probability Density Function

The cumulative distribution function is handy, but some of its information can be hard to see. For example, how would you answer the question, “What is the most common weight of a cow?” You would squint at the CDF and try to determine where it was steepest.

To make these sorts of questions easier to answer, we take the derivative of the CDF to get the *probability density function* (or PDF). For the cows, it would look like this:

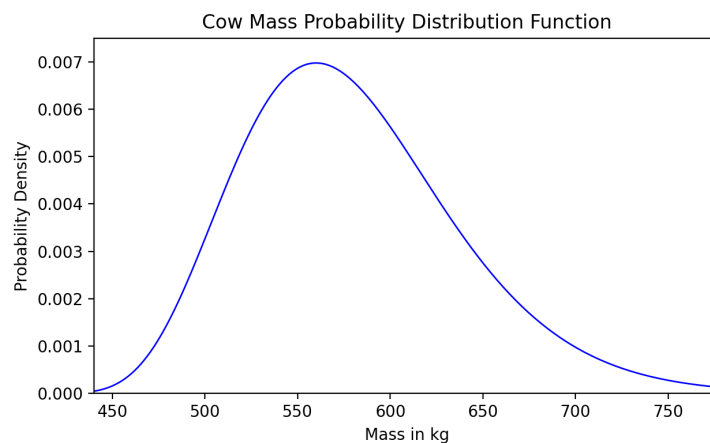


Figure 4.5: Probability density function for cow weights.

Now you can easily see that the CDF was steepest at about 560 kg. The value where the PDF is highest is called the *mode*: the most common weight. You might say “560 kg is the mode of cow mass.”

Note that the mode is often different from the mean or the median. In this case, for example, the distribution is skewed right — the long tail of heavy cows pulls the mean and median above the mode.

Once again, let us say you want to know what proportion of cows weigh between 620 kg and 635 kg. This is more difficult with a PDF than it is with a CDF. With a PDF, you have to find the area under the curve between $x = 620$ and $x = 635$.

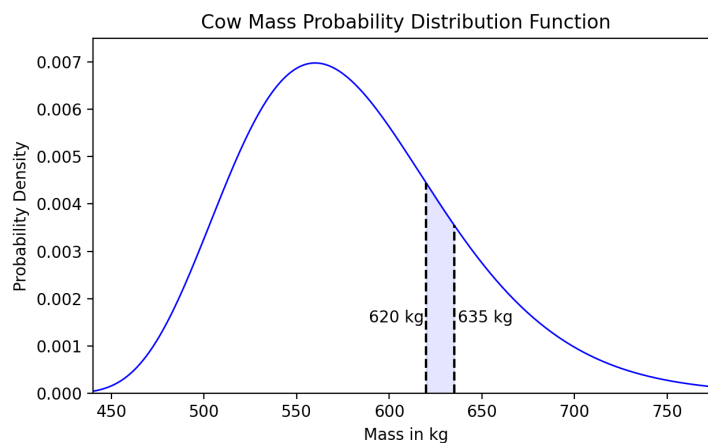


Figure 4.6: Probability density function with the integral from 620 kg to 635 kg.

This is why it is called a “probability density” — to get a true probability you need to multiply the density by the width of the region.

What is the area under the entire curve? If you integrated it, you would get the CDF. The CDF goes from 0 to 1.0. The area under a PDF is *always* 1.0.

4.1.3 The Continuous Uniform Distribution

The most simple continuous distribution is the uniform distribution between two numbers a and b . The CDF is a straight line from zero at a to 1 at b . For example, here is the CDF for the uniform distribution between 6 and 9:

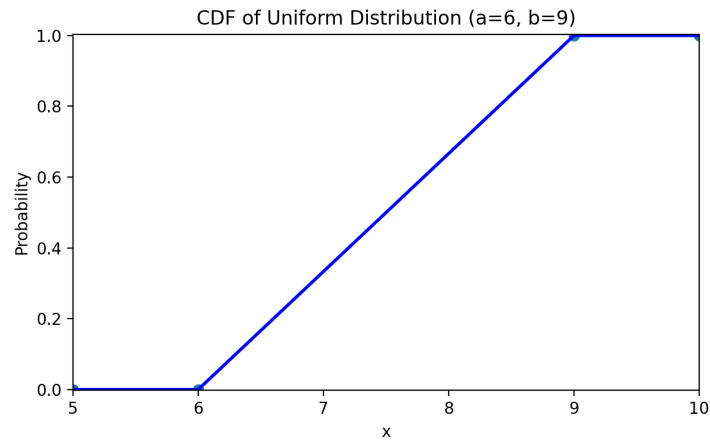


Figure 4.7: Cumulative distribution function for a uniform distribution between 6 and 9.

That line goes from 0 to 1 over a distance of 3, so its slope is $\frac{1}{3}$ between 6 and 9 and zero everywhere else. Thus, the PDF (its derivative) looks like this:

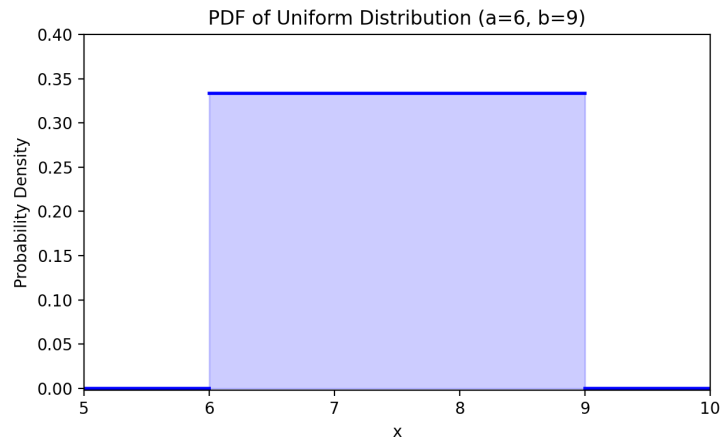


Figure 4.8: Probability density function for a uniform distribution from the continuous uniform distribution.

So, we can write the probability distribution of a continuous uniform distribution between a and b as:

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b. \end{cases}$$

Notice that if a and b are less than 1 apart, the value of $p(x)$ will be greater than 1. This is an important difference between a probability and a probability density:

- A probability will always be in the interval $[0, 1]$.
- A probability density will never be less than 0, but can be much larger than 1.

That said, the probability density will always integrate to 1.

Here is the PDF for a uniform distribution between 0.6 and 0.9:

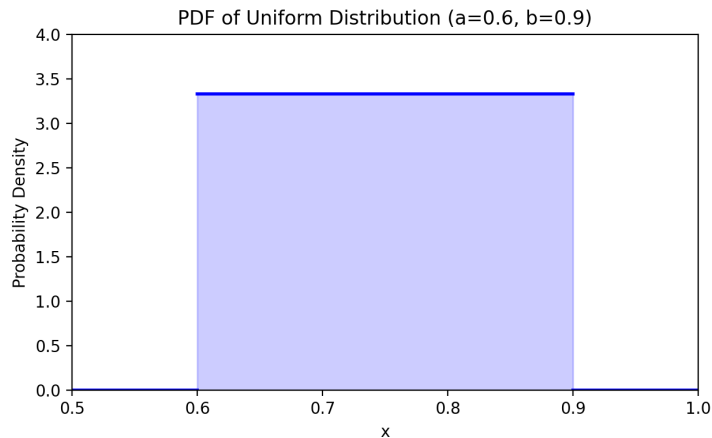


Figure 4.9: Uniform probability density function.

The mean and median of a uniform distribution between a and b is its midpoint: $\frac{a+b}{2}$.

The variance (σ^2) is $\frac{(b-a)^2}{12}$.

4.1.4 Continuous Distributions In Python

The SciPy library has functions that let a programmer work with a large collection of different probability distributions.

For example, if you wanted to work with a continuous uniform distribution between 6 and 9, you would import the relevant functions like this:

```
1 from scipy.stats import uniform
```

Now, if you wanted a numpy array containing a sample of 300 numbers generated randomly from that distribution:

```
1 samples = uniform.rvs(loc=6, scale=3, size=300)
```

The `loc` argument is `a`. The `scale` argument is `b - a`.

If you wanted to know the value of the probability density function at 8 and 10, you could use the `pdf` function:

```
1 x_values = np.array([8, 10])
2 p_values = uniform.pdf(x_values, loc=6, scale=3)
```

Now, `p_values` contains 0.33333 and 0.0.

To get the value of the cumulative distribution function at those points, you would use the `cdf` function:

```
1 cdf_values = uniform.cdf(x_values, loc=6, scale=3)
```

Now, `cdf_values` contains 0.666667 and 1.0.

The inverse of the CDF is very useful. It answers questions like “How heavy does a cow have to be to be in top 1%?”

```
1 bottom_top_percentiles = np.array([0.01, 0.99])
2 boundaries = uniform.ppf(bottom_top_percentiles, loc=6, scale=3)
```

Now, `boundaries` contains 6.03 and 8.97.

The SciPy library supplies these functions (`rvs`, `pdf`, `cdf`, and `ppf`) for over a hundred common continuous probability distributions.

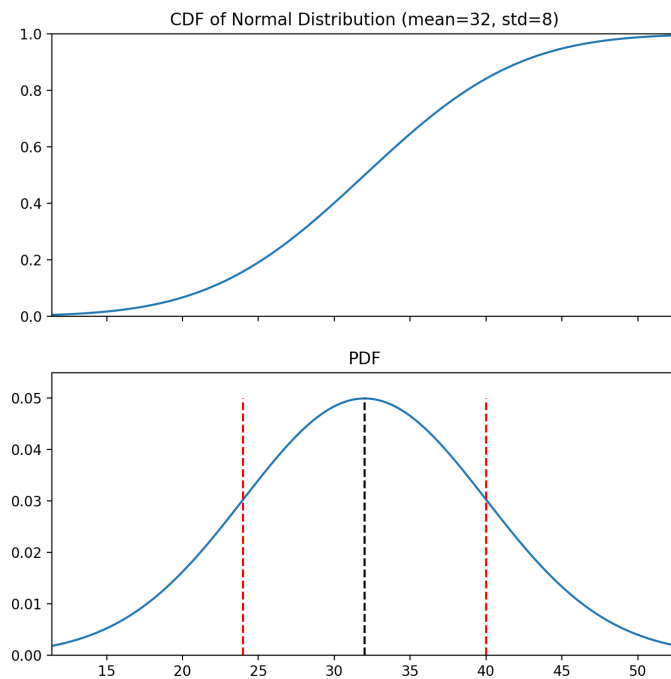
The common “bell curve” shaped distribution is called a Gaussian or Normal distribution. It is described by its mean (the midpoint of the bell) and its standard deviation. For the normal distribution, about 68% of the population lies within one standard deviation of the mean. We will talk a lot more about the normal distribution in other chapters, but let us take this opportunity to plot the CDF and PDF of a normal distribution with a mean of 32 and a standard deviation of 8.

Create a file called `plot_norm.py` and add the following lines:

```
1 import numpy as np
2 from scipy.stats import norm
3 import matplotlib.pyplot as plt
4
5 # Constants
```

```
6 MEAN = 32
7 STD = 8
8
9 # Plotting from the 0.5 percentile to the 99.5 percentile
10 x_min = norm.ppf(0.005, loc=MEAN, scale=STD)
11 x_max = norm.ppf(0.995, loc=MEAN, scale=STD)
12
13 # Make 200 points between x_min and x_max
14 x_values = np.linspace(x_min, x_max, 200)
15
16 # Get CDF for each x value
17 cdf_values = norm.cdf(x_values, loc=MEAN, scale=STD)
18
19 # Get PDF for each x value
20 pdf_values = norm.pdf(x_values, loc=MEAN, scale=STD)
21
22 # What is the highest density?
23 max_density = norm.pdf(MEAN, loc=MEAN, scale=STD)
24
25 # Make a figure with two axes
26 fig, axs = plt.subplots(nrows=2, sharex=True, figsize=(8, 8), dpi=200)
27 axs[0].set_xlim(left=x_min, right=x_max)
28
29 # Draw the CDF on the first axis
30 axs[0].set_title("CDF of Normal Distribution (mean=32, std=8)")
31 axs[0].set_ylim(bottom=0.0, top=1.0)
32 axs[0].plot(x_values, cdf_values)
33
34 # Draw the PDF on the second axis
35 axs[1].set_title("PDF")
36 axs[1].set_ylim(bottom=0.0, top=max_density * 1.1)
37 axs[1].plot(x_values, pdf_values)
38
39 # Add lines for mean, mean-std, and mean+std
40 axs[1].vlines(MEAN - STD, 0, max_density, "r", linestyle="dashed")
41 axs[1].vlines(MEAN + STD, 0, max_density, "r", linestyle="dashed")
42
43 # Save out the figure
44 fig.savefig("norm_32_8.png")
```

The resulting plot should look like this:



What do those vertical lines mean? An ornithologist might tell you, “The wingspan of adult robins are normally distributed with a mean of 32 cm and a standard deviation of 8 cm.” This means 68% of the population of adult robins would have wingspans between the two red lines.

Exercise 24 SciPy Stats

Working Space

Globally, the height of adult women is approximately normally distributed. The mean is 164.7 cm. The standard deviation is 7.1 cm.

Use Python and SciPy stats to answer these questions:

- To be in the tallest decile (the top 10%) of adult women, how tall does one need to be?
- What percentage of adult women are

between 160 cm and 165 cm?

(In case you are wondering: For men the mean is 178.4 cm and the standard deviation is 7.6 cm.)

Answer on Page 92

4.2 Probabilities as Areas

Earlier in this chapter, you saw that for a single variable, X , the likelihood that variable X has a certain value, x , is given by the probability distribution function (PDF), $f(x)$. Additionally, the total area under a PDF is always 1. Mathematically, this is because the PDF is the derivative of the cumulative density function (CDF), which always starts at 0 and ends at 1. Logically, the area under the PDF from $x = a$ to $x = b$ represents the likelihood that X will be between a and b .

Probability that X lies between a and b

For a single variable, X , with a continuous density function, $f(x)$, the probability that X lies between a and b is given by:

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx$$

Suppose that we know the PDF for human body mass and we wanted to know the probability that a randomly-selected person has a body mass between 70 and 80 kgs. That probability would be represented by the integral of the PDF from $x = 70$ to $x = 80$ (see figure 4.10).

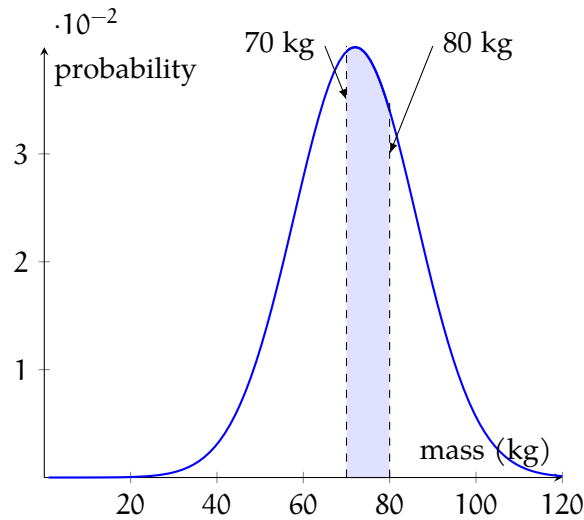


Figure 4.10: The highlighted area is equal to the probability that a person has a mass between 70 and 80 kgs

If we expand that integral to include all possible masses, what would the integral evaluate to? Since that would represent the probability that a random person has mass (which all people do), the integral should be equal to 1. In fact, the integral of a PDF over all possible values of the variable must be equal to 1, since that represents the sum of all the probabilities that the variable will hold all possible values.

For any probability density function, $f(x)$, the integral over the domain of x is always 1.

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Example: Let $f(x) = kx(10 - x)$ for $0 \leq x \leq 10$ and 0 for all other values of x . Find a value for k . Then find the probability that x lies between 7 and 10.

Solution: Since $f(x)$ is a PDF, we know that:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Solving for k :

$$\begin{aligned} 1 &= \int_0^{10} kx(10 - x) dx = k \int_0^{10} 10x - x^2 dx \\ &= k \left[5x^2 - \frac{1}{3}x^3 \right]_{x=0}^{x=10} = k \left[500 - \frac{1000}{3} \right] = k \left(\frac{500}{3} \right) = 1 \end{aligned}$$

Which implies that $k = \frac{3}{500} = 0.006$

Therefore, in order to be a PDF, $f(x) = 0.006x(10 - x)$ and we can find the probability that X lies between 7 and 10:

$$\begin{aligned} P(7 \leq X \leq 10) &= \int_7^{10} 0.006x(10 - x) \, dx \\ &= 0.006 \left[5x^2 - \frac{1}{3}x^3 \right]_{x=7}^{x=10} = 0.006 \left[5(100 - 49) - \frac{1}{3}(1000 - 343) \right] \\ &= 0.006 \left[255 - \frac{657}{3} \right] = 0.006(36) = 0.216 = 21.6\% \end{aligned}$$

Therefore, there is a 21.6% chance that X lies between 7 and 10.

Exercise 25 Representing Probabilities as Integrals

Let $f(t)$ be the probability density function for the time it takes you to complete your weekly chores at home. Express the following probabilities as integrals or state what the given integral represents in words.

1. that you spend less than 15 minutes on your chores.
2. that you spend between half an hour and an hour on your chores.
3. that you spend more than 45 minutes on your chores.
4. $\int_{20}^{\infty} f(t) \, dt$
5. $\int_{60}^{120} f(t) \, dt$

Working Space

Answer on Page 92

Exercise 26 Using PDFs to Find Probabilities

For each $f(x)$, find k such that $f(x)$ is a PDF. Then, find the requested probability.

Working Space

$$1. f(x) = \begin{cases} \frac{ke^{2-x}}{(1+e^{2-x})^2}, & -2 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases},$$

$$P(X \geq 2)$$

$$2. f(x) = \frac{k}{2+4x^2}, P(1 \leq X \leq 4)$$

$$3. f(x) = k(6x - 3x^2) \text{ for } 0 \leq x \leq 2 \text{ and } 0 \text{ otherwise, } P(X > 3/4)$$

Answer on Page 93

4.2.1 Exponential PDFs

There are two common PDFs used to model real-world probability distributions: exponential and normal distributions. We address exponential PDFs in this section and will discuss normal PDFs in the next.

The time it takes for a radioactive particle to decay, or a bank teller to serve a customer, or a piece of equipment to fail are all commonly modeled with exponentially decreasing PDFs. These take the form:

$$f(t) = \begin{cases} 0, & \text{if } t < 0 \\ ce^{-ct}, & \text{if } t \geq 0 \end{cases}$$

Example: Carbon-14 has a half-life of 5700 years, and the probability that a single particle will undergo decay after a given amount of time, t , is given by:

$$f(t) = \begin{cases} 0, & \text{if } t < 0 \\ \frac{1}{5700} e^{-t/5700} & \text{if } t \geq 0 \end{cases}$$

Where t is in years. Find the probability that a particle of carbon-14 decays before 5700 years.

Solution:

$$\begin{aligned}P(T \leq 5700) &= \int_0^{5700} \frac{1}{5700} e^{-t/5700} dt \\&= \frac{1}{5700} (-5700) \left[e^{-t/5700} \right]_{t=0}^{t=5700} \\&= -1 \left[e^{-1} - e^0 \right] = 1 - \frac{1}{e} \approx 63.2\%\end{aligned}$$

The next chapter is about the most important continuous distribution of all: the normal distribution.

The Normal Distribution

Measure the heights of everyone in your grade, and plot them. Most people are near the middle, a few are very tall or very short, and the picture is a symmetric hill. Test scores, birth weights, and errors in measurement make the same hill. It is called the *normal distribution*, or the bell curve, and it is the most important distribution in statistics. In the Continuous Distributions chapter, you learned to describe a continuous variable with a CDF and a PDF. This chapter applies those tools to the bell curve.

5.1 The Normal Distribution

The Normal distribution, also known as the Gaussian distribution, is a type of continuous probability distribution for a real-valued random variable. It is one of the most important probability distributions in statistics due to its several unique properties and usefulness in many areas.

5.1.1 Defining the Normal Distribution

The Normal distribution is defined by its mean (μ) and standard deviation (σ). The probability density function (pdf) of a Normal distribution is given by:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

where:

- x is a value of the variable,
- μ is the mean or expectation of the distribution,
- σ is the standard deviation,
- σ^2 is the variance.

5.1.2 Importance of the Normal Distribution

There are several reasons why the Normal distribution is crucial in statistics:

- **Central Limit Theorem:** One of the main reasons for the importance of the Normal distribution is the Central Limit Theorem (CLT). The CLT states that the distribution of the sum (or average) of a large number of independent, identically distributed variables approaches a Normal distribution, regardless of the shape of the original distribution. The next chapter, Sampling Distributions and the Central Limit Theorem, shows this happening in a simulation.
- **Symmetry:** The Normal distribution is symmetric, which simplifies both the theoretical analysis and the interpretation of statistical results.
- **Characterized by Two Parameters:** The Normal distribution is fully characterized by its mean and standard deviation. The mean determines the center of the distribution, and the standard deviation determines the spread or girth of the distribution.
- **Common in Nature:** Many natural phenomena follow a Normal distribution. This includes characteristics like people's heights or IQ scores, measurement errors in experiments, and many others.

Given its properties, the Normal distribution serves as a foundation for many statistical procedures and concepts, including hypothesis testing, confidence intervals, and linear regression analysis.

Example

A set of course grades is modeled by a normal distribution with mean $\mu = 75$ and standard deviation $\sigma = 8$. Using a 90–80–70–60 grading scale, find the percentage of students who earned a B (scores from 80 to less than 90) using Excel and Python.

Solution

Let X be a student's grade. Then

$$X \sim N(75, 8).$$

A B corresponds to the interval $80 \leq X < 90$, so we want

$$P(80 \leq X < 90).$$

Excel computes cumulative probabilities using

`NORM.DIST(x, mean, standard_dev, TRUE)`

Therefore,

$$P(80 \leq X < 90) = P(X < 90) - P(X < 80).$$

In Excel, enter:

`=NORM.DIST(90,75,8,TRUE) - NORM.DIST(80,75,8,TRUE)`

Excel returns approximately

0.2356.

So about 23.56% of students earned a B.

Now compute the same probability in Python.

```
1 from scipy.stats import norm
2
3 mu = 75
4 sigma = 8
5
6 probability = norm.cdf(90, mu, sigma) - norm.cdf(80, mu, sigma)
7
8 print(probability)
9 print(probability * 100)
```

Python returns approximately

0.2356

which confirms the Excel result.

Both Excel and Python are computing the same cumulative distribution function.

Exercise 27

Working Space

A batch of parts has diameters modeled
by a normal distribution with mean $\mu =$

10 mm and standard deviation $\sigma = 2$ mm:

$$X \sim N(10, 2).$$

A part is considered *acceptable* if its diameter is between 6.72 mm and 12.28 mm.

- (a) Write the probability statement that represents the fraction of acceptable parts.
- (b) Use Excel to compute this probability using NORM.DIST.
- (c) Convert your probability to a percentage.

Answer on Page 94

5.2 The Normal Curve as a Density

Other natural phenomena, such as annual snowfall, height of individuals in a population, or birth weight are best modeled with a normal distribution. Professors often find that test scores are normally-distributed as well. Normal distributions have three key properties:

1. unimodal - there is one "peak" on a graph of a normal distribution
2. symmetric - there is an equal probability represented on either side of the "peak"
3. equal mean and median - the mean and median of a normal distribution are the same

An example normal distribution is shown in figure ?? . Notice that there is one peak and the graph is symmetric about that peak.

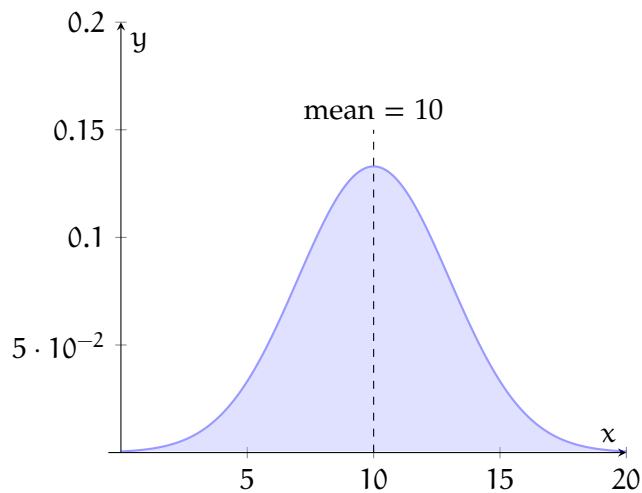


Figure 5.1: A normal distribution with a mean of 10 and a standard deviation of 3

Normal PDFs take the form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Where σ is the *standard deviation* and μ is the *mean*. Recall that for a normal distribution, the standard deviation is the distance you have to go from the mean to reach 68% of the population. That is, if you look at the population within one standard deviation on either side of the mean, it represents approximately 68% of the total population (see figure ??). The lower σ is, the narrower the distribution. The higher σ is, the wider the distribution (see figure ??).

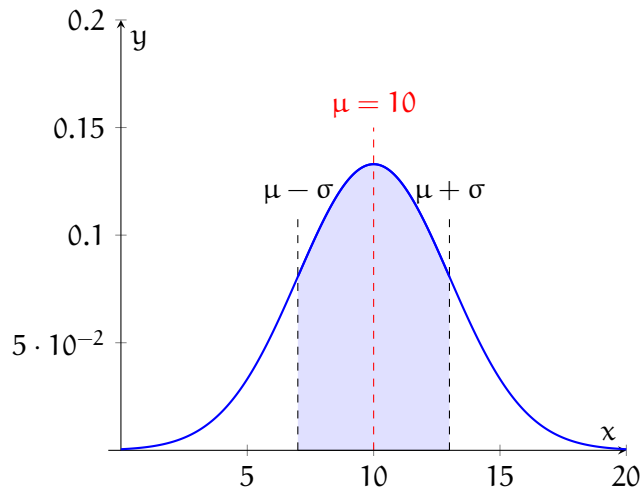


Figure 5.2: The shaded area represents approximately 68% of the population

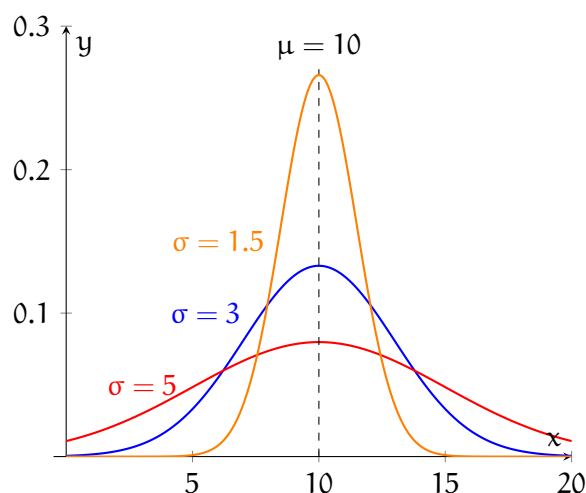


Figure 5.3: Changing σ affects the width of the distribution, but not its center

Example: Birth weights in the United States are normally distributed with a mean of 7 lbs 6 oz and a standard deviation 1 lb 2 oz. Write a PDF to represent this data. What percentage of US babies are born at less than 5 lbs? Over 9 lbs?

Solution: For simplicity, let us express everything in ounces. Recall that there are 16 oz in a pound. Therefore,

$$\mu = 118 \text{ oz}$$

$$\sigma = 18 \text{ oz}$$

And we can write the PDF for US birth weight:

$$f(x) = \frac{1}{18\sqrt{2\pi}} e^{-(x-118)^2/(2(18)^2)}$$

To find the percentage of babies with a birth weight below 5 lbs, we integrate from 0 (since a baby must have some weight) to 80 oz (the equivalent of 5 lbs):

$$P(X \leq 80) = \int_0^{80} \frac{1}{18\sqrt{2\pi}} e^{-(x-118)^2/(2(18)^2)} dx$$

Unfortunately for us, there is no anti-derivative to e^{-x^2} , so we cannot evaluate the integral by hand. However, you can use a calculator or computer to find the approximate value of the integral. Inputting the integral into a calculator (such as a TI-89), we find that:

$$P(X \leq 80) \approx 0.0174$$

And therefore, approximately 1.74% of US babies are born weighing less than 5 lbs.

To find the percentage of babies born weighing more than 9 lbs, we integrate from 144 oz to infinity (obviously, babies are not born with infinite weight, but this will do just fine for an approximation):

$$P(X \geq 144) = \int_{144}^{\infty} \frac{1}{18\sqrt{2\pi}} e^{-(x-118)^2/(2(18)^2)} dx$$

Again, using a calculator or computer, we find that:

$$P(X \geq 144) \approx 0.0743$$

And therefore approximately 7.43% of US babies are born weighing more than 9 lbs.

Sampling Distributions and the Central Limit Theorem

A national poll asks about 1,000 people a question and then announces what more than 150 million voters think, “plus or minus 3 percentage points.” How can 1,000 people possibly speak for 150 million? And where does that 3 come from?

The answer is one of the most surprising facts in all of mathematics. Averages behave far more predictably than the individual values they come from. Take almost any population, as lopsided and strange as you like, draw a large random sample, and compute its average. If you did that over and over, the averages would pile up in a bell-shaped normal curve, every time. That fact is the *Central Limit Theorem* (CLT), and it is the reason every confidence interval and significance test in the chapters that follow works.

6.1 Statistics Vary from Sample to Sample

A *parameter* is a number that describes a whole population, like the true average watch time μ of every video on a streaming site. A *statistic* is a number computed from a sample, like the average $\bar{x}$ of 30 randomly chosen videos. We use statistics to estimate parameters, because measuring a whole population is usually impossible.

Draw a different random sample, and you get a different $\bar{x}$. This is *sampling variability*, and it is not a mistake: it is what random samples do. The key question is *how much* a statistic varies. To answer it, imagine taking every possible sample of the same size and computing the statistic for each. The distribution of all those values is the statistic’s *sampling distribution*.

Nobody can take every possible sample. But a computer can take ten thousand, which is close enough to see the pattern.

6.2 A Simulation

On a video site, most videos are watched for only a few minutes, but a few are watched for an hour. Suppose watch times follow a strongly skewed distribution with mean $\mu = 10$ minutes and standard deviation $\sigma = 10$ minutes. (The program uses the exponential distribution from the Poisson and Exponential chapter, which has exactly that shape.) For

each sample size $n = 1, 5$, and 30 , the program draws 10,000 random samples, averages each one, and draws a histogram of the 10,000 averages:

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy.stats import norm
4
5 rng = np.random.default_rng(1)
6 MEAN = 10          # average watch time, in minutes
7 SAMPLES = 10_000   # how many samples we draw for each size
8
9 fig, axes = plt.subplots(1, 3, figsize=(10, 3))
10 for ax, n in zip(axes, [1, 5, 30]):
11     # 10,000 samples of n watch times each; average each sample
12     data = rng.exponential(MEAN, size=(SAMPLES, n))
13     means = data.mean(axis=1)
14     print(f"n = {n:2}: mean of x-bar = {means.mean():.2f}, "
15           f"sd of x-bar = {means.std():.2f}, theory = {MEAN / np.sqrt(n):.2f}")
16
17     ax.hist(means, bins=50, density=True, color="#005892")
18     x = np.linspace(0, means.max(), 300)
19     ax.plot(x, norm.pdf(x, MEAN, MEAN / np.sqrt(n)), color="orange")
20     ax.set_title(f"sample size n = {n}")
21     ax.set_xlabel("sample mean (minutes)")
22 fig.tight_layout()
23 fig.savefig("clt_simulation.png")

```

$n = 1$: mean of $\bar{x}$ = 10.01, sd of $\bar{x}$ = 10.00, theory = 10.00

$n = 5$: mean of $\bar{x}$ = 9.96, sd of $\bar{x}$ = 4.43, theory = 4.47

$n = 30$: mean of $\bar{x}$ = 9.96, sd of $\bar{x}$ = 1.82, theory = 1.83

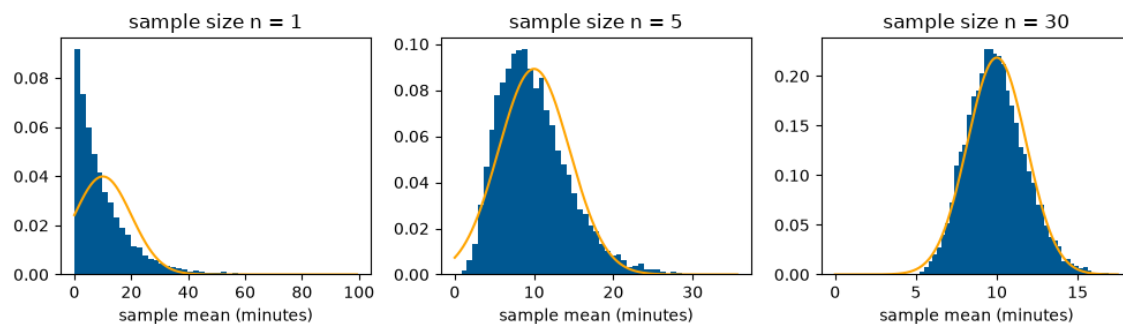


Figure 6.1: Histograms of 10,000 sample means for samples of size 1, 5, and 30, drawn from a skewed population. The orange curve is the normal distribution the CLT predicts.

Three things happen as n grows, and they are the whole story of this chapter:

1. **Center:** the sample means are always centered at the population mean, 10.
2. **Spread:** they cluster more tightly, with a standard deviation that matches $\frac{10}{\sqrt{n}}$.
3. **Shape:** they become normal. With $n = 1$, the histogram is just the skewed population. With $n = 5$, it is still lopsided. With $n = 30$, it fits the normal curve almost perfectly, even though no individual watch time is anywhere near normal.

6.3 The Center and Spread of $\bar{x}$

The first two facts follow from the rules in the Discrete Probability and Random Variables chapter. Think of a sample as n independent random variables $X_1, \dots, X_n$, each with mean μ and standard deviation σ . Then $\bar{x} = \frac{1}{n}(X_1 + \dots + X_n)$, and

- by linearity, $E[\bar{x}] = \frac{1}{n}(\mu + \dots + \mu) = \frac{1}{n}(n\mu) = \mu$;
- because variances of independent variables add, the sum has variance $n\sigma^2$, and dividing by n divides the variance by n^2 : $\text{Var}(\bar{x}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

Mean and Standard Deviation of $\bar{x}$

For random samples of size n from a population with mean μ and standard deviation σ ,

$$\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Because $\mu_{\bar{x}} = \mu$, we say $\bar{x}$ is an *unbiased* estimator: it does not tend to guess too high or too low. The formula $\frac{\sigma}{\sqrt{n}}$ has a sobering consequence, the *square-root law*: to cut the spread in half, you need four times as many observations. To make an estimate 10 times more precise, you need 100 times the data. This is the same $\frac{1}{\sqrt{n}}$ you saw in the Monte Carlo Simulation chapter, and it is the same reason.

6.4 The Central Limit Theorem

The third fact, about shape, is the deep one.

The Central Limit Theorem

When n is large, the sampling distribution of $\bar{x}$ is approximately normal, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, no matter what shape the population has.

How large is “large”? It depends on the population. If the population is itself normal, $\bar{x}$

is exactly normal for every n . If it is roughly symmetric, $n = 10$ or so is plenty. If it is strongly skewed, you need more. The common AP rule of thumb is $n \geq 30$, which is why the Statistical Inferences for Means chapter accepts $n \geq 30$ as a way to satisfy its “normal or large sample” condition.

The CLT also explains why so many things in nature are close to normal. A person’s height is the sum of many small effects from many genes and from nutrition. A measurement error is the sum of many small disturbances. Sums of many small, independent pieces are approximately normal, whatever the pieces look like.

Example: An elevator sign says “Maximum 12 people or 1,000 kg.” Adult weights have mean 80 kg and standard deviation 15 kg. If 12 randomly chosen adults get on, what is the chance they exceed 1,000 kg?

Solution: Twelve people exceed 1,000 kg exactly when their average weight exceeds $\frac{1000}{12} \approx 83.3$ kg. By the CLT, $\bar{x}$ is approximately normal with mean 80 and standard deviation $\frac{15}{\sqrt{12}} \approx 4.33$. So

$$z = \frac{83.3 - 80}{4.33} \approx 0.77, \quad P(\bar{x} > 83.3) = P(Z > 0.77) \approx 0.22$$

More than a one-in-five chance. That is why real elevators are built to carry far more than the posted limit.

Notice how differently individuals and averages behave. The chance that one random adult weighs more than 95 kg is $P(Z > 1) \approx 0.16$. The chance that the *average* of 4 adults is more than 95 kg is $P\left(Z > \frac{95-80}{15/\sqrt{4}}\right) = P(Z > 2) \approx 0.023$, about seven times less likely. Averages are hard to push to extremes.

6.5 The Sampling Distribution of a Proportion

A sample proportion $\hat{p}$ (“p-hat”) is secretly a mean. Code each person as 1 for “yes” and 0 for “no.” Then $\hat{p}$ is the average of those 0s and 1s, and the count of yeses, $X = n\hat{p}$, is binomial (from the Binomial and Geometric Distributions chapter). Dividing the binomial mean np and standard deviation $\sqrt{np(1-p)}$ by n :

Mean and Standard Deviation of $\hat{p}$

$$\mu_{\hat{p}} = p, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

The sampling distribution of $\hat{p}$ is approximately normal when $np \geq 10$ and $n(1-p) \geq 10$.

That last line is the CLT for 0/1 data, and it is the **large counts** condition you will check

before every confidence interval for a proportion in the next chapter.

Now we can answer the opening question. For a poll of $n = 1000$ with p near 0.5, $\sigma_{\hat{p}} = \sqrt{\frac{0.5 \cdot 0.5}{1000}} \approx 0.0158$. About 95% of a normal distribution lies within 2 standard deviations of the center, and $2(0.0158) \approx 0.03$: the famous “plus or minus 3 points.” Notice that the population size, 150 million, appears nowhere in the formula. A well-drawn random sample of 1,000 is just as accurate for a whole country as for a single large city.

Example: A candidate truly has 55% support. What is the chance that a random poll shows them losing ($\hat{p} < 0.5$)?

```

1 import numpy as np
2
3 rng = np.random.default_rng(2)
4 p = 0.55          # true support for the candidate
5
6 for n in [100, 1000]:
7     # 10,000 polls of n voters; p-hat is the fraction who say yes
8     p_hat = rng.binomial(n, p, size=10_000) / n
9     wrong = np.mean(p_hat < 0.5)
10    print(f"n = {n:4}: mean {p_hat.mean():.3f}, sd {p_hat.std():.4f}, "
11          f"polls showing a loss: {wrong:.4f}")

```

```

n = 100: mean 0.550, sd 0.0496, polls showing a loss: 0.1338
n = 1000: mean 0.550, sd 0.0158, polls showing a loss: 0.0005

```

With 100 voters, about 13% of polls get the winner wrong. With 1,000, only about 1 in 2,000 do. The normal model agrees well for $n = 1000$, where $P(Z < \frac{0.5 - 0.55}{0.0158}) = P(Z < -3.18) \approx 0.0007$. For $n = 100$, it gives 0.157, while the exact binomial answer is 0.135. The gap comes from $\hat{p}$ moving in steps of 0.01 with only 100 voters, so a smooth curve is a rougher fit. Larger samples make the normal approximation better.

Exercise 28 Center and Spread

A population has mean 50 and standard deviation 12. You take random samples of size 36.

Working Space

1. What are the mean and standard deviation of the sampling distribution of $\bar{x}$?
2. Find $P(\bar{x} > 53)$.
3. How large a sample would make $\sigma_{\bar{x}}$

at most 1?

Answer on Page 94

Exercise 29 **Watch Times**

Watch times on the video site have mean 10 minutes and standard deviation 10 minutes, and are strongly skewed. For a random sample of 40 videos, estimate the probability that the average watch time is more than 12 minutes. Why is it reasonable to use a normal model even though watch times are not normal?

Working Space

Answer on Page 95

Exercise 30 **True or False?**

A student says, “The Central Limit Theorem says that if I take a big enough sample, my data will be normally distributed.” Is that right? Explain.

Working Space

Answer on Page 95

Exercise 31 **Dice Averages**

Change `clt_simulation.py` so that each observation is a roll of a fair die, using `rng.integers(1, 7, size=(SAMPLES, n))`, and use sample sizes 1, 2, and 10. Describe the three histograms. (For one

Working Space

die, $\mu = 3.5$ and $\sigma \approx 1.71$.)

Answer on Page 95

Simulation-Based Inference

You survey 20 classmates about how much they sleep, and the average is 6.8 hours. How far off from the true average for your whole school might that be? A friend runs an experiment and finds that students who studied in silence scored 2 points higher on a quiz than students who studied with music. Is that a real effect, or could it be luck?

Statisticians have formulas for questions like these, and the next chapters teach them. But the ideas behind the formulas are simpler than the formulas themselves, and with a computer you can use the ideas directly. This chapter shows two techniques that answer both questions with nothing but resampling and shuffling: the *bootstrap* and the *permutation test*. They were invented only about 50 years ago, once computers became fast enough, and today they are everyday tools in science and data analysis.

7.1 The Bootstrap

In the Sampling Distributions and the Central Limit Theorem chapter, we learned how much a sample mean varies by drawing thousands of samples from a population. In real life, you cannot do that: you have one sample, and the population is unknown.

The bootstrap's trick is to let the sample stand in for the population. If the sample was chosen at random, it is our best picture of what the population looks like. So we draw new samples *from the sample*, of the same size, **with replacement**. Each *bootstrap sample* uses the same values, but some appear twice or more and some not at all, just as a fresh random sample from the population would contain different people. Computing the statistic for thousands of bootstrap samples shows how much it bounces around. The name comes from the phrase “pulling yourself up by your own bootstraps”: the sample estimates its own uncertainty.

Example: Twenty randomly chosen students report how many hours they slept. Estimate the average for the whole school, with a range of plausible values.

```
1 import numpy as np
2
3 rng = np.random.default_rng(4)
4
5 # Hours of sleep reported by 20 randomly chosen students
6 sleep = np.array([6.5, 7.0, 5.5, 8.0, 6.0, 7.5, 6.5, 9.0, 5.0, 7.0,
7                  6.0, 8.5, 7.0, 6.5, 4.5, 7.5, 6.0, 7.0, 8.0, 6.5])
8 n = len(sleep)
```

```

9 print(f"sample mean: {sleep.mean():.3f}")
10
11 # 10,000 bootstrap samples: n values drawn WITH replacement from the sample
12 resamples = rng.choice(sleep, size=(10_000, n), replace=True)
13 boot_means = resamples.mean(axis=1)
14
15 low, high = np.percentile(boot_means, [2.5, 97.5])
16 print(f"bootstrap SE: {boot_means.std():.3f}")
17 print(f"s / sqrt(n): {sleep.std(ddof=1) / np.sqrt(n):.3f}")
18 print(f"95% bootstrap interval: ({low:.2f}, {high:.2f})")

```

```

sample mean: 6.775
bootstrap SE: 0.250
s / sqrt(n): 0.253
95% bootstrap interval: (6.28, 7.25)

```

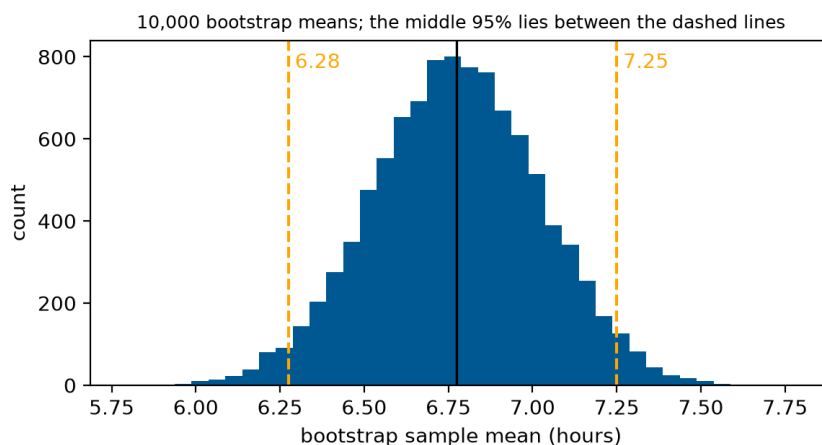


Figure 7.1: The bootstrap distribution of the sample mean. The black line is the original sample mean, and the dashed lines cut off the lowest and highest 2.5%.

Three things to notice:

- The bootstrap means pile up in a bell shape around the sample mean, just as the CLT predicts.
- Their standard deviation, the *bootstrap standard error*, is 0.250 hours. The CLT chapter's formula $\frac{\sigma}{\sqrt{n}}$, using the sample's standard deviation s in place of the unknown σ , gives 0.253. The simulation rediscovered the formula.
- The middle 95% of the bootstrap means runs from 6.28 to 7.25 hours. This is a *95% bootstrap percentile interval*: a range of plausible values for the school's true average sleep. We are fairly confident the true average is somewhere in it. The next chapter calls this kind of range a *confidence interval* and gives formulas for it.

7.1.1 Beyond the Mean

The real power of the bootstrap is that the same code works for any statistic, even ones with no simple formula. Change one line to bootstrap the *median*, the 90th percentile, or the correlation between two variables, and you get an interval for that too. Statisticians with only formulas would be stuck.

7.1.2 What the Bootstrap Cannot Do

The bootstrap can only reuse the information already in the sample. If the sample was not random (say, you asked only your friends), resampling it thousands of times reproduces the same bias thousands of times. And with a very small sample, a handful of values cannot represent the population's shape well, so the bootstrap interval can be too narrow. Random sampling still comes first.

Exercise 32 Bootstrap Samples by Hand

A tiny sample has the values 3, 5, and 8.

Working Space

1. Write down three different possible bootstrap samples.
2. Can a bootstrap sample be 8, 8, 8? Can it be 3, 5, 9? Explain.
3. What is the largest possible bootstrap sample mean?

Answer on Page 95

Exercise 33 Bootstrapping a Median

Change `bootstrap.py` to compute a 95% bootstrap interval for the median hours of sleep. (Hint: `np.median(resamples, axis=1)`.)

Working Space

Answer on Page 95

7.2 The Permutation Test

Now the experiment. Twenty-four students were randomly assigned to study for a quiz either in silence or with music, 12 in each group.

Silence	15, 17, 14, 16, 18, 13, 17, 15, 16, 19, 14, 16
Music	13, 15, 12, 16, 14, 11, 15, 13, 17, 14, 12, 15

The silence group's mean is 15.83 and the music group's is 13.92, a difference of 1.92 points.

Here is the skeptic's argument. "Music makes no difference. Each student would have gotten the same score in either group. The groups differ only because of which students the random assignment happened to put where." That claim is the *null hypothesis*.

If the skeptic is right, the labels "silence" and "music" are meaningless. So we can test the claim directly: pool all 24 scores, shuffle them, deal them into two new groups of 12, and compute the difference in means. That is one *permutation*. Repeat 10,000 times, and you see what differences random assignment alone produces when music truly does nothing.

```

1 import numpy as np
2
3 rng = np.random.default_rng(5)
4
5 # Quiz scores (out of 20) in a randomized experiment
6 silence = np.array([15, 17, 14, 16, 18, 13, 17, 15, 16, 19, 14, 16])
7 music = np.array([13, 15, 12, 16, 14, 11, 15, 13, 17, 14, 12, 15])
8 observed = silence.mean() - music.mean()
9 print(f"observed difference: {observed:.3f}")
10
11 # If music makes no difference, the labels are meaningless: shuffle them.
12 scores = np.concatenate([silence, music])
13 diffs = np.empty(10_000)
14 for i in range(10_000):
15     shuffled = rng.permutation(scores)
16     diffs[i] = shuffled[:12].mean() - shuffled[12:].mean()
17
18 p_value = np.mean(diffs >= observed)
19 print(f"p-value: {p_value:.4f}")

```

```

observed difference: 1.917
p-value: 0.0083

```

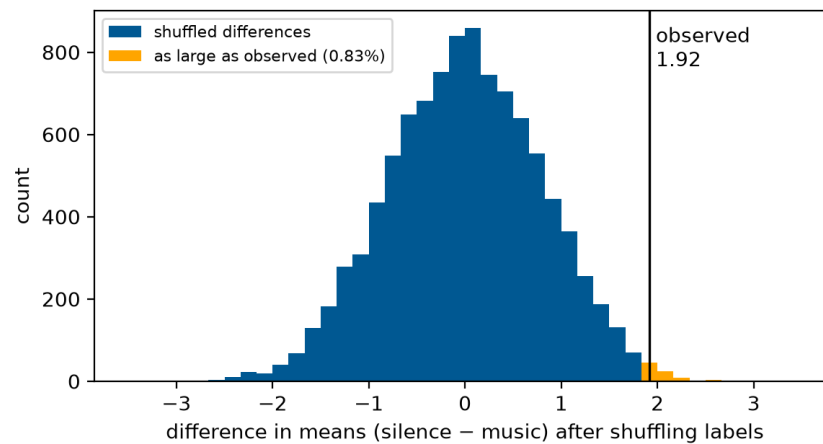


Figure 7.2: Differences in means after shuffling the labels 10,000 times. Only 0.83% of the shuffles produced a difference as large as the real one.

The shuffled differences cluster around 0, as they should if labels do not matter. A difference of 1.92 or more came up in only 83 of the 10,000 shuffles. That fraction, 0.0083, is the *p-value*: the probability, if music makes no difference, of seeing a difference at least as large as the one observed. It is so small that the skeptic's story is hard to believe. We have convincing evidence that, for students like these, studying in silence leads to higher quiz scores. Because the groups were formed by random assignment, the difference was caused by the music, not by some other difference between the groups.

The Inference for Two Means chapter later in this volume teaches a formula-based version of this test, the two-sample *t*-test. For these data, it gives a *p*-value of about 0.007, close to our simulation's 0.008.

7.2.1 Why Shuffling Works

Shuffling is not a trick; it repeats the experiment's own randomness. The random assignment is the only reason the two groups could differ if music had no effect. By re-doing the assignment thousands of times, we build the distribution of differences that the null hypothesis predicts, with no formulas and no assumptions about bell curves. That is why permutation tests are trusted even for small samples and skewed data.

Exercise 34 A New Warm-Up Routine

A coach randomly assigns 40 runners to a new warm-up routine or the old one, 20 each. In the new group, 10 runners improved their times; in the old group, 4

Working Space

did. Write a permutation test to estimate the p-value for the claim that the new routine helps. What do you conclude?

Answer on Page 96

Exercise 35 **What the p-Value Means**

A student says, “The p-value of 0.0083 means there is a 0.83% chance that music has no effect.” Explain what is wrong with that statement, and give a correct interpretation.

Working Space

Answer on Page 96

Exercise 36 **Bootstrap or Permutation?**

For each question, say whether a bootstrap or a permutation test fits better.

Working Space

1. From a random sample of 50 phone users, estimate the median daily screen time with a range of plausible values.
2. In a randomized experiment, did plants given a new fertilizer grow taller than plants given plain water?

Answer on Page 96

Answers to Exercises

Answer to Exercise 1 (on page 5)

Following the structure shown in the formal definition of a definite integral, we can set $f(x) = x^3 + x \sin x$ and rewrite the limit of the sum as $\lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x = \int_0^\pi f(x) dx$. Therefore, the full definite integral would be written as $\int_0^\pi (x^3 + x \sin x) dx$.

Answer to Exercise 2 (on page 15)

By the difference property,

$$\int_0^1 (5 - 6x^2) dx = \int_0^1 5 dx - \int_0^1 6x^2 dx$$

By the constant multiple property,

$$\int_0^1 5 dx - \int_0^1 6x^2 dx = \int_0^1 5 dx - 6 \int_0^1 x^2 dx$$

By the integral of a constant,

$$\int_0^1 5 dx = 5(1 - 0) = 5$$

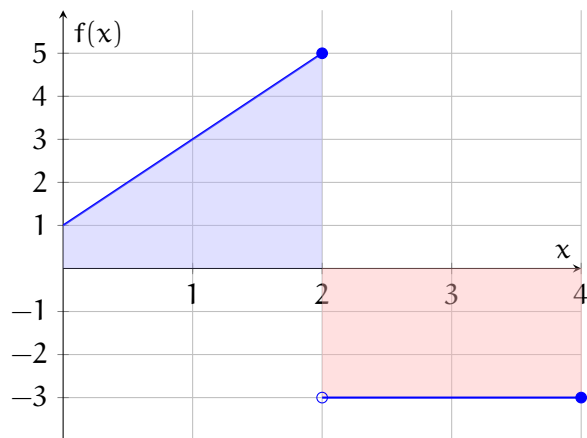
Putting it all together, we see that

$$\int_0^1 (5 - 6x^2) dx = 5 - 6\left(\frac{1}{3}\right) = 5 - 2 = 3$$

Answer to Exercise 3 (on page 15)

We can break the integral into two parts: from $x = 0$ to $x = 2$ (shaded in blue), and from $x = 2$ to $x = 4$ (shaded in red). The blue portion is a trapezoid, so it has a total area of $\frac{1}{2}(b_1 + b_2)(h) = \frac{1}{2}(1 + 5)(2) = 6$. Because it is above the x -axis, the area is positive. The

red portion is a rectangle and has a total area of $2 \times 3 = 6$ and is *negative*, because it lies below the x -axis. Therefore, the total area is $6 + -6 = 0$.



Answer to Exercise 4 (on page 15)

Using the properties of integrals, we can rewrite $\int_{-1}^9 3g(x) + 2 \, dx$ as $3 \int_{-1}^9 g(x) \, dx + 2(9 - (-1))$. From the graph, we can determine $\int_{-1}^9 g(x) \, dx = 2.5$. Therefore, $\int_{-1}^9 3g(x) + 2 \, dx = 3(2.5) + 2(10) = 27.5$.

Answer to Exercise 5 (on page 16)

(B). If $f'(x) > 0$ for all x , then f must be increasing for $4 < x < 7$. Since (C) decreases from $x = 5$ to $x = 7$, we can eliminate it. We can also eliminate (D), since $f(4) = f(5) = f(7)$, which implies either the slope of f is zero or changes from positive to negative. If $\int_4^7 f(x) \, dx = 0$, then some portion of $f(x)$ lies above the x -axis, while some other portion lies below (we must have positive and negative areas for the sum to be zero). This eliminates (A) and (E), since the integral of (A) would have a negative value and the integral of (E) would have a positive value. This leaves (B).

Answer to Exercise 6 (on page 20)

(D) 2. First, we try to compute the limit directly: $\lim_{x \rightarrow 1} \frac{\int_1^x g(t) \, dt}{g(x) - 6} = \frac{\int_1^1 g(t) \, dt}{6 - 6} = \frac{0}{0}$, which is undefined. Because g is continuous and differentiable, we can apply l'Hôpital's rule, differentiating the top and the bottom separately. By the FTC, the derivative of the top is $g(x)$, and the derivative of the bottom is $g'(x)$: $\lim_{x \rightarrow 1} \frac{\int_1^x g(t) \, dt}{g(x) - 6} = \lim_{x \rightarrow 1} \frac{g(x)}{g'(x)} = \frac{g(1)}{g'(1)} =$

$$\frac{6}{3} = 2.$$

Answer to Exercise 7 (on page 21)

1. $g'(x) = \sqrt{x + x^3}$
2. $F(x) = -\int_0^x \sqrt{1 + \sec t} \, dt$ and therefore $F'(x) = -\sqrt{1 + \sec x}$
3. setting $u = e^x$ and noting $\frac{du}{dx} = e^x$, then $h'(x) = \frac{d}{du} \int_1^u \ln t \, dt \left(\frac{du}{dx} \right)$ Taking the derivative and substituting for $\frac{du}{dx}$, we find $h'(x) = \ln u \cdot e^x = \ln(e^x) \cdot e^x = x \cdot e^x$
4. $y = -\int_{\frac{\pi}{4}}^{\sqrt{x}} \theta \tan \theta \, d\theta$. Setting $u = \sqrt{x}$ and noting that $\frac{du}{dx} = \frac{1}{2\sqrt{x}}$, we see that $y' = -\frac{d}{du} \left[\int_{\frac{\pi}{4}}^u \theta \tan \theta \, d\theta \right] \frac{du}{dx} = -u \tan u \cdot \frac{1}{2\sqrt{x}} = -\sqrt{x} \tan \sqrt{x} \cdot \frac{1}{2\sqrt{x}} = \frac{-\sqrt{x} \tan \sqrt{x}}{2\sqrt{x}} = -\frac{1}{2} \tan \sqrt{x}$

Answer to Exercise 8 (on page 22)

1. The antiderivative of $t^{-3/2}$ is $\frac{-2}{\sqrt{t}}$. Therefore, the integral is equal to $\left[\frac{-2}{\sqrt{t}} \right]_1^4 = \frac{-2}{\sqrt{4}} - \frac{-2}{\sqrt{1}} = -1 + 2 = 1$.

Answer to Exercise 9 (on page 23)

According to FTC, $h'(x) = f(x)$ and $h''(x) = f'(x)$. Examining the graph, we see that the curve lies below the x -axis for $0 < x < 6$, which means that $h(6) = \int_0^6 f(t) \, dt < 0$. $h'(6) = f(6) = 0$ and $h''(6) = f'(6) > 0$. Therefore, $h(6) < h'(6) < h''(6)$.

Answer to Exercise 10 (on page 23)

We know that $f(2) = \int_{-5}^2 f'(x) \, dx + f(-5)$. Examining the graph, we know that $\int_{-5}^2 f'(x) \, dx = \frac{1}{2}(3)(2) - \frac{1}{2}\pi(2^2)$ (the area of the triangle above the x -axis less the area of the semi-circle below the axis). Therefore, $f(-5) = f(2) - \int_{-5}^2 f'(x) \, dx = 1 - (3 - 2\pi) = 2\pi - 2$

Answer to Exercise 11 (on page 24)

11.71. The particle's velocity will be given by its initial velocity plus the integral of its

acceleration over the time period. Therefore, $v(3) = v(0) + \int_0^3 a(t) dt = 5 + \int_0^3 \frac{t+3}{\sqrt{t^3+1}} dt \approx 5 + 6.71 = 11.71$.

Answer to Exercise 12 (on page 25)

We begin by setting up the integral for average value of a function with $a = 0$, $b = \frac{\pi}{2}$, and $f(x) = \sqrt{\cos x}$:

$$\begin{aligned} \frac{1}{\pi/2 - 0} \int_0^{\pi/2} \sqrt{\cos x} dx \\ = \frac{2}{\pi} \int_0^{\pi/2} \sqrt{\cos x} dx \end{aligned}$$

There is not an obvious way to evaluate this integral by hand. Luckily, this question allows for the use of a calculator. Entering this integral into a calculator (such as a TI-89 or Wolfram Alpha), we find that:

