Integration

UIUC Math Camp 2020

Miles D. Williams

Some useful resources:

- See this great series on integrals by the site *Math is Fun*. I generally go to this site any time I have a question about calculus, linear or matrix algebra, etc. It's an accessible resource that's not overly jargon-y.
- Did you know that you don't have to do calculus by hand? Wolfram Alpha is going to be your *new best friend*.

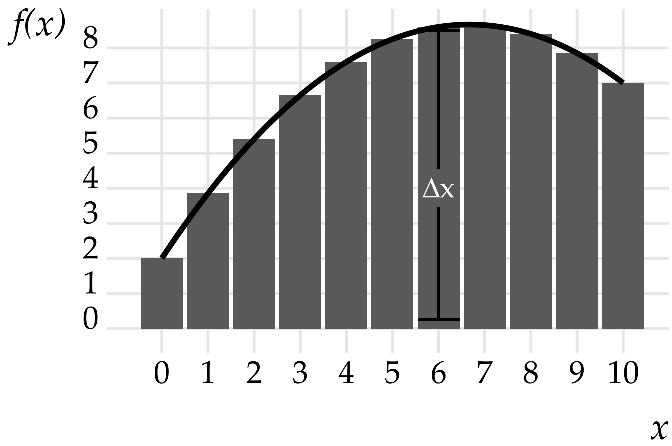
What is an integral?

In this session we will cover *integration*. Integration is a useful tool for a host of reasons, and as a political scientist, you'll find yourself reading more than a few peer reviewed journal articles that use integration—perhaps in analyzing a formal mathematical model of citizen, political leader, or state behavior—and you even may apply integration in your own research. And I should add that integration is *integral* (ha! a pun!) to probability theory. Without the ability to integrate functions, we wouldn't have things like cumulative probability functions to summarize the cumulative probability of an event over a range of values of some random variable.

But perhaps a more common application of integration that you may have encountered prior to now is in *finding the area under the curve* of some function.

Consider, for example, a quadratic function $f : x \to y$. Say we would like to know the area under this function's curve.

One strategy to find the area would be to sum over the areas of individual rectangular "slices" under the curve along x:



We define the area of a given slice as $\Delta x = 1 \times f(x)$ where 1 is the width of a given slice, and where $f(x)$ is the value of the function at a given value along x.

In this particular example, denote the set of values along x that we're going to use as x' where $x' = \{0, 1, 2, ..., 9, 10\} \in \mathbb{Z}^{0+}$ (that is, x' is a vector of non-negative integers ranging from 0 to 10).

Further denote $\Delta x'_i$ as the area of a given slice indexed by $i = 1, 2, ..., |x'|$ where $\Delta x'_i = f(x'_i)$ and where $|x'|$ tells us the number of elements or values of x'_i in the vector x'.

Since in this example a slice has width of 1, $\Delta x'_i$ is simply equal to the value of $f(x'_i)$. We thus now have the vector $\Delta x' = \{\Delta x'_1, \Delta x'_2, ..., \Delta x'_{|x'|}\}$, where for each element of the vector $\Delta x'_i = f(x'_i)$. We can then approximate the area under the curve as

$$\text{Area} \approx \sum_{i=1}^{|\Delta x'|} \Delta x'_i.$$

Though this approach is simple enough, it has one glaring weakness: *it won't give us exactly the right answer for the area under the curve.* We could try to use narrower or more finely grained slices to get an answer closer to the truth, but these again would fail to tell us the true area under the curve.

It's only as $\Delta x \to 0$ that the answer gets closer to the truth. Integration is a convenient and elegant way to identify what this correct answer is at the limit as Δx becomes infinitesimally small.

Another definition for integration: It's the reverse of a derivative

You might also know of another convenient definition of integration: *it is simply the reverse of finding the derivative of a function*.

Say we have the function

$$f(x) = x^2.$$

Using the *power rule*, we know that the derivative of this function is

$$\frac{df(x)}{dx} = 2x.$$

Suppose that we then define the function $h(x) = df(x)/dx$. What would be the integral of $h(x)$?

It turns out that the integral of $h(x)$ is $x^2 + C$. This is just our original function, plus C.

The C, here, represents the "constant of integration." This is just a placeholder for any potential constant (including zero, as was the case with our original function). We include this because the derivative $df(x)/dx = 2x$ in principal could be the derivative for any number of possible functions:

$$\begin{aligned} f(x) &= x^2? \\ f(x) &= x^2 + 1? \\ f(x) &= x^2 - 999? \end{aligned}$$

Using C just helps to keep us honest about the fact that any time we integrate a function, said function could include any addition by $C \in \mathbb{R}$ as well.

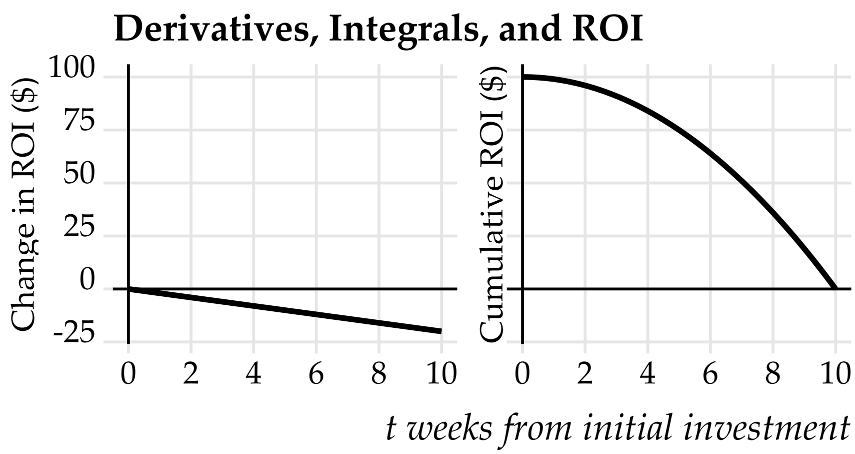
Some more intuition

Just like how derivatives tell us something about a particular function—the first derivative tells us the slope of a function at a given point, for example—the integral of a function also provides useful information.

For example, say that the function $k(t) = -2t$ tells us the rate at which private investment in a mid-sized paper supply firm in northeast Pennsylvania yields returns on investment (ROI) at week t after the initial investment. The integral of $k(t)$, in turn, will tell us about the cumulative ROI at month t:

$$\underbrace{\overbrace{-2t}^{\text{ROI rate at } t}}_{\color{red}{\text{Derivative}}} \longleftrightarrow \overbrace{\underbrace{C - t^2}_{\text{Total ROI at } t}}^{\color{blue}{\text{Integral}}}$$

Plotting these out below, we see on the left the rate of change in ROI communicated by $k(t)$, while on the right we see the cumulative ROI at a given t. In this example, I assume $C = 100$, denoting an initial investment of 100 dollars.



While $k(t)$ tells us how quickly ROI changes at a given point in time, its integral tells us cumulatively how much money we'll have made by a given point in time—or in this case, *how much money we'd use*.

Notation

Just like how derivatives have a special notation, so do integrals. For instance, we represent the derivative of a function by writing

$$\frac{df(x)}{dx}.$$

Similarly, when we want to represent the integral of a function, we write

$$\int f(x)dx.$$

The above notation has some interesting features, so let's break them down really quickly:

$$\overbrace{\int}^{\text{integral symbol}} \qquad \underbrace{f(x)}_{\text{the function to integrate}} \qquad \overbrace{dx}^{\text{slices over x}} \quad .$$

Think of the $\int$ symbol like a fancy version of $\sum$, which is used to denote the *sum* over some set of values—like in our example with taking the sum over $\Delta x_i'$. But, while $\sum$ is used when there is a *finite* set of values to take the sum of, $\int$ is used when there is an *infinite* set of values.

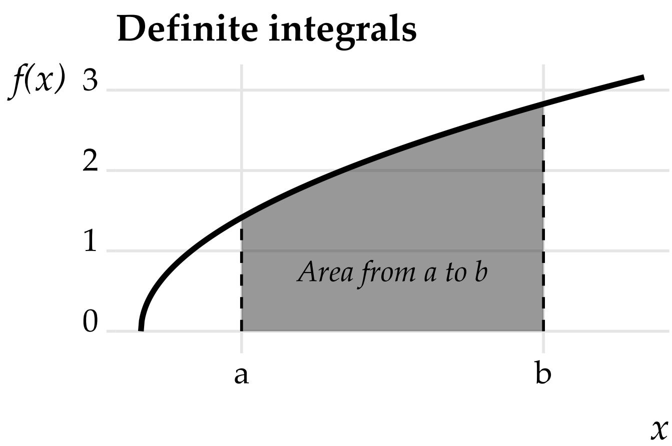
The notation dx signifies that as we're integrating, we're taking the sum over an infinite set of values. While Δx denotes a slice with width greater than zero, dx means that the slices are approaching zero in width.

Finally, the function to be integrated, $f(x)$, is sandwiched between the integral symbol and the notation for slices over x.

Note that you could also use $g(x)$ or $g(y)$, or any other notation for representing a function and its inputs. $\int g(z)dz$ is just as appropriate as $\int f(x)dx$—just make sure that you're consistent!

Definite and indefinite integrals

When we use the integral symbol, we will sometimes use $\int_a^b$. Using this notation clarifies that we're dealing with a *definite* integral, as opposed to an *indefinite* integral. Just like the notation $\sum_{i=1}^n$ clarifies that we are summing over some finite set of values starting with the first value up to value n, $\int_a^b$ clarifies that we're interested in getting the area under a curve for values $a < x \leq b$.



When we use definite integrals to find the area under a defined range of values for a curve, it's helpful to know that we can ignore the constant of integration (C). This is because when we use $\int_a^b$, what we mean is that we want to take the integral of some function where $x = b$, and then subtract the integral of the function where $x = a$. When we do this, the Cs will cancel out. Let me show you what I mean.

Say we have $g(x) = 2x^3$ and we want the definite integral from 2 to 7. We know that the indefinite integral of this function is

$$\int g(x)dx = \frac{x^4}{2} + C.$$

Meanwhile, the definite integral is given as

$$\int_2^7 g(x)dx = \left(\frac{7^4}{2} + C\right) - \left(\frac{2^4}{2} + C\right),$$

which simplifies to

$$\frac{7^4}{2} - \frac{2^4}{2}.$$

Notice that the Cs cancel out. This leaves us with the following *definite* answer for the area under the curve of $g(x)$ from $2 < x \leq 7$:

$$\int_2^7 g(x)dx = 1,192.5.$$

To be a little more general, let $h(x) + C$ denote the indefinite integral of $g(x)$. The definite integral is then given as

$$\int_a^b g(x)dx = h(b) - h(a).$$

Another way that we can write the definite integral is with square brackets like so

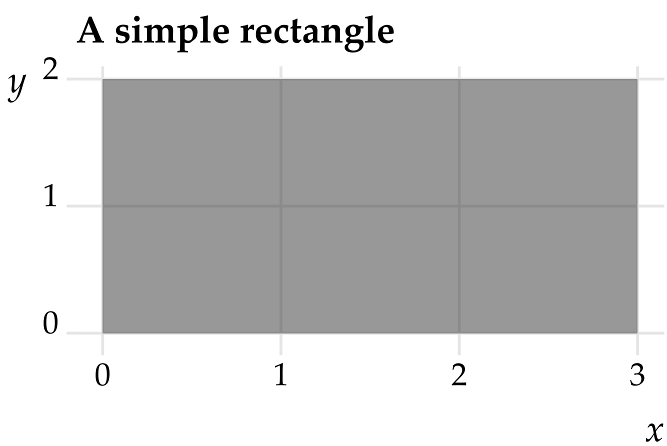
$$\int_a^b g(x)dx = [h(x)]_a^b.$$

Inside the square brackets we have the indefinite integral without C, and along the right square bracket we subscript the starting value (a) and superscript the end value (b).

Finding the area of a rectangle

Some of the above is pretty abstract, so let's apply integration to a really simple problem: *finding the area of a rectangle.*

Consider the below 2×3 rectangle:



If we wanted to find the area of this rectangle, we could take a number of approaches.

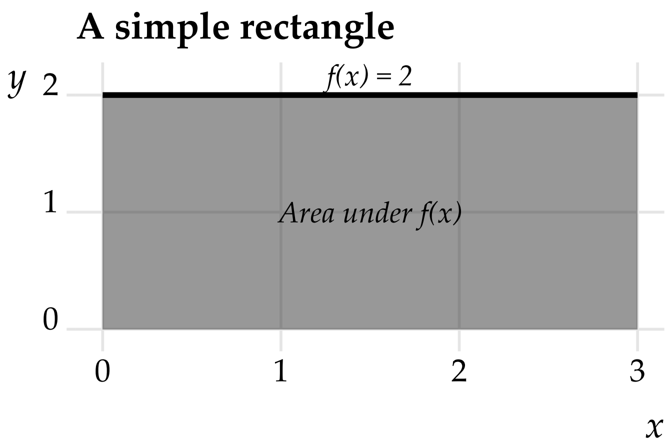
One straightforward strategy would be to just add up the number of 1×1 squares or pieces that make up the rectangle. There are six of these squares in total, so that means the area of the rectangle is six—simple enough.

We also could multiply the rectangle's length and width. The length (or height) is 2, and the width is 3. $2 \times 3 = 6$, so, again, we know the area is six.

However, we can also use integration to find the rectangle's area. Say we treated the top-most edge of the rectangle as the line of a function that maps values of x to y. Such a function for the top line of the rectangle would look like this:

$$f(x) = y = 2.$$

Plotting out this function gets us the following figure:



To find the area of this rectangle (the area under the curve of $f(x)$), we can just take the definite integral of $f(x)$.

Using the power rule, we know that the indefinite integral of $f(x)$ is

$$\int f(x)dx = \int 2dx = 2x + C.$$

When then have the following definite integral

$$\int_0^3 2dx = [2x]_0^3 = 2 \cdot 3 - 2 \cdot 0,$$

which has the solution

$$\int_0^3 2dx = 6.$$

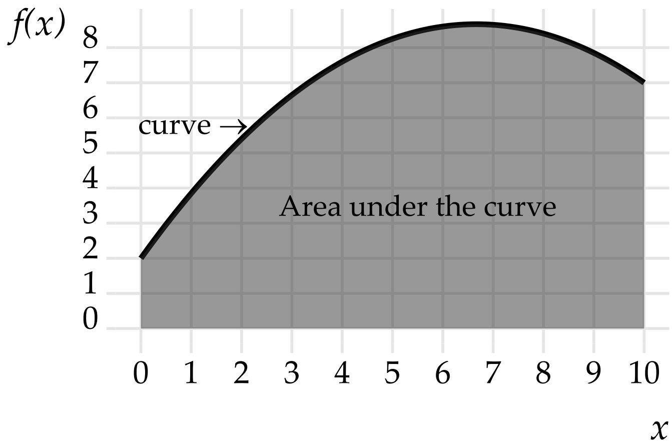
This is same answer as before!

A More Complicated Example

For a problem like finding the area of a rectangle, or the area under a line with slope $= 0$, using integration seems a little excessive.

However, integration is especially useful when we're dealing with irregularly shaped functions (or at least shapes with curves instead of sharp corners and all straight lines[1]).

Consider the quadratic equation used in the intro for this document. The curve for this function, recall, looks like this:



The particular function that gives rise to this figure is given as

$$f(x) = 2 + 2x - \frac{3}{20}x^2.$$

Say we want to know the area under $f(x)$ from $x = 0$ to $x = 10$. We first find the integral of $f(x)$:

$$\int f(x)dx = 2x + x^2 - \frac{1}{20}x^3 + C.$$

The definite integral then is given as

$$\int_0^{10} f(x)dx = \left[2x + x^2 - \frac{1}{20}x^3\right]_0^{10} = 20 + 100 - 50 = 70.$$

The area under the curve is 70!

How does this solution compare to the addition of slices approach highlighted at the beginning of this document?

Recall that we defined $\Delta x'$ as a vector of areas for particular slices under the curve of $f(x)$ along x. x' was a vector of non-negative integers ranging from 0 to 10, and $\Delta x'_i = f(x'_i)$ is the area for the i^{th} slice.

[1]I had a calculus teacher who hated the phrase "straight line."

Using this approach, the area under the curve is approximated as

$$\text{Area} \approx \sum_{i=1}^{|\Delta x'|} \Delta x'_i = \sum_{i=1}^{|x'|} 2 + 2x'_i + \frac{3x'^2_i}{20}.$$

I won't make you suffer through the process of calculating each $\Delta x'_i$ and then summing them up. Instead, I'll just tell you that the answer is 74.25. This solution is *relatively* close to the correct answer, which is 70.

But still, using this approach we're overestimating the true area under the curve by 4.25—that's not ideal. This approach is also considerably more computationally intensive. The solution with integration not only provides the correct answer, it's also just more elegant.

That makes integration incredibly appealing—hopefully you can see that now, too.

When you *can't* integrate

Before we wrap things up, I hasten to point out one important caveat about integrals: the function to be integrated must be a *continuous* function. There *cannot* be any kind of break, or jump, or vertical asymtote.

Problem Sets

Now it's your turn to play around with integrals. You can find the problem sets for this lesson here.