Calculus for Epidemiology:
From Area to Slope and Back Again

Eben Kenah

August 19, 2019

Abstract

This is a very complicated case, Maude. You know, a lotta ins, a lotta outs, a lotta what-have-yous. And a lot of, uh, strands to keep in my head, man. You know, a lotta strands in old Duder's head.

The Dude
The Big Lebowski (1998)

The Force will be with you. Always.

Obi-Wan Kenobi
Star Wars (1977)

Calculus is not easy, but it will be with you always. I have tried to explain why it is useful before giving any details of how it is done. The sections with an asterisk (*) are about the details. They will be there when you need them even if you don't read them the first time. Suggestions for further reading are at the end.

I could never have gone far in any science because on the path of every science the lion Mathematics lies in wait for you.

C. S. Lewis
Surprised by Joy (1955)

Wrong will be right when Aslan comes in sight,
At the sound of his roar, sorrows will be no more,
When he bares his teeth, winter meets its death,
And when he shakes his mane, we shall have spring again.

C. S. Lewis
The Lion, the Witch, and the Wardrobe (1950)

In *The Lion, the Witch, and the Wardrobe*, a children's novel by C. S. Lewis (1898–1963), the lion Aslan saves the land of Narnia from perpetual winter. If the world of science has an Aslan, it is mathematics. It can be demanding and intimidating, but it is also magnificent and good. Learning basic calculus will give you new eyes to see the landscape of epidemiology, helping you avoid confusion and dogma.

1 Areas and slopes in the world

Calculus is traditionally divided into two parts: *Integral* calculus is about calculating the area under a curve, where area above the x-axis is positive and area below is negative. *Differential* calculus is about calculating the slope of a curve, which is positive when the curve is increasing and negative when it is decreasing. These two branches have a profound connection that has made much of the modern world possible.

Before we talk about calculus itself, it will help to understand why area and slope are such useful concepts. We will begin by thinking carefully about a car driving on the x-axis starting at $x = 0$. The headlights of the car point toward positive values of x, and it has a transmission but no steering wheel.

1.1 Units

If our car goes forward at 25 miles per hour for one hour, it will end up at $x = 25$ miles. This is what "25 miles per hour" means:

$$25\,\frac{\text{miles}}{\text{hour}} \times 1\text{ hour} = 25\text{ miles}. \tag{1}$$

Notice how the time units cancel out just like equal numbers would, leaving a result in miles. To see what could go wrong, consider the following:

$$25\,\frac{\text{miles}}{\text{hour}} \times 60\text{ minutes} \neq 1{,}500\text{ miles}. \tag{2}$$

We have two options for fixing equation (2):

- We can change the units for speed to miles per minute:

$$\frac{25}{60}\ \frac{\text{miles}}{\text{minute}} \times 60 \text{ minutes} = 25 \text{ miles}. \tag{3}$$

- We can change the units for time to hours:

$$25\ \frac{\text{miles}}{\text{hour}} \times 60 \text{ minutes} \times \frac{1 \text{ hour}}{60 \text{ minutes}} = v \text{ miles}. \tag{4}$$

Units always multiply and cancel out like numbers. When adding things, they should always be expressed in identical units (i.e., add apples to apples). Using these rules to keep track of units is a great way to catch mistakes.

Not keeping track of units can cause serious problems. The Mars Climate Orbiter was a space probe designed, built, and launched at a cost of $327.6 million (in 1998 dollars) to study the Martian atmosphere.[1] On September 23, 1999, it fired its rockets to slow down and enter an orbit around Mars after a 9.5-month journey of 416 million miles. It went behind the planet 49 seconds earlier than expected and disappeared forever. An investigation found that the error occurred because Lockheed Martin software produced results in units of pound-force × seconds that were passed to NASA software written for units of newtons × seconds. There are roughly 4.448 newtons in one unit of pound-force. The orbiter got 33 miles too close to the surface of Mars, where it either burned up in the atmosphere or bounced back into space like a skipping stone.

Most equations in science do not explicitly specify units because they are valid for any sets of constants and units that are consistent with each other. Conversely, inconsistent units lead to errors like those that doomed the Mars Climate Orbiter. The creative and consistent use of units is the key to using areas and slopes to describe the world. In our car example, we will use *miles per hour* for speed and *hours* for time.

1.2 Area: Position from velocity

Let $v(t)$ be the velocity of our car at time t. We say "velocity" instead of "speed" because we let $v(t)$ be negative when the car is moving backwards.[2] If the car travels forward at a constant speed of 25 miles per hour, then

$$v(t) = 25 \text{ for each } t \in [0, 1] \tag{5}$$

where $t \in [0, 1]$ means that $0 \leq t \leq 1$. Figure 1 shows the graph of this function. Because the area under $v(t)$ forms a rectangle, we can calculate it by multiplying the height and width of the rectangle:

$$25\ \frac{\text{miles}}{\text{hour}} \times 1 \text{ hour} = 25 \text{ miles}. \tag{6}$$

[1] See `https://en.wikipedia.org/wiki/Mars_Climate_Orbiter`.

[2] The speed of the car is the absolute value of the velocity. If your car has an electronic speedometer, it will show your speed when driving in reverse.

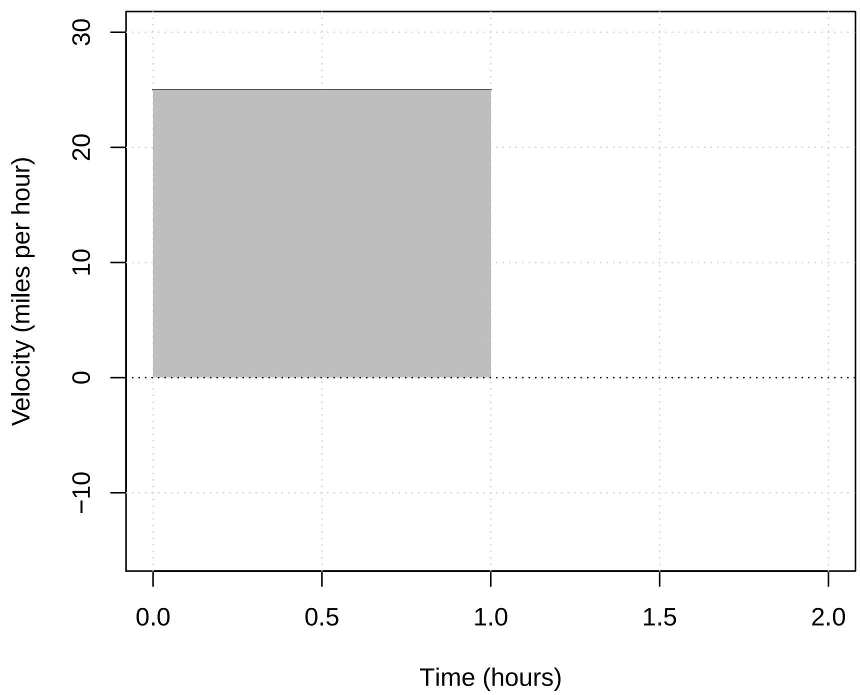

Figure 1: Graph of $v(t)$, the velocity of the car at time t in miles per hour, from $t = 0$ to $t = 1$ hour. The area under the graph is shaded.

Now suppose the car does not stop instantly at the end of one hour. Instead, it continues forward at 10 miles per hour from $t = 1$ hour to $t = 1.5$ hours. Our function $v(t)$ is now

$$v(t) = \begin{cases} 25 & \text{for } t \in [0, 1], \\ 10 & \text{for } t \in (1, 1.5] \end{cases} \tag{7}$$

where $t \in (1, 1.5]$ means $1 < t \leq 1.5$.[3] The position of the car at $t = 1.5$ is

$$25 \text{ miles} + \Big(0.5 \text{ hours} \times 10 \, \frac{\text{miles}}{\text{hour}}\Big) = 25 \text{ miles} + 5 \text{ miles} = 30 \text{ miles}. \tag{8}$$

Note how we added miles to miles to get a result in miles. The new plot of $v(t)$ is shown in Figure 2. The total area under the new graph of $v(t)$ is the sum of two rectangles:

$$\Big(25 \, \frac{\text{miles}}{\text{hour}} \times 1 \text{ hour}\Big) + \Big(10 \, \frac{\text{miles}}{\text{hour}} \times 0.5 \text{ hours}\Big) = 30 \text{ miles}. \tag{9}$$

Again, the area under the graph of $v(t)$ corresponds to the position of the car at the end of its journey.

Now suppose the car goes into reverse at $t = 1.5$ hours. It travels backward (toward lower values of x) at 10 miles per hour for half an hour. The velocity function is now

$$v(t) = \begin{cases} 25 & \text{for } t \in [0, 1], \\ 10 & \text{for } t \in (1, 1.5], \\ -10 & \text{for } t \in (1.5, 2]. \end{cases} \tag{10}$$

The position of the car at $t = 2$ hours is

$$30 \text{ miles} + \Big(-10 \, \frac{\text{miles}}{\text{hour}} \times 0.5 \text{ hours}\Big) = 30 \text{ miles} - 5 \text{ miles} = 25 \text{ miles}. \tag{11}$$

We added miles to miles as before, using negative miles to account for the car going backwards. The new graph of $v(t)$ is shown in Figure 3. The area under the graph of $v(t)$ now consists of two rectangles with positive area and one with negative area:

$$\Big(25 \, \frac{\text{miles}}{\text{hour}} \times 1 \text{ hour}\Big) + \Big(10 \, \frac{\text{miles}}{\text{hour}} \times 0.5 \text{ hours}\Big) + \Big(-10 \, \frac{\text{miles}}{\text{hour}} \times 0.5 \text{ hours}\Big) = 25 \text{ miles}. \tag{12}$$

As expected, the position of the car at $t = 2$ hours equals the area under the graph of $v(t)$ between $t = 0$ and $t = 2$ hours.

The position of the car at any time $T \in [0, 2]$ is the area under the graph of $v(t)$ between $t = 0$ and $t = T$. Remember that we assumed the car started at $x = 0$ at time $t = 0$. If the car started at $x = x_0$, we would have to add x_0 to the area under $v(t)$ between $t = 0$ and $t = T$ to get its position at time $t = T$.

[3] Does it matter if we say $v(1) = 25$ or $v(1) = 10$? Why or why not?

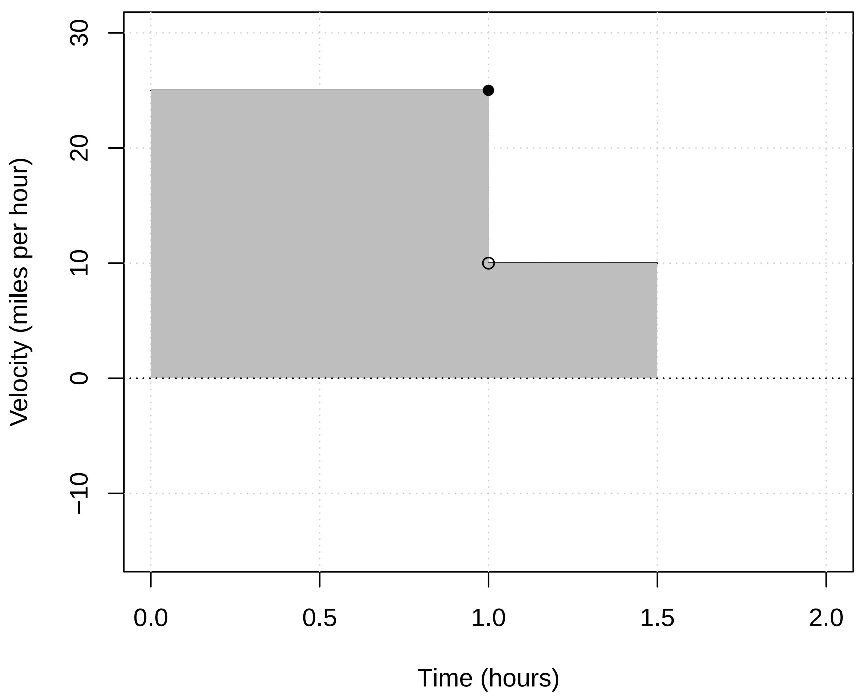

Figure 2: Graph of $v(t)$, the velocity of the car at time t in miles per hour, from $t = 0$ to $t = 1.5$ hours. The shaded area under the graph now consists of two rectangles, both with positive area. The solid and open circles at $t = 1$ hour indicate that $v(1) = 25$ miles per hour as in equation (7).

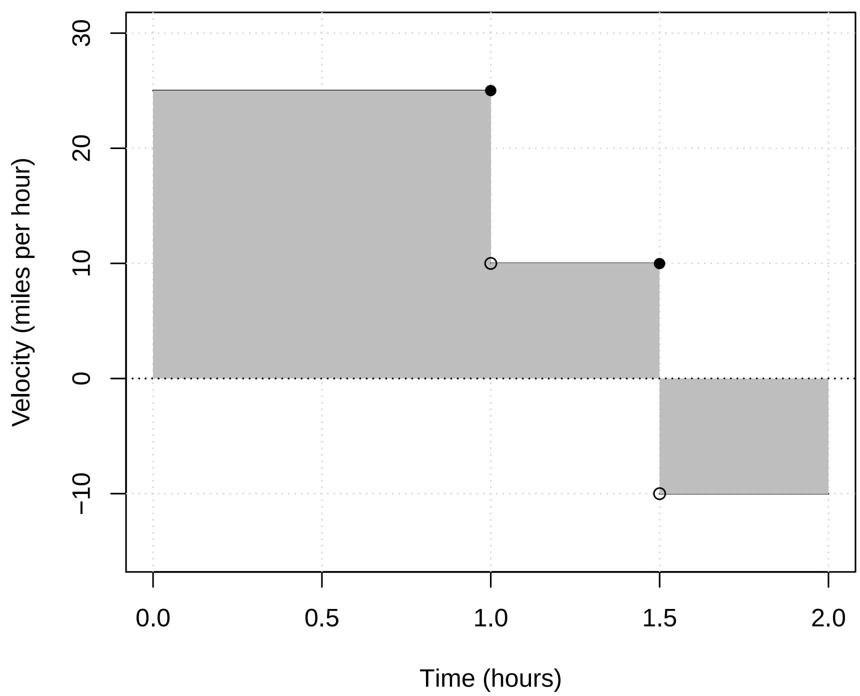

Figure 3: Graph of $v(t)$, the velocity of the car at time t in miles per hour, from $t = 0$ to $t = 2$ hours. The shaded area under the graph now consists of three rectangles, of which the last has negative area. The solid and open circles at $t = 1.5$ hours indicate that $v(1.5) = 10$ miles per hour as in equation (10).

1.3 Slope: Velocity from position

Now let $x(t)$ denote the position of the car at time t. In the last section, we figured out that we can calculate $x(t)$ using the area under the graph of the velocity function $v(t)$. In calculus notation, we have

$$x(t) = \int_0^t v(u)\,\mathrm{d}u. \tag{13}$$

There are three important things to notice about equation (13):

- The velocity function from the last section is written "$v(u)$" instead of "$v(t)$". The variable inside the parentheses is called the *argument* of the function, and can use anything that is convenient. The argument "t" was the first letter in "time", but the meaning stays exactly the same if we use "u" instead. This lets us use "t" as the argument of the position function.
- You can think of the "$v(u)\,\mathrm{d}u$" on the right as the area of a rectangle with height $v(u)$ and width $\mathrm{d}u$. The function $v(u)$ is called the **integrand**, and the $\mathrm{d}u$ (the "change in u") is called the **differential**.
- The **integral sign** $\int$ tells us to add up the areas of these rectangles between time 0 and time t, which are called the **limits of integration**. The lower limit is at the bottom and the upper limit is at the top of the integral sign. It is important to distinguish between the upper limit of integration t, which is the argument of $x(t)$, and the argument of the integrand $v(u)$.

The integral notation is a concise—almost poetic—description of everything we did in the last section.

Using the velocity function in equation (10) and Figure 3, we can calculate $x(t)$ for all $t \in [0, 2]$ by calculating the areas represented by the integral in equation (13). This gives us

$$x(t) = \begin{cases} 25t & \text{for } t \in [0,1], \\ 25 + 10(t-1) & \text{for } t \in (1, 1.5], \\ 30 - 10(t-1.5) & \text{for } t \in (1.5, 2]. \end{cases} \tag{14}$$

Figure 4 shows a plot of $x(t)$. In each interval during which the velocity $v(t)$ is constant, the plot of $x(t)$ is just a line segment. Thus, the graph of $x(t)$ is just three connected line segments.

The slope of a line segment is its vertical change (the "rise") divided by its horizontal change (the "run"). For the first line segment, we get

$$m_1 = \frac{25 \text{ miles}}{1 \text{ hour}} = 25\,\frac{\text{miles}}{\text{hour}}. \tag{15}$$

The correct units for speed came back when we calculated the slope! The slope of the second line segment is

$$m_2 = \frac{5 \text{ miles}}{0.5 \text{ hours}} = 10\,\frac{\text{miles}}{\text{hour}}, \tag{16}$$

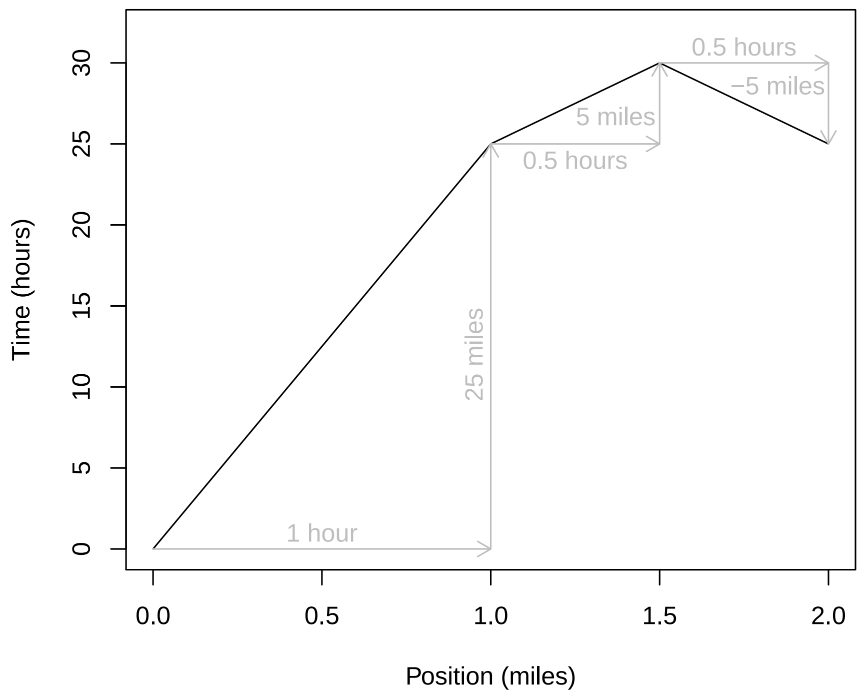

Figure 4: Graph of $x(t)$, the position of the car at time t in miles forward from the starting point. At each $t \in [0, 2]$, the position equals the area under the velocity function $v(u)$ from $u = 0$ to $u = t$, which is represented by the integral in equation (13). Over each interval in which velocity is constant, a horizontal arrow shows the change in time and a vertical arrow shows the change in position.

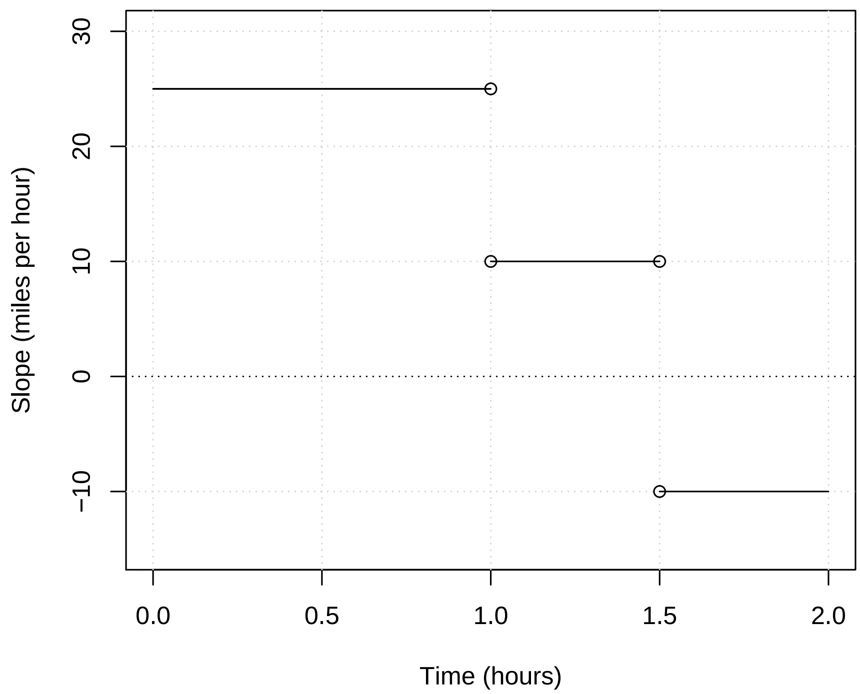

Figure 5: Graph of $x'(t)$, the slope of the position function $x(t)$. Note that the units for $x'(t)$ and the velocity $v(t)$ are the same. Inside each interval in which velocity is constant, $x'(t) = v(t)$. Open circles indicate that $x'(t)$ is undefined at $t = 1$ and $t = 1.5$, which is the only difference between it and $v(t)$.

and the slope of the third line segment is

$$m_3 = \frac{-5 \text{ miles}}{0.5 \text{ hours}} = -10 \ \frac{\text{miles}}{\text{hour}}. \tag{17}$$

We can define a function that gives us the slope of the graph of $x(t)$ in Figure 4 at each time t (except for $t = 1$ and $t = 1.5$ where two segments join):

$$x'(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t) = \begin{cases} 25 & \text{for } t \in [0, 1), \\ 10 & \text{for } t \in (1, 1.5), \\ -10 & \text{for } t \in (1.5, 2]. \end{cases} \tag{18}$$

The function $x'(t)$ is called the *derivative* of $x(t)$. Figure 5 shows a plot of $x'(t)$.

Comparing Figure 5 to Figure 3, we can see that they are the same at each point where $v(t)$ does not "jump". In other words:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^t v(u)\,\mathrm{d}u = v(t) \tag{19}$$

whenever $v(t)$ does not jump. This connection between area and slope is a version of the *fundamental theorem of calculus*, which is one of the most important problem-solving tools ever discovered. To see why it is so useful, we have to move beyond the world of rectangles and line segments.[4]

2 Basic concepts revisited

In Section 1, the velocity of the car was constant over time intervals and then changed instantly. Real cars can't change velocity instantly without violating almost every law of physics and every rule of the road. To develop a useful tool for describing and analyzing the world we live in, we need definitions of area, slope, and other basic concepts that are less fragile.

2.1 Sets and intervals

In mathematics, a **set** is a collection of objects called **elements**. There are two common ways to describe a set, both enclosed in braces ({}). The first is to list the elements of the set, such as

$$A = \{2, 3, 4, 5, 6\}. \tag{20}$$

The second is to describe the elements of the set, replacing the phrase "such that" with a colon (:). For example,

$$A = \{x : x \text{ is an integer and } 2 \leq x \leq 6\}. \tag{21}$$

Which approach to use is a matter of convenience and clarity. For small sets, a list is often more concise. For large sets, a description is usually easier and sometimes necessary.

By convention, sets are usually denoted by capital letters and elements by lower-case letters. If a set A contains an element a, we write

$$a \in A. \tag{22}$$

If every element of A is also an element of a set B, we say that A is a *subset* of B and write

$$A \subseteq B. \tag{23}$$

Sets A and B are *equal* if $A \subseteq B$ and $B \subseteq A$, which means that they contain exactly the same elements. If $A \subseteq B$ and $A \neq B$, then A is a *proper subset* of B because A does not contain all of the elements of B.

[4] In defense of simple examples: "*. . . we need a dream-world in order to discover the features of the real world we think we inhabit* (and which may actually be just another dream-world)." [Paul Feyerabend, *Against Method* (1975), italics original] In my opinion, mathematics is our most ancient, useful, and beautiful collective dream-world.

Set operations There are three important operations for sets: complement, union, and intersection. If A is a set, its complement A^{C} is the set of all objects that are not elements of A. This is defined with respect to some "universal set" Ω (which may not be specified explicitly), so

$$A^{\mathsf{C}} = \{\omega \in \Omega : \omega \notin A\}, \tag{24}$$

which is logically equivalent to "not A". If A and B are sets, their *union* is

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}, \tag{25}$$

which is logically equivalent to "A or B" where "or" includes the possibility that both A and B occur. Their *intersection* is

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}, \tag{26}$$

which is logically equivalent to "A and B".

Intervals The sets we will use most often are the real numbers (denoted $\mathbb{R}$) and intervals in $\mathbb{R}$. An interval may or may not contain its endpoints. An endpoint with a square bracket is included in the interval; an endpoint with a parenthesis is not included. If $a < b$, there are four possible intervals with endpoints a and b:

$$\begin{array}{ll} [a,b] = \{x \in \mathbb{R} : a \leq x \leq b\} & (a,b] = \{x \in \mathbb{R} : a < x \leq b\} \\ [a,b) = \{x \in \mathbb{R} : a \leq x < b\} & (a,b) = \{x \in \mathbb{R} : a < x < b\} \end{array} \tag{27}$$

An interval is allowed to have one or both endpoints at infinity. Infinite endpoints must have parentheses because $\pm\infty$ are not actually in $\mathbb{R}$. For example, $(-\infty, a]$ denotes all numbers less than or equal to a, (a, ∞) denotes all numbers greater than a, and $(-\infty, \infty) = \mathbb{R}$.

Intervals of the form $[a, b]$ are called **closed intervals**, and intervals of the form (a, b) are called **open intervals**. An open interval containing x is called a **neighborhood** of x.

2.2 Functions

If A and B are sets, a **function** from A to B assigns an element $f(x) \in B$ to each element $x \in A$. A is called the *domain* of f. The input x is called the **argument** of f, and the output $y = f(x)$ is called the **value** of f at x. There are two important details hidden in this picture:

1. For each $x \in A$, $f(x)$ must be exactly one element of B. The domain A is part of the definition of the function, although it may not always be specified explicitly.

2. Elements of B can be associated with any number of elements of A, including zero and infinity.

When we have a function f defined on a set A that takes values in a set B, we say that "f maps A to B" and write

$$f : A \to B. \tag{28}$$

The set of all elements in B that are output from f for at least one $x \in A$ is written $f(A)$. We always have $f(A) \subseteq B$, but often $f(A) \neq B$. For example, the function $f(x) = x^2$ has $f(\mathbb{R}) = [0, \infty)$.

Composition of functions It is often useful to combine functions, using the output of one as the input for another. If A, B, and C are sets such that $f : A \to B$ and $g : B \to C$, then we can use the output of f as input for g. The combined function

$$g \circ f : A \to C \tag{29}$$

is called "g of f" and defined as

$$(g \circ f)(x) = g\big(f(x)\big). \tag{30}$$

For example, let $f(x) = x + 1$ and $g(x) = x^2$. Then $(g \circ f)(x) = (x + 1)^2$. This new function can then be combined with other functions, and so on. This lets us build complex functions out of simple pieces.

Real-valued functions A **real-valued function** is a function whose values are in $\mathbb{R}$. In Section 1, the velocity function $v(t)$ and the position function $x(t)$ are both real-valued functions defined on the interval $[0, 2]$. The derivative $x'(t)$ in equation (18) is a real-valued function defined on a union of three intervals, $[0, 1) \cup (1, 1.5) \cup (1.5, 2]$, because $x'(1)$ and $x'(1.5)$ are undefined. From now on, we will deal only with real-valued functions defined on intervals in $\mathbb{R}$.

2.3 Limits

This is where the real calculus begins. Moving beyond the rectangles and line segments of Section 1 requires us to be precise when we talk about "approaching" a number—either as a function argument or a function value.

A function f has a **limit** ℓ at x if $f(u)$ is guaranteed to be close to ℓ whenever $f(u)$ is defined and u is close enough to x. More precisely: For each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|f(u) - \ell| < \varepsilon \text{ whenever } |u - x| < \delta. \tag{31}$$

In other words: Given any ε, we can find a δ that makes equation (31) true. Note that we only need to consider values of u where $f(u)$ is defined. If the limit of f at x exists and equals ℓ, we write

$$\lim_{u \to x} f(u) = \ell. \tag{32}$$

For a limit, it doesn't matter what happens at x itself. We do not require that $f(x) = \ell$ or even that f is defined at x.

It is important to remember that limits might not exist. A famous example of this is called the *Dirichlet function*. *Rational* numbers are numbers that can be written $^m/_n$ where m and n are integers. The set of all rational numbers is usually denoted $\mathbb{Q}$ (for "quotient"). All other real numbers are called *irrational*. The Dirichlet function is

$$D(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases} \tag{33}$$

Because all intervals contain both rational numbers and irrational numbers, this function has no limit anywhere.

Limits equal to $\pm\infty$ Unlike functions, limits can take values of $\pm\infty$. If for each $Y \in \mathbb{R}$, there exists a $\delta > 0$ such that

$$f(u) > Y \text{ whenever } |u - x| < \delta, \tag{34}$$

then

$$\lim_{u \to x} f(x) = \infty. \tag{35}$$

Here, "Y" was chosen to remind us of the y-axis because the large numbers (positive or negative) occur in the *value* of the function. Limits that equal $-\infty$ are defined similarly except that $f(u) < Y$ whever $|u - x| < \delta$.

Limits at $\pm\infty$ Strictly speaking, ∞ and $-\infty$ are not part of $\mathbb{R}$. We cannot get "close" to them in the same sense that we can get close to a real number. Thus, limits at $\pm\infty$ require a slightly different definition than limits at finite arguments. If for each $\varepsilon > 0$, there exists an $X \in \mathbb{R}$ such that

$$|f(u) - \ell| < \epsilon \text{ whenever } u > X, \tag{36}$$

then we write

$$\lim_{u \to \infty} f(u) = \ell. \tag{37}$$

Limits at $-\infty$ are similar except that we need an X such that $|f(u) - \ell| < \varepsilon$ whenever $u < X$. Here, "X" was chosen to remind us of the x-axis because the large numbers (positive or negative) occur in the *argument* of the function. Limits at $\pm\infty$ can take values of $\pm\infty$ just like limits at finite arguments.

One-sided limits The function f has a **limit from the right** ℓ at x if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|f(u) - \ell| < \varepsilon \text{ whenever } x < u < x + \delta. \tag{38}$$

In this case, we write

$$\lim_{u \to x^+} f(u) = \ell. \tag{39}$$

This is almost the same as the definition leading to equation (37), but we only consider $u > x$ (i.e., to the right of x on a number line). A **limit from the left** is defined similarly except that we consider only $u < x$ (i.e., to the left of x on a number line) and we write

$$\lim_{u \to x^-} f(u) = \ell. \tag{40}$$

Limits from the right and left can be different, and one or both of them might not exist. If $\lim_{u \to x} f(u) = \ell$, then both one-sided limits at x exist and equal ℓ.

2.4 Continuous functions

Functions can be incredibly weird.[5] We need to move beyond rectangles and line segments, but how far will we go? A **continuous** function is a function that changes smoothly, with small changes in input resulting in small changes in output. A function $f : A \to \mathbb{R}$ is **continuous at** x if

$$\lim_{u \to x} f(u) = f(x). \tag{41}$$

The function f is **continuous on** a set A if it is continuous at each $x \in A$. The position function $x(t)$ from equation (14) and Figure 4 is continuous on the interval $[0, 2]$.

A function f is **continuous from the left** if

$$\lim_{u \to x^-} f(u) = f(x), \tag{42}$$

which means that the limit from the left at x equals $f(x)$. It is **continuous from the right** if

$$\lim_{u \to x^+} f(u) = f(x), \tag{43}$$

which means that the limit from the right at x equals $f(x)$. If f is continuous at x, then it is continuous from both sides at x. The velocity function $v(t)$ from equation (10) and Figure 3 is continuous from the left on $[0, 2]$.

A function f on an interval I is **piecewise continuous** if I can be broken into a finite number of subintervals such that f is continuous *inside* each subinterval (i.e., at all points except possibly the endpoints). The velocity function $v(t)$ from equation (10) and Figure 3 is piecewise continuous on $[0, 2]$ because it is continuous inside each of the subintervals $[0, 1]$, $(1, 1.5]$, and $(1.5, 2]$.

From now on, we will deal only with continuous or piecewise continuous real-valued functions defined on intervals in $\mathbb{R}$.

Discontinuities To better understand continuous functions, it is helpful to see what can go wrong. There are three types of discontinuities:

[5] The Dirichlet function in equation (33) is weird, the Cantor function (a.k.a. the Devil's staircase) and Thomae's function (a.k.a. the popcorn function) are weirder, and so on.

- A *removable discontinuity* occurs at x when both one-sided limits at x are finite and equal, but they are not equal to $f(x)$.
- A *jump discontinuity* occurs at x when both one-sided limits at x are finite but unequal.
- An *essential discontinuity* occurs at x when at least one of the one-sided limits is infinite or does not exist. The Dirichlet function in equation (33) has an essential discontinuity at each $x \in \mathbb{R}$ because it has no one-sided limits anywhere.

Figure 6 shows a function with all three types of discontinuity. [6]

Composition of continuous functions When we combine continuous functions, the result is usually another continuous function:

- If f and g are continuous at x, so are $f(x) + g(x)$ and $f(x) \times g(x)$.
- If $g(x) \neq 0$, then ${}^{f(x)}/_{g(x)}$ is also continuous at x.
- If f is continuous at x and g is continuous at $f(x)$, then $(g \circ f)(x) = g\big(f(x)\big)$ is continuous at x.

Although we are dealing only continuous or piecewise continuous functions, this leaves us a great deal of freedom. For almost all applications of calculus in epidemiology, it lets us do everything we need to do.

3 Derivatives

In Section 1, we defined the "derivative" of $x(t)$ to be the slope of the line segment that contained the point $\big(t, x(t)\big)$. Clearly, this will not work for functions whose graphs do not break down into line segments. But when does it make sense to talk about the slope of a graph at a point? If f is function defined on an interval containing x, then the **derivative** of f at x is

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{44}$$

if this limit exists. If $f'(x)$ exists, then f is said to the **differentiable at** x. The function f is **differentiable on** a set A if is differentiable at each $x \in A$.

For a given value of h, the expression on the right is the slope of the line segment connecting $\big(x, f(x)\big)$ and $\big(x+h, f(x+h)\big)$. The limit tells us to calculate the slopes of smaller and smaller line segments—in both positive and negative directions from x. Some simple examples will help make this clear:

[6] What types of discontinuities are there in $v(t)$, $x(t)$, and $x'(t)$ from Section 1? Where do limits exist? Where do one-sided limits exist? Where are they continuous, continuous from the right, or continuous from the left?

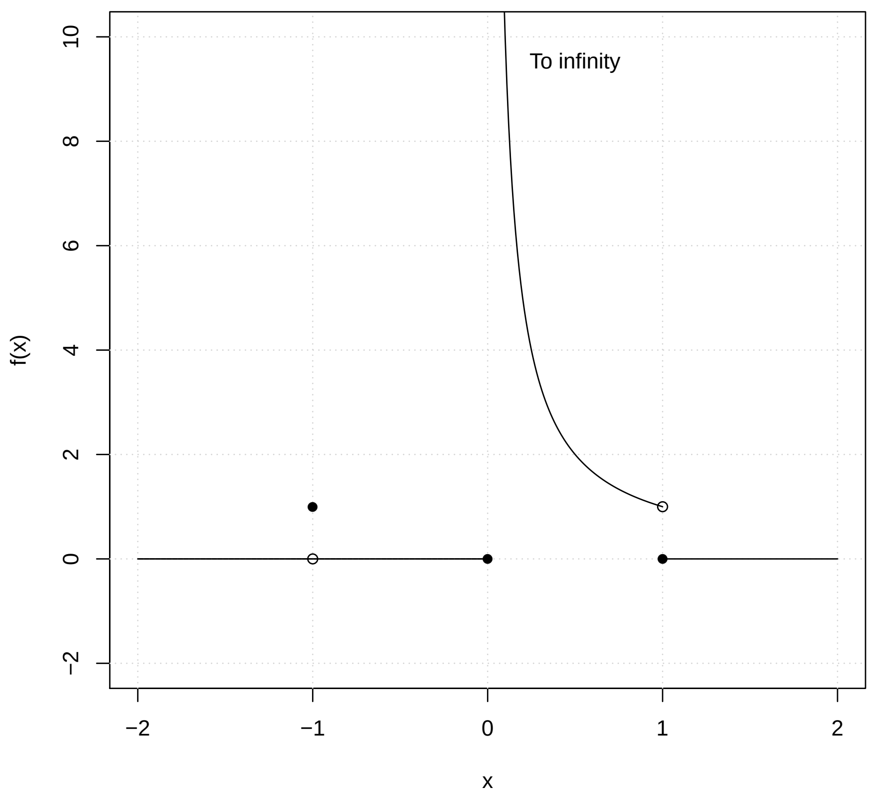

Figure 6: A function with all three types of discontinuities: a removable discontinuity at $x = -1$, an essential discontinuity at $x = 0$, and a jump discontinuity at $x = 1$. The function is continuous from the left at $x = 0$ and continuous from the right at $x = 1$. It has limits from the right and left everywhere, but they are unequal at $x = 0$ and $x = 1$.

- Let $f(x) = c$ for some $c \in \mathbb{R}$. Then
$$f'(x) = \lim_{h \to 0} \frac{c - c}{h} = 0, \tag{45}$$
so the derivative of the constant function $f(x) = c$ is $f'(x) = 0$.
- Let $f(x) = x$. Then
$$f'(x) = \lim_{h \to 0} \frac{(x + h) - x}{h} = \frac{h}{h} = 1, \tag{46}$$
so the derivative of $f(x) = x$ is $f'(x) = 1$.
- Let $f(x) = x^2$. Then
$$f'(x) = \lim_{h \to 0} \frac{(x + h)^2 - x}{h} = \lim_{h \to 0} \frac{2xh + h^2}{h} = \lim_{h \to 0} (2x + h) = 2x, \tag{47}$$
so the derivative of $f(x) = x^2$ is $f'(x) = 2x$.

All three of these examples are differentiable on all of $\mathbb{R}$.

Higher derivatives The derivative of a real-valued function is another real-valued function. If f is a function and its derivative f' is differentiable, then the function
$$f''(x) = \frac{\mathrm{d}}{\mathrm{d}x} f'(x) \tag{48}$$
is called the **second derivative** of f. For example, let $f(x) = x^2$. The first derivative of $f(x) = x^2$ is $f'(x) = 2x$, and the second derivative of f is $f''(x) = 2$. The first two derivatives are generally the most useful, but we can also define third derivatives, fourth derivatives, and so on. The k^{th} derivative of f at x can be written $f^{(k)}(x)$, especially for $k \geq 3$.

3.1 Approximation

If f is a function defined on an interval is differentiable at x, then
$$f(x + h) \approx f(x) + f'(x)h \tag{49}$$
for small values of h (which can be positive or negative). We can rewrite this approximation as a function
$$T_x(u) = f(x) + f'(x)(x - u) \tag{50}$$
called the **first-order Taylor series** approximation to f at x. When higher derivatives exist, they can be used to calculate higher-order Taylor series that are better approximations to f near x than equation (50). Often, these polynomials are much easier to deal with than the original functions. Taylor series allow computers to calculate functions like sine, cosine, logarithms, and exponents to very high accuracy.

The graph of $T_x(u)$ is a line passes through the point $\big(x, f(x)\big)$ and has slope $f'(x)$. This is called the **tangent line** to the graph of f at x. Figure 7 shows several tangent lines to the graph of $f(x) = x^3$.

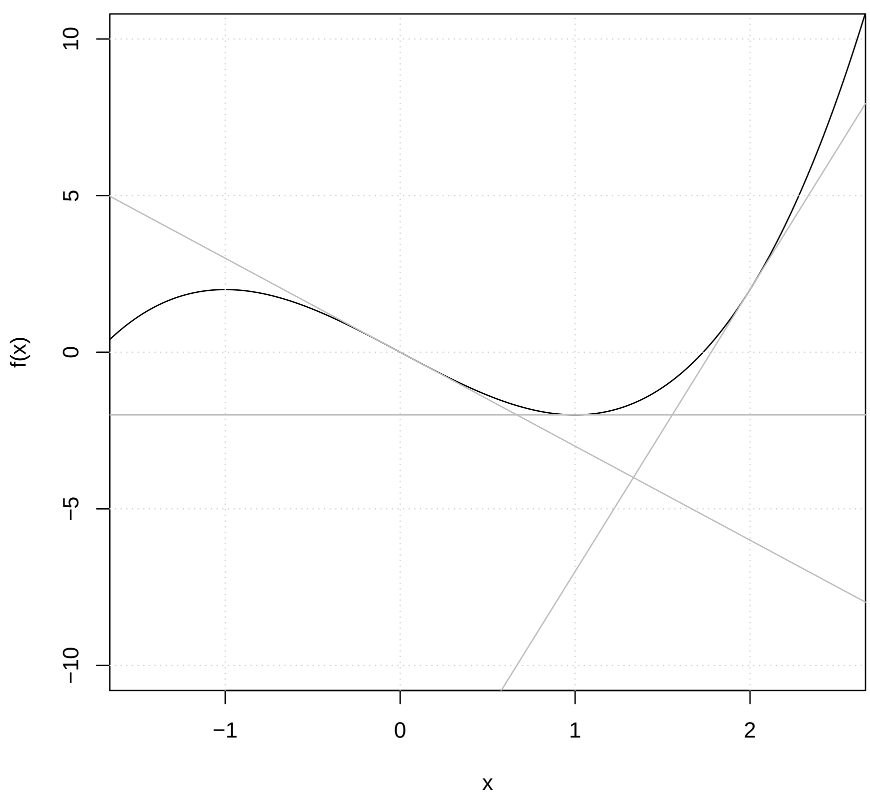

Figure 7: A plot of the function $f(x) = x^3 - 3x$ with tangent lines at $x = 0$, $x = 1$, and $x = 2$. The tangent line at x goes through $\big(x, f(x)\big)$ and has slope $f'(x) = 3x^2 - 3$. The calculation of this derivative is covered in Section 3.4.

3.2 Optimization

A function f has a **local maximum** at x if $f(u) \leq f(x)$ for all u in a neighborhood of x. It has a **local minimum** at x if $f(u) \leq f(x)$ for all u in a neighborhood of x. If f is differentiable, its derivative can be used to find local minima and maxima. An x such that $f'(x) = 0$ or $f'(x)$ is undefined is called a **critical point** of f. Local minima or maxima can only occur at critical points or at the endpoints of the domain of f.

If f has a local minimum or maximum at an x such that $f'(x)$ is defined and x is not an endpoint of the domain of f, then $f'(x) = 0$. To see why this is true, take a another look at the definition of $f'(x)$ in equation (44). If $f'(x) > 0$, then we must have

$$f(x+h) - f(x) > 0 \text{ for small enough } h > 0 \tag{51}$$

or

$$f(x+h) - f(x) < 0 \text{ for small enough } h < 0. \tag{52}$$

In the first case, f cannot be a local maximum because $f(x+h) > f(x)$ for small positive h. In the second case, f cannot be a local minimum because $f(x+h) < f(x)$ for small negative h. Similar arguments show that we cannot have $f'(x) < 0$ if x is a local maximum or minimum. Because we assumed that $f'(x)$ is defined, the only possibility left is $f'(x) = 0$.

It is possible to have a critical point that is not a local maximum or minimum. A simple example of this is $f(x) = x^3$, which has $f'(0) = 0$ even though it is strictly increasing on all of $\mathbb{R}$.

Second derivative test If f has a critical point x inside its domain, the second derivative test can help determine whether it is a local maximum or a local minimum:

- If $f''(x) < 0$, then x is a local maximum. As u increases and we cross x, $f'(u)$ goes from positive (uphill) to negative (downhill).
- If $f''(x) > 0$, then x is a local minimum. As u increases and we cross x, $f'(u)$ goes from negative (downhill) to positive (uphill).
- If $f''(x) = 0$, then x could be a local maximum, a local minimum, neither, or even both (e.g., a constant function). An x where $f''(x) = 0$ is called an **inflection point**.

When the second derivative test is inconclusive, it is usually easiest to look at the function itself to determine what is happening at x. Sometimes, this is easier than doing the second derivative test at all.

3.3 Calculating derivatives*

Now that we know how derivatives can be useful, we will cover some basic rules for calculating them. In Section 2.2, we saw how functions can be combined to make new functions. This turns out to be extremely useful for calculating derivatives. Once we know the derivatives some basic functions, the following rules can be used to calculate the derivatives of functions built up from them.

Adding and muliplying constants Suppose f is differentiable at x and we choose any constant $c \in \mathbb{R}$. Then

- The derivative of $f(x) + c$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x) + c\big) = \lim_{h \to 0} \frac{\big(f(x+h) + c\big) - \big(f(x) + c\big)}{h}$$
$$= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x), \tag{53}$$

  so adding a constant c does not change the derivative.

- The derivative of $cf(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(cf(x)\big) = \lim_{h \to 0} \frac{cf(x+h) - cf(x)}{h}$$
$$= c \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = cf'(x), \tag{54}$$

  so multiplying by a constant c multiplies the derivative by c.

In the first case, adding c moves the graph of the function up or down but the slope stays the same. In the second case, multiplying by c changes the rise (vertical distance) of each line segment but leaves the run (horizontal distance) unchanged, so the slope is muliplied by c.

Adding and muliplying functions Now suppose f and g are functions that are both differentiable at x. Then

- The derivative of $f(x) + g(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x) + g(x)\big) = \lim_{h \to 0} \frac{\big(f(x+h) + g(x+h)\big) - \big(f(x) + g(x)\big)}{h}$$
$$= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}$$
$$= f'(x) + g'(x), \tag{55}$$

  so the derivative of a sum is the sum of the derivatives.

- The derivative of $f(x)g(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x)g(x)\big) = f'(x)g(x) + f(x)g'(x). \tag{56}$$

  To see why, imagine a rectangle with sides $f(x)$ and $g(x)$.[7]

Note that the rules for adding and muliplying constants in equations (53) and (54) are special cases of these rules if we let $g(x) = c$.

[7] Can you draw a picture of what happens when x changes to $x + h$ for a small h?

The chain rule If f and g are functions such that $g(f(x))$ is defined, f is differentiable at x, and g is differentiable at $f(x)$, then the chain rule says that

$$\frac{\mathrm{d}}{\mathrm{d}x} g\big(f(x)\big) = g'\big(f(x)\big) f'(x). \tag{57}$$

This rule is more mysterious than the others. The easiest way to see why it works is to use an approximation like that in equation (49):

$$\begin{aligned} g\big(f(x+h)\big) &\approx g\big(f(x) + hf'(x)\big) \\ &\approx g\big(f(x)\big) + [hf'(x)]g'\big(f(x)\big). \end{aligned} \tag{58}$$

In the second line, $hf'(x)$ is used just like the h in equation (49). Rearranging, we get

$$\frac{g\big(f(x+h)\big) - g\big(f(x)\big)}{h} \approx g'\big(f(x)\big) f'(x) \tag{59}$$

for small h. This is an explanation, not a proof.

3.4 Useful derivatives*

To calculate derivatives, we first use the definition in equation (44) to figure out the derivatives of some basic functions. We then use the rules of Section 3.3 to calculate the derivatives of functions built out of them. The set of functions you can differentiate will grow like a tree from a tiny seed.

Positive integer powers of x In equation (46), we saw that the derivative of $f(x) = x$ is $f'(x) = 1$. If we let $k = 1$, then we have

$$\frac{\mathrm{d}}{\mathrm{d}x} x^k = kx^{k-1} \tag{60}$$

because $x^0 = 1$. This is our *base case.* Now suppose that $k \geq 1$ and equation (60) above is true. Using the rule from equation (56) for differentiating a product of functions, we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} x^{k+1} &= \frac{\mathrm{d}}{\mathrm{d}x} \big(x \times x^k\big) \\ &= 1x^k + x \frac{\mathrm{d}}{\mathrm{d}x} x^k \\ &= x^k + x \times kx^{k-1} \qquad (61) \\ &= (k+1)x^k. \qquad (62) \end{aligned}$$

We have shown that if equation (60) is true for k, it is also true for $k+1$. This is called the *induction step.* Since the base case of $k = 1$ was true, we have shown that equation (60) holds for all x to the power 1, 2, or any other positive integer k. This is an example of a **proof by mathematical induction**.

Using this result plus the fact that the derivative of a constant is zero, we can now take the derivative of any polynomial.

Other integer powers of x When $x \neq 0$, we define $x^{-k} = {}^1\!/_{x^k}$. The result we just proved about positive integer powers of x also tells us how to differentiate negative integer powers. When k is any positive integer,

$$x^k x^{-k} = 1. \tag{63}$$

If we take derivatives on both sides, we get

$$k x^{k-1} x^{-k} + x^k \frac{\mathrm{d}}{\mathrm{d}x} x^{-k} = 0. \tag{64}$$

Simplifying the left-hand side gives us

$$\frac{k}{x} + x^k \frac{\mathrm{d}}{\mathrm{d}x} x^{-k} = 0. \tag{65}$$

Solving this, we get

$$\frac{\mathrm{d}}{\mathrm{d}k} x^{-k} = (-k) x^{(-k)-1}. \tag{66}$$

This is the same pattern as we had in equation (60) for positive integer powers of x. Because $x^0 = 1$,

$$\frac{\mathrm{d}}{\mathrm{d}x} x^0 = 0 = 0 \times x^{-1}. \tag{67}$$

when $x \neq 0$. Thus, the pattern holds for all integers k when $x \neq 0$.

Quotients of functions Equation (56) gave us a rule for differentiating products of functions, but it is also useful to differentiate quotients of functions.[8] We just proved that $\frac{\mathrm{d}}{\mathrm{d}x} x^{-1} = -x^{-2}$. Using the chain rule in equation (57), we get that

$$\frac{\mathrm{d}}{\mathrm{d}x} \frac{1}{f(x)} = -\frac{f'(x)}{f(x)^2}. \tag{68}$$

Now we can use the product rule in equation (56) to get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} \frac{f(x)}{g(x)} &= \frac{\mathrm{d}}{\mathrm{d}x} \left(f(x) \times \frac{1}{g(x)} \right) \\ &= \frac{f'(x)}{g(x)} + \frac{f(x) g'(x)}{g(x)^2} \\ &= \frac{f'(x) g(x) + f(x) g'(x)}{g(x)^2}. \end{aligned} \tag{69}$$

This is usually painful to remember and calculate. It is more convenient to treat it as a special case of the product rule.

[8] Remember that the quotient ${}^{f(x)}\!/_{g(x)}$ is defined only when $g(x) \neq 0$.

Exponents and logarithms Infinite series must be handled with care, but the ones we will talk about in this section are behave like finite sums and polynomials. *Euler's number* is

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots \approx 2.718281828459\ldots \tag{70}$$

where "k!" denote k *factorial*. The **exponential function** is defined by the infinite series

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots . \tag{71}$$

For each $x \in \mathbb{R}$, the denominator $k!$ grows much faster than the numerator x^k, so the value of the sum is determined mostly by the first few terms. The following properties of e^x are useful:

- For all $x \in \mathbb{R}$, $e^x > 0$.
- By definition, $e^0 = 1$ and $e^1 = e$.
- For $x > 0$, $e^x < 1$. For $x > 0$, $e^x > 1$.

Using the derivative of x^k, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}e^x &= \frac{1}{1!} + 2\frac{1}{2!}x + 3\frac{1}{3!}x^2 + 4\frac{1}{4!}x^3 + \cdots \\ &= 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots \\ &= e^x. \end{aligned} \tag{72}$$

Thus, e^x is its own derivative!! It is the only function with this property. Because $e^x > 0$ for all x, it is strictly increasing.

The **natural logarithm** of x is the logarithm of x to base e. It is often denoted $\ln x$. It is defined for all $x > 0$, and for all such x we have

$$x = e^{\ln x}. \tag{73}$$

For example, $\ln 1 = 0$ and $\ln e = 1$. Differentiating both sides of equation (73), we get

$$1 = e^{\ln x}\frac{\mathrm{d}}{\mathrm{d}x}\ln x = x\frac{\mathrm{d}}{\mathrm{d}x}\ln x. \tag{74}$$

Rearranging gives us

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x}. \tag{75}$$

Therefore, $\ln x$ is differentiable at all $x > 0$. Equation (75) also tells us that $\ln x$ is strictly increasing because $x^{-1} > 0$ for all $x > 0$, Note that e^x grows very quickly and $\ln x$ grows very slowly.

Combining the exponential and natural logarithm functions lets us take derivatives with respect to a positive base for any real exponent. Let $r \in \mathbb{R}$ be an exponent and $x > 0$ be a positive real base. Because

$$x^r = e^{r \ln x}, \tag{76}$$

its derivative is

$$\frac{\mathrm{d}}{\mathrm{d}x}x^r = e^{r\ln x}\frac{r}{x} = rx^{r-1}. \tag{77}$$

This is the same pattern from equation (60) that we proved for integer exponents $k \neq 0$ and any real base x, but now we have proven it for all real exponents when the base $x > 0$.

We can also differentiate with respect to exponents when the base is positive. Let $b > 0$ be a positive real base and x be a real exponent. Because

$$b^x = e^{x\ln b}, \tag{78}$$

its derivative is

$$\frac{\mathrm{d}}{\mathrm{d}x}b^x = e^{x\ln b}\ln b = b^x \ln b. \tag{79}$$

When $b = e$, this reduces to equation (72) because $\ln e = 1$.

4 Integrals

Imagine that our car from Section 1 rolls down a very long, straight hill such that

$$v(t) = 10t \text{ for } t \in [0, 2]. \tag{80}$$

The graph of $v(t)$ is shown in Figure 8. It is a triangle with a base of 2 hours and a height of 20 miles per hour, so basic geometry tells us that the area under the curve is

$$\frac{1}{2} \times 2 \text{ hours} \times 20\ \frac{\text{miles}}{\text{hour}} = 20 \text{ miles}. \tag{81}$$

If the connection between area under $v(t)$ and position from Section 1 still holds, the car should be at position $x = 20$ miles at $t = 2$ hours.

4.1 Approximation by rectangles

But how do we know that this connection between area and position still works? We have only really showed that the position of the car and the area under $v(t)$ are the same when the velocity of the car is piecewise constant. However, we could *approximate* the area under $v(t)$ with rectangles and calculate the area in the rectangles. For example, let's start with one rectangle for each half hour. If we use the speed at the beginning of each interval (the lowest speed during that half hour), our rectangles will stay below the graph of $v(t)$. We will get

$$\frac{1}{2} \text{ hour} \times (0 + 5 + 10 + 15)\ \frac{\text{miles}}{\text{hour}} = 15 \text{ miles}. \tag{82}$$

This would be the position of a car that traveled 0 miles per hour in the first half hour and then 5, 10, and 15 miles per hour in the next three half hours. Our car was almost always traveling faster than this, so it must have gone further

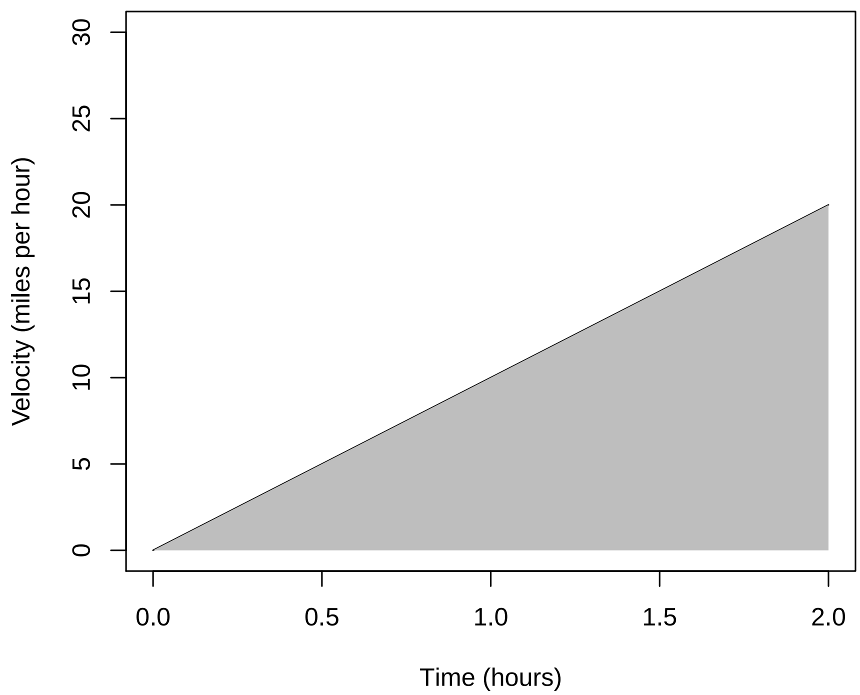

Figure 8: Graph of $v(t)$ when the car is constantly accelerating. The area under the graph makes a triangle.

down the x-axis. If we use the speed at the end of each interval (the highest speed during that half hour), we get

$$\frac{1}{2}\text{ hour} \times (5 + 10 + 15 + 20)\,\frac{\text{miles}}{\text{hour}} = 25\text{ miles.} \tag{83}$$

This is the position of a car that went 5 miles per hour in the first half hour and then 10, 15, and 20 miles per hour in the next three half hours. We know the car did not travel this far, because this is the position of a car with a velocity that is almost always greater than the actual car. Our hypothesis that the car is at $x = 20$ miles is within these bounds, which is encouraging. Figure 10 shows the velocity function of the car with the rectangles giving upper and lower bounds on the area.

Smaller rectangles should give us a better approximation. If we repeat the process from the last paragraph with ten rectangles instead of four, then each rectangle has a base of 0.2 hours (12 minutes). The lower bound is

$$\frac{1}{5}\text{ hours} \times (0 + 2 + 4 + 6 + 8 + 10 + 12 + 14 + 16 + 18)\,\frac{\text{miles}}{\text{hour}} = 18\text{ miles,} \tag{84}$$

and the upper bound is

$$\frac{1}{5}\text{ hours} \times (2 + 4 + 6 + 8 + 10 + 12 + 14 + 16 + 18 + 20)\,\frac{\text{miles}}{\text{hour}} = 22\text{ miles.} \tag{85}$$

These bounds differ by four miles instead of ten, and they still contain 20 miles. However, this process gets cumbersome quickly as our rectangles get smaller.

Integrable functions If we can make the upper and lower bounds as close to each other as we want by using small enough rectangles to approximate the area below the graph of v, then v is said to be **integrable** on the interval $[0, 2]$. All continuous and piecewise continuous functions are integrable.

The Dirichlet function in equation (33) is not integrable. In any interval, its minimum value is zero and its maximum value is one. If we try to calculate the area under $D(x)$ for $x \in [a, b]$, the lower bound is zero and the upper bound is $b - a$ no matter how small the rectangles are. When a function is not integrable, there is a sense in which the area under its graph doesn't even exist.[9]

4.2 Differentiating an integral

In Section 3.1, we learned how the derivative of a function f at x could be used to approximate f in a small neighborhood of x. When h is small,

$$x(t + h) \approx x(t) + hx'(t). \tag{86}$$

[9] "Whereof one cannot speak, thereof one must be silent." [Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (1918)] The notion of integrability here is *Riemann integrability*. The Dirichlet function is *Lebesgue integrable*, and its integral over any interval is zero. The Lebesgue integral exists and equals the Riemann integral whenever the latter exists. A function that is not Lebesgue integrable doesn't have an area under its graph that we can talk about sensibly.

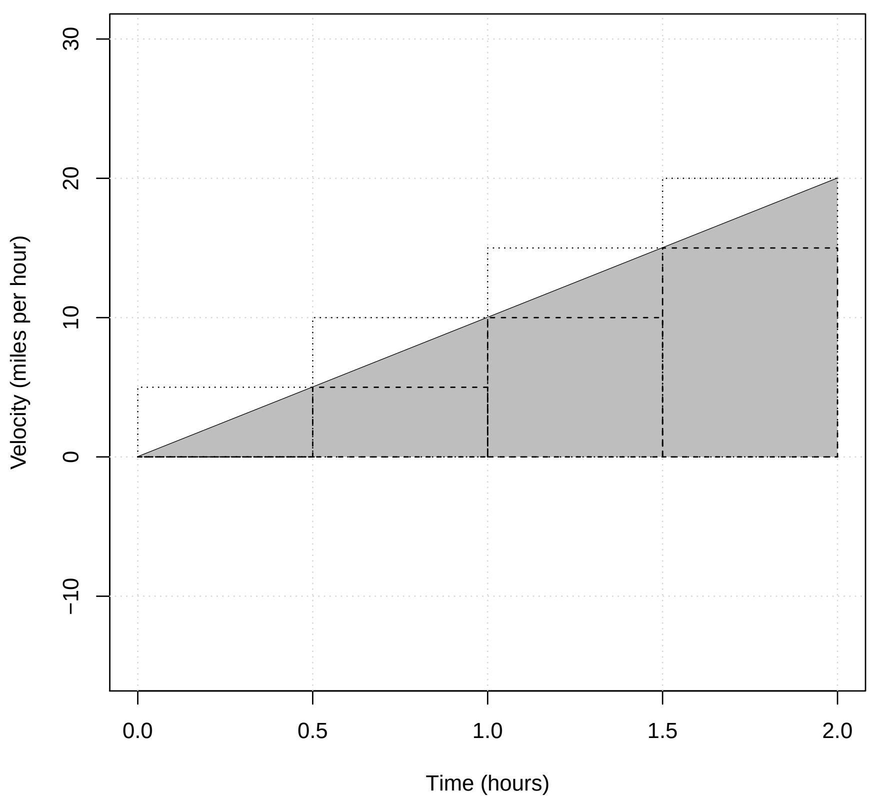

Figure 9: Graph of $v(t)$ when the car is rolling down a hill. Rectangles where the car moves at constant velocity for half-hour time intervals give upper and lower bounds on the area under $v(t)$.

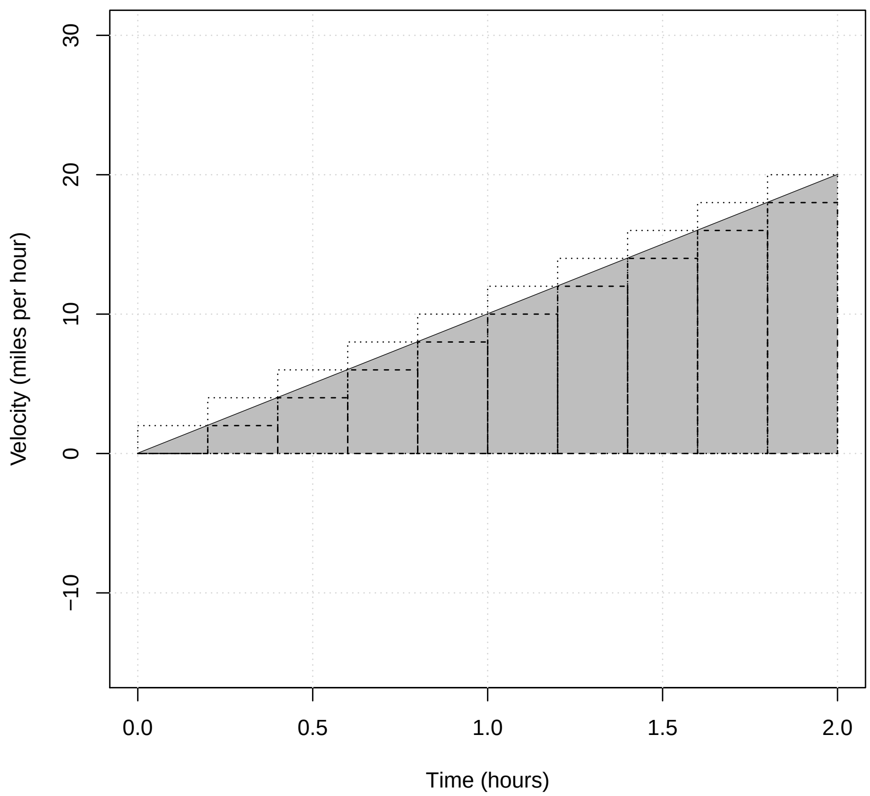

Figure 10: Graph of $v(t)$ when the car is rolling down a hill. Rectangles where the car moves at constant velocity over 12-minute intervals give upper and lower bounds on the area under $v(t)$. This approximation is more accurate than the one in Figure 10.

Using this approximation to estimate $x'(t)$, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x(t) &= \lim_{h\to 0}\frac{x(t+h)-x(t)}{h}\\ &\approx \lim_{h\to 0}\frac{v(t)h}{h} = v(t).\end{aligned} \tag{87}$$

When $v(t)$ is continuous at t, this equality is exact.[10] A useful way to think of this result is to imagine a rectangle with a base on the interval $[t, t+h]$ at a t where $v(t)$ is continuous. The area under the graph of $v(t)$ between t and $t+h$ is approximately $v(t)h$ when h is small.[11] Even though we don't know how to calculate $x(t)$, we know that $x'(t)$ equals $v(t)$ whenever v is continuous at t.

The velocity of our car at time t is $v(t) = 10t$, which is continuous at all t. Using what we learned about derivatives in Section 3, let's guess that

$$x(t) = 5t^2 + C \tag{88}$$

where C is some constant. Then

$$x'(t) = 10t = v(t), \tag{89}$$

so the velocity function is correct. To figure out what C is, we use the fact that the car started at $x = 0$, so

$$x(0) = C = 0. \tag{90}$$

Thus, $x(t) = 5t^2$ for $t \in [0, 2]$.

The first fundamental theorem of calculus The result in equation (87) is a special case of the first fundamental theorem of calculus: If $g : [a, b] \to \mathbb{R}$ is integrable and

$$G(x) = \int_a^x g(u)\,\mathrm{d}u, \tag{91}$$

then G is continuous on $[a, b]$ and $G'(x) = g(x)$ at all x where g is continuous.

Position and velocity again In our car example, the first fundamental theorem of calculus lets us calculate the position of the car each each $t \in [a, b]$ whenever:

1. We know where the car starts at time $t = a$.
2. The velocity function $v(t)$ is integrable on $[a, b]$.

The position of the car is

$$x(t) = x(a) + \int_a^t v(u)\,\mathrm{d}u \tag{92}$$

[10] Given any $\varepsilon > 0$, we can find an h small enough to ensure that $v(u) \in (v(t)-\varepsilon, v(t)+\varepsilon)$ whenever $u \in (t-h, t+h)$.

[11] Drawing a picture of this will help you understand it better.

at each $t \in [a, b]$. As in Section 1, $x(t)$ is continuous and $x'(t) = v(t)$ for all t where v is continuous. However, we have taken a huge step forward because we now know that this works for any velocity function that is integrable.

We also get an additional result: Given a position function $x(t)$, then we must have $v(t) = x'(t)$ wherever $x'(t)$ exists. Therefore:

- We can calculate the position from the velocity by *integrating* whenever the velocity function is integrable.
- We can calculate the velocity from the position by *differentiating* whenever the position function is differentiable.

A similar relationship exists between velocity and acceleration, which is the *first derivative* of the velocity and the *second derivative* of the position function.

4.3 Integrating a derivative*

Calculating a derivative is usually a straighforward process of applying the rules of Section 3.3 to break down a function until we know the derivative of each part and then putting it back together. Calculating an integral, which is called **integration**, means working backwards to find a function has a derivative equal to the integrand. Integration usually requires more skill and imagination than differentiation. This is why calculus classes almost always teach derivatives first even though integrals came first historically.[12]

Antiderivatives The antiderivative of a function f on an interval $[a, b]$ is a function F such that $F'(x) = f(x)$ at each $x \in [a, b]$. The antiderivative of a function $f(x)$ is often written as an integral without limits of integration, and a **constant of integration** (usually "C") is added because this does not change the derivative. For example:

- Powers of x have the antiderivative

$$\int x^k \,\mathrm{d}x = \frac{1}{k+1} x^{k+1} + C. \tag{93}$$

  This works when the corresponding derivative of x^{k+1} is correct: positive integer powers of all x, integer powers of $x \neq 0$, and real powers of $x > 0$.

- The exponential function has the antiderivative

$$\int e^x \,\mathrm{d}x = e^x + C. \tag{94}$$

For more complex functions, there are a number of integration techniques used to find antiderivatives. The most important are *integration by substitution* (based on the chain rule) and *integration by parts* (based on the product rule).

[12] In the 3^{rd} century BC, Archimedes calculated the area enclosed by a parabola and a line using an approximation by triangles. The idea of the derivative did not fully appear until the 17^{th} century AD. See `https://en.wikipedia.org/wiki/The_Quadrature_of_the_Parabola`.

The second fundamental theorem of calculus The first fundamental theorem of calculus was about the derivative of an integral. The second is about the integral of a derivative. If g is an integrable function on $[a, b]$ and G is an antiderivative of g on $[a, b]$, then

$$\int_a^b g(x)\,\mathrm{d}x = G(b) - G(a). \tag{95}$$

In other words the integral of g over $[a, b]$ equals the change in its antiderivative from a to b. Note that the constant of integration C cancels out in the subtraction on the right-hand side of equation (95).

Position and velocity one more time If we are given an integrable velocity function $v(t)$ and we find an antiderivative $x(t)$, then the change in the position of the car from a to b is

$$\int_a^b v(t)\,\mathrm{d}t = x(b) - x(a). \tag{96}$$

This equation is valid no matter where the car starts at time a.

The second fundamental theorem is useful for functions that are defined piecewise. For example, suppose we can find an antiderivative x_1 for the interval $[a, b]$ and an antiderivative x_2 for the interval $[b, c]$. Then

$$\begin{aligned}\int_a^c v(t)\,\mathrm{d}t &= \int_a^b v(t) + \int_b^c v(t)\,\mathrm{d}t \\ &= [x_1(b) - x_1(a)] + [x_2(c) - x_2(b)]. \end{aligned} \tag{97}$$

To get the area under the curve on $[a, c]$, we just add the areas under the curve on the subintervals $[a, b]$ and $[b, c]$. We can use the second fundamental theorem separately on each subinterval for as many pieces as we need.

Further reading

If you would like to learn more about calculus, the following two books are highly recommended:

- *The Cartoon Guide to Calculus* by Larry Gonick (HarperCollins, 2012) is a great book for beginners. It explains a lot of ideas with pictures, honestly points out areas where deep mathematics is being skipped, and has good examples and problems.

- *Understanding Analysis, Second Edition* by Stephen Abbott (Springer, 2015) is a much more advanced book. It covers real analysis, which is the mathematical theory that underlies and generalizes calculus. It goes over the details that *The Cartoon Guide to Calculus* skips over, showing where our intuitions fail and how we avoid getting lost.

Both are relatively inexpensive, and *The Cartoon Guide to Calculus* covers almost all the calculus you could possibly need to use as an epidemiologist.