$$\frac{2}{\pi} \int_0^{\pi/2} \sqrt{\cos x} dx \approx 0.763$$

Answer to Exercise 13 (on page 26)

If $F = bt$, then $a(t) = \frac{bt}{m}$. Since the particle starts from rest, we know that $v_0 = 0 \frac{m}{s}$. And we know the relationship between acceleration and velocity:

$$v(t) = v_0 + \int_0^t a(s) ds$$

Substituting for v_0 and a :

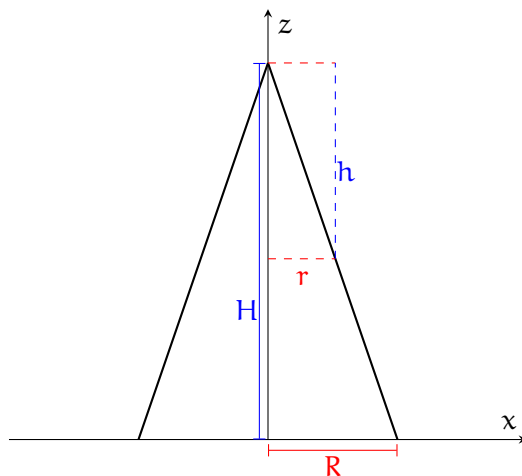
$$v(t) = 0 + \int_0^t \frac{bs}{m} ds = \frac{bt^2}{2m}$$

Answer to Exercise 14 (on page 28)

Imagine a side view of the cone (see figure below), an isosceles triangle with height H and base $2R$. If we take horizontal cross-sections, then each cross-section is a circle h from the top with a radius r . Because the triangles are similar (the small triangle from the tip down to the slice and the whole side view share the angle at the tip), we also know that $\frac{H}{h} = \frac{R}{r}$. Therefore, we can define r in terms of h : $r = \frac{hR}{H}$ and the volume of each

subsequent cross-section is $\pi r^2 dh = \pi \frac{h^2 R^2}{H^2} dh$. We start with $h = 0$ and end with $h = H$:

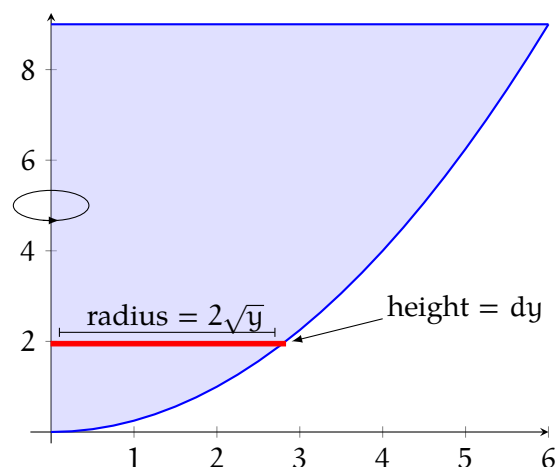
$$\begin{aligned} V_{\text{cone}} &= \int_0^H \pi \frac{h^2 R^2}{H^2} dh = \pi \frac{R^2}{H^2} \int_0^H h^2 dh \\ &= \pi \frac{R^2}{H^2} \left[\frac{1}{3} h^3 \right]_{h=0}^{h=H} = \pi \frac{R^2}{3H^2} [H^3 - 0^3] \\ &= \pi \frac{R^2}{3H^2} H^3 = \frac{\pi}{3} R^2 H \end{aligned}$$



Answer to Exercise 15 (on page 30)

If we are rotating about the y axis, we should make our slices horizontal, so their width is dy (see graph below). Then, the volume of each cylinder is given by $V = \pi r^2 dy$ and the total volume is given by:

$$\begin{aligned} V &= \int_0^9 \pi [2\sqrt{y}]^2 dy \\ V &= 4\pi \int_0^9 y dy = 2\pi y^2 \Big|_{y=0}^{y=9} \\ V &= 2\pi (9)^2 = 162\pi \end{aligned}$$



Answer to Exercise 16 (on page 31)

Since the graph is rotated around the x -axis, we will take vertical slices with width dx , and rotate them to make cylinders with radius $f(x)$ and height dx . The volume of each egg is given by:

$$\int_{-1}^1 \pi [f(x)]^2 dx$$

To determine our limits of integration, we note that $\sqrt{1-x^2} = 0$ (and therefore, $f(x) = 0$) when $x = \pm 1$.

For the chicken:

$$V_{\text{chickenegg}} = \pi \int_{-1}^1 \left[(-0.02x^3 + 0.03x^2 + 0.12x + 0.454) \sqrt{1-x^2} \right]^2 dx$$

For the mallard duck:

$$V_{\text{duckegg}} = \pi \int_{-1}^1 \left[(-0.06x^3 + 0.04x^2 + 0.1x + 0.54) \sqrt{1-x^2} \right]^2 dx$$

Using a calculator, we find that $V_{\text{chickenegg}} \approx 0.897$ and $V_{\text{duckegg}} \approx 1.263$. Therefore, mallard ducks lay larger eggs than chickens do.

Answer to Exercise 17 (on page 32)

First, since we are revolving around the y -axis, we know our slices will have width dy .

We will rewrite the functions as x in terms of y :

$$x = \sqrt{y}$$

$$x = \frac{y^2}{4}$$

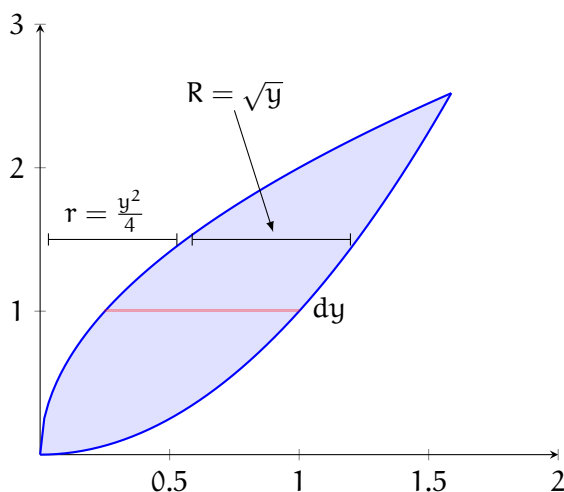
Setting them equal to each other to find the y -value at which they intercept:

$$\sqrt{y} = \frac{y^2}{4}$$

$$4 = \frac{y^2}{\sqrt{y}} = y^{3/2}$$

$$y = \sqrt[3]{4^2} = 2\sqrt[3]{2}$$

Examining a graph (shown below), we see that the outer radius is $x = \sqrt{y}$ and the inner radius is $x = \frac{y^2}{4}$.



So, the total volume of the solid of revolution is given by:

$$V = \pi \int_0^{2\sqrt[3]{2}} (\sqrt{y})^2 - \left(\frac{y^2}{4}\right)^2 dy$$

$$V = \pi \int_0^{2\sqrt[3]{2}} \left[y - \frac{y^4}{16} \right] dy$$

$$V = \pi \left[\frac{1}{2}y^2 - \frac{1}{80}y^5 \right]_{y=0}^{y=2\sqrt[3]{2}}$$

$$V = \pi \left[\frac{6}{5}2^{2/3} \right] \approx 5.9844$$

Answer to Exercise 18 (on page 33)

If each cross section is a square, then the volume of each cross section is given by $s^2 dx$, where s is the side length of the square. Since the side length is equal to the distance between the graph of y and the x -axis, we can see that $s = y = \ln(3 - x)$. And, therefore, the total volume of all the cross sections is given by $\int_0^2 [\ln(3 - x)]^2 dx$. Using a calculator, this integral evaluates to ≈ 1.029 .

Answer to Exercise 19 (on page 34)

Since the cross-sections are perpendicular to the x -axis, they will have width dx and we will integrate across the domain of the ellipse. Setting $y = 0$ to find the domain of the ellipse:

$$9x^2 = 25 \rightarrow x^2 = \frac{25}{9} \rightarrow x = \pm \frac{5}{3}$$

A right isosceles triangle with hypotenuse h has area $\frac{1}{4}h^2$. In this case, each triangle's hypotenuse is given by the distance between the top and bottom of the ellipse. The top of the ellipse is defined by $y = \frac{1}{4}\sqrt{25 - 9x^2}$ and the bottom by $y = -\frac{1}{4}\sqrt{25 - 9x^2}$. Therefore, the length of each hypotenuse is $\frac{1}{2}\sqrt{25 - 9x^2}$.

Then, each cross-section has a total volume of $\frac{1}{4}h^2 dx = \frac{1}{4} \left(\frac{1}{2}\sqrt{25 - 9x^2} \right)^2 dx$ and the volume of the solid is:

$$\begin{aligned} V_{\text{solid}} &= \int_{-5/3}^{5/3} \frac{1}{4} \left(\frac{1}{2}\sqrt{25 - 9x^2} \right)^2 dx \\ &= \frac{1}{4} \int_{-5/3}^{5/3} \frac{1}{4} (25 - 9x^2) dx \\ &= \frac{1}{16} \int_{-5/3}^{5/3} (25 - 9x^2) dx = \frac{1}{16} [25x - 3x^3]_{x=-5/3}^{x=5/3} \\ &= \frac{1}{16} \left[\left(25 \left(\frac{5}{3} \right) - 25 \left(\frac{-5}{3} \right) \right) - \left(3 \left(\frac{5}{3} \right)^3 - 3 \left(\frac{-5}{3} \right)^3 \right) \right] \\ &= \frac{1}{16} \left[\frac{250}{3} - \left(\frac{375}{27} + \frac{375}{27} \right) \right] = \frac{1}{16} \left[\frac{250}{3} - \frac{250}{9} \right] = \frac{1}{16} \left[\frac{750}{9} - \frac{250}{9} \right] = \frac{1}{16} \left[\frac{500}{9} \right] = \frac{125}{36} \end{aligned}$$

Answer to Exercise 20 (on page 42)

1. $L = \int_1^3 \sqrt{1 + \frac{1}{x^2}} \, dx$
2. $L = \int_0^\pi \sqrt{1 + \cos^2 x} \, dx$
3. $L = \int_1^4 \sqrt{1 + (x^2 - \frac{1}{4x^2})^2} \, dx$
4. $L = \int_0^{\frac{\pi}{3}} \sqrt{1 + (\frac{1}{\cos x} \times -\sin x)^2} \, dx = \int_0^{\frac{\pi}{3}} \sqrt{1 + \tan^2 x} \, dx$

Answer to Exercise 21 (on page 43)

From looking at the structure of the given arc length integral, we can see that $f'(t) = \sqrt{3t+4}$. Taking the antiderivative, we find that $f(x) = \frac{2}{9}(3x+4)^{3/2} + C$. Substituting $f(0) = 2$, we can solve for C .

$$2 = \frac{2}{9}(3(0) + 4)^{3/2} + C$$

$$2 = \frac{2}{9}(4)^{3/2} + C$$

$$2 = \frac{2}{9}(2)^3 + C$$

$$2 = \frac{16}{9} + C$$

$$\frac{18}{9} = \frac{16 + 9C}{9}$$

$$18 = 16 + 9C$$

$$2 = 9C$$

$$C = \frac{2}{9}$$

Therefore, $f(x) = \frac{2}{9}(3x+4)^{3/2} + \frac{2}{9}$. To find the coordinate point where $s(x) = 3$, we first note that the antiderivative of $\sqrt{3t+5}$ is $\frac{2}{9}(3t+5)^{3/2} + C$. Therefore, $s(x) = \frac{2}{9}(3x+5)^{3/2} - \frac{2}{9}(5)^{3/2}$. Setting $s(x) = 3$ and solving for x , we find that $x = 1.159$. The point on f is $(1.159, 4.765)$.

Answer to Exercise 22 (on page 43)

Recall that arc length is given by $\int_a^b \sqrt{1 + [f'(x)]^2} \, dx$. Since $f(x) = \ln x$, then $f'(x) = \frac{1}{x}$. Taking $a = 1$, $b = 2$, and $f'(x) = \frac{1}{x}$, the integral that gives the length of the curve $y = \ln x$ on the specified interval is $\int_1^2 \sqrt{1 + \frac{1}{x^2}} \, dx$.

Answer to Exercise 23 (on page 43)

Since we are told the vector-valued velocity of the ship, we know that $\frac{dx}{dt} = -1412 \sin t$, $\frac{dy}{dt} = 1412 \cos t$, and $\frac{dz}{dt} = t$. The distance traveled in the first 60 seconds is given by $\int_0^{60} \sqrt{(-1412 \sin t)^2 + (1412 \cos t)^2 + t^2} dt$. Using a calculator, the integral evaluates to 84745 meters. The distance traveled in the second 60 seconds is given by $\int_{60}^{120} \sqrt{(-1412 \sin t)^2 + (1412 \cos t)^2 + t^2} dt$. Using a calculator, this integral evaluates to 84898 meters.

Answer to Exercise 24 (on page 54)

```

1  from scipy.stats import norm
2
3  # Constants
4  MEAN = 164.7
5  STD = 7.1
6
7  # What is the cutoff for the top decile?
8  cutoff = norm.ppf(0.9, loc=MEAN, scale=STD);
9  print(f"To be in the top 10 percent, you must be at least {cutoff:.2f} cm")
10
11 # What proportion of women are between 160cm and 165cm?
12 shorter_than_160 = norm.cdf(160, loc=MEAN, scale=STD)
13 shorter_than_165 = norm.cdf(165, loc=MEAN, scale=STD)
14 between = shorter_than_165 - shorter_than_160
15 print(f"{between * 100.0:.2f}% of adult women are between 160 and 165 cm.")

```

When run, this will give you:

```

1  > python3 women.py
2  To be in the top 10 percent, you must be at least 173.80 cm
3  26.29% of adult women are between 160 and 165 cm.

```

Answer to Exercise 25 (on page 57)

1. $\int_0^{15} f(t) dt$
2. $\int_{30}^{60} f(t) dt$
3. $\int_{45}^{\infty} f(t) dt$

4. the probability that you spend more than 20 minutes on your weekly chores.
5. the probability that you spend between 1 and 2 hours on your weekly chores.

Answer to Exercise 26 (on page 58)

1. We know that

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-2}^6 \frac{ke^{2-x}}{(1+e^{2-x})^2} dx$$

Let $u = e^{2-x}$, then $-du = e^{2-x} dx$. Substituting:

$$\begin{aligned} &= \int_{x=-2}^{x=6} \frac{-k}{(1+u)^2} du = k \left[\frac{1}{1+u} \right]_{x=-2}^{x=6} = k \left[\frac{1}{1+e^{2-x}} \right]_{x=-2}^{x=6} \\ &= k \left[\frac{1}{1+e^{-4}} - \frac{1}{1+e^4} \right] = \frac{k(e^4-1)}{1+e^4} = 1 \end{aligned}$$

Which implies that $k = \frac{1+e^4}{e^4-1}$. Substituting for k , we find the probability that $X \geq 2$ is:

$$\begin{aligned} P(X \geq 2) &= \frac{1+e^4}{e^4-1} \int_2^6 \frac{e^{2-x}}{(1+e^{2-x})^2} dx = \frac{1+e^4}{e^4-1} \left[\frac{1}{1+e^{2-x}} \right]_{x=2}^{x=6} \\ &= \frac{1+e^4}{e^4-1} \left(\frac{1}{1+e^{-4}} - \frac{1}{1+1} \right) = \frac{1+e^4}{e^4-1} \left(\frac{1}{1+e^{-4}} - \frac{1}{2} \right) \approx 0.5 \end{aligned}$$

There is a 50% chance the variable will be greater than 2.

2. We know that:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \frac{k}{2+4x^2} dx = \frac{k}{2} \int_{-\infty}^{\infty} \frac{1}{1+2x^2} dx \\ &= \frac{k}{2} \left[\frac{\arctan(\sqrt{2}x)}{\sqrt{2}} \right]_{x=-\infty}^{x=\infty} = \frac{k}{2} \left[\frac{\arctan(\infty)}{\sqrt{2}} - \frac{\arctan(-\infty)}{\sqrt{2}} \right] \\ &= \frac{k}{2} \left(\frac{\pi}{2\sqrt{2}} - \frac{-\pi}{2\sqrt{2}} \right) = \frac{k}{2} \left(\frac{\pi}{\sqrt{2}} \right) = \frac{\pi k}{2\sqrt{2}} = 1 \end{aligned}$$

Which implies that $k = \frac{2\sqrt{2}}{\pi}$. Therefore $f(x) = \frac{\sqrt{2}}{\pi(1+2x^2)}$. Additionally,

$$\begin{aligned} P(1 \leq X \leq 4) &= \int_1^4 \frac{\sqrt{2}}{\pi(1+2x^2)} dx = \frac{\sqrt{2}}{\pi} \int_1^4 \frac{1}{1+2x^2} dx \\ &= \frac{\sqrt{2}}{\pi} \left[\frac{\arctan(\sqrt{2}x)}{\sqrt{2}} \right]_{x=1}^{x=4} = \frac{1}{\pi} \left[\arctan(4\sqrt{2}) - \arctan(\sqrt{2}) \right] \approx \frac{0.441}{\pi} \approx 0.140 = 14\% \end{aligned}$$

3. We know that:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} k(6x - 3x^2) dx = 3k \int_0^2 2x - x^2 dx \\ &= 3k \left[x^2 - \frac{1}{3}x^3 \right]_{x=0}^{x=2} = 3k \left(4 - \frac{8}{3} \right) = 3k \left(\frac{4}{3} \right) = 4k = 1 \end{aligned}$$

Which implies that $k = 1/4$. And therefore,

$$\begin{aligned} P(X > \frac{3}{4}) &= \int_{3/4}^2 \frac{6x - 3x^2}{4} dx = \frac{3}{4} \int_{3/4}^2 2x - x^2 dx \\ &= \frac{3}{4} \left[x^2 - \frac{1}{3}x^3 \right]_{x=3/4}^{x=2} = \frac{3}{4} \left[(2)^2 - \left(\frac{3}{4}\right)^2 - \frac{1}{3} \left(2^3 - \left(\frac{3}{4}\right)^3 \right) \right] \\ &= \frac{3}{4} \left(\frac{175}{192} \right) = \frac{175}{256} \approx 68.4\% \end{aligned}$$

Answer to Exercise 27 (on page 63)

(a) The fraction of acceptable parts is

$$P(6.72 \leq X \leq 12.28).$$

(b) Since $\text{NORM.DIST}(x, \text{mean}, \text{sd}, \text{TRUE})$ returns $P(X \leq x)$, we compute

$$P(6.72 \leq X \leq 12.28) = P(X \leq 12.28) - P(X \leq 6.72).$$

In Excel, enter:

$$=\text{NORM.DIST}(12.28, 10, 2, \text{TRUE}) - \text{NORM.DIST}(6.72, 10, 2, \text{TRUE}).$$

(c) Excel returns approximately 0.8224, so the percentage acceptable is

$$82.24\%.$$

Answer to Exercise 28 (on page 73)

1. $\mu_{\bar{x}} = 50$ and $\sigma_{\bar{x}} = \frac{12}{\sqrt{36}} = 2$.

2. $z = \frac{53-50}{2} = 1.5$, so $P(\bar{x} > 53) = P(Z > 1.5) \approx 0.067$.
3. $\frac{12}{\sqrt{n}} \leq 1$ means $\sqrt{n} \geq 12$, so $n \geq 144$.

Answer to Exercise 29 (on page 74)

$\sigma_{\bar{x}} = \frac{10}{\sqrt{40}} \approx 1.58$, so $z = \frac{12-10}{1.58} \approx 1.26$ and $P(\bar{x} > 12) \approx P(Z > 1.26) \approx 0.10$. The normal model is reasonable because of the CLT: $n = 40$ is large enough that the sample mean is approximately normal even though the population is skewed. (The exact answer for this population is about 0.107, so the approximation is close.)

Answer to Exercise 30 (on page 74)

No. A bigger sample makes the *data* look more like the population, so a sample of watch times stays skewed no matter how large it is. The CLT is about the *sampling distribution of the mean*: if you took many samples and averaged each one, those averages would be approximately normal.

Answer to Exercise 31 (on page 74)

With $n = 1$, the histogram is flat: six equal bars. With $n = 2$, it is a triangle peaking at 3.5, the shape of the dice sums from the Discrete Probability chapter. With $n = 10$, it is already close to a bell curve centered at 3.5 with standard deviation about $\frac{1.71}{\sqrt{10}} \approx 0.54$. A flat, symmetric population needs only a small n for the CLT to kick in.

Answer to Exercise 32 (on page 79)

1. For example, 3, 5, 8; 5, 5, 3; and 8, 3, 8. Each has three values drawn with replacement from {3, 5, 8}.
2. Yes, 8, 8, 8 is possible, because values are drawn with replacement. No, 3, 5, 9 is impossible, because 9 is not in the original sample.
3. 8, from the bootstrap sample 8, 8, 8.

Answer to Exercise 33 (on page 79)

```
1 import numpy as np
2
3 rng = np.random.default_rng(6)
4 sleep = np.array([6.5, 7.0, 5.5, 8.0, 6.0, 7.5, 6.5, 9.0, 5.0, 7.0,
5                  6.0, 8.5, 7.0, 6.5, 4.5, 7.5, 6.0, 7.0, 8.0, 6.5])
6 resamples = rng.choice(sleep, size=(10_000, len(sleep)), replace=True)
7 boot_medians = np.median(resamples, axis=1)
8 print(np.median(sleep), np.percentile(boot_medians, [2.5, 97.5]))
```

The sample median is 6.75 hours, and the 95% bootstrap interval is from 6.25 to 7.25 hours. (The bootstrap medians can only take a few values here, because the data are all multiples of 0.5.)

Answer to Exercise 34 (on page 81)

```
1 import numpy as np
2
3 rng = np.random.default_rng(7)
4 # 1 = improved, 0 = did not; first 20 got the new routine, last 20 the old one
5 outcomes = np.array([1] * 10 + [0] * 10 + [1] * 4 + [0] * 16)
6 observed = 10 - 4          # 6 more students improved with the new routine
7
8 diffs = np.empty(10_000)
9 for i in range(10_000):
10     shuffled = rng.permutation(outcomes)
11     diffs[i] = shuffled[:20].sum() - shuffled[20:].sum()
12 print(np.mean(diffs >= observed))
```

It prints a p-value of about 0.047. If the routine made no difference, a gap of 6 or more improvers would happen only about 5% of the time, so this is some evidence that the new routine helps, though not overwhelming. (This is right at the usual 0.05 cutoff, which the next chapters explain.)

Answer to Exercise 35 (on page 82)

The p-value is computed *assuming* music has no effect, so it cannot be the probability of that assumption. Correct: if music had no effect on quiz scores, only about 0.83% of random assignments would produce a difference in means of 1.92 points or more.

Answer to Exercise 36 (on page 82)

1. Bootstrap: you want an interval for one statistic (the median) from one sample.
2. Permutation test: you want to know whether a difference between two randomly assigned groups is larger than chance alone would produce.



INDEX

- 95% bootstrap percentile interval, 78
- arc length, 35
 - formula, 38
 - parametric, 40
 - vector-valued, 39
- Archimedes, 35
- area under a constant, 7
- average value of a function, 24
- bootstrap, 77
- bootstrap sample, 77
- bootstrap standard error, 78
- Central Limit Theorem, 69
- central limit theorem, 62, 69
 - simulation, 69
- chain rule
 - with integrals, 20
- cross-sections, 32
- cumulative distribution function, 47, 63
 - python, 51
- definite integral, 5
 - antiderivatives, 21
 - riemann sum as a limit, 5
- definite integrals
 - areas of, 6
 - properties of, 6
 - sums and differences, 8
- disk method, 29
- displacement, 41
 - from velocity, 13
- distance traveled, 41
- Fresnel function, 20
- FTC, 17
- fundamental theorem of calculus, 17
 - first part, 17
 - second part, 17
- Gaussian distribution, 52
- half angle formula, 30
- impulse, 13
- integral
 - bounds estimate, 10
 - reversed limits, 12
- inverse CDF, 52
- mean, 65
- mode, 49
- net area, 5
- NORM.DIST, 62
- normal distribution, 61
 - example, 62
 - parameters, 61
- null hypothesis, 80
- p-value, 81
- parameter, 69
- percentile, 52

- permutation, 80
- permutation test, 77, 80
- probability density
 - versus probability, 50
- probability density function, 48, 61

- sample proportion, 72
- sampling distribution, 69
- sampling variability, 69
- simulation-based inference, 77
- solids of revolution, 29
 - disk method, 29
- speed, 40
 - of a parametric curve, 40
- square-root law, 71
- squeeze theorem, 19
- standard deviation, 65
- statistic, 69

- unbiased, 71
- uniform distribution, 49

- volume
 - cone, 28
- Volume Single Integral, 27
 - volume of a sphere, 27

- washer method, 31
- work
 - integral, 13