

1 Numerical Integration

Integrating a function is crucial to many problems in science and engineering, and yet it is often the case that finding the analytic form of an integral is difficult. If what one has is just data that defines a function (data that might come from conducting an experiment, for example), then one doesn't even have the actual function expressed in an analytic form.

The solution to this is *numerical integration*, also called *quadrature*. In essence, if we can compute the value of a function given values of the arguments to that function, then we can compute the integral of that function, at least to some degree of numerical accuracy.

It is important to mention that numerical integration is not the same thing as finding the area under the curve. If the function is always positive, then the two are the same, but one must remember that for functions that go negative the integral produces a “positive area” value and a “negative area” value that are summed to get the integral. This is an obvious thing and not conceptually difficult. The problem with numerical integration, unlike doing a calculus exercise, is that with a black box for the function one might not be aware that the function takes on negative values and thus might overlook the distinction.

1.1 The Simplest Method for Numerical Integration

Consider a function known to be positive over an interval x_L to x_R . (This means that we can speak of the integral and the area under the curve interchangeably.) In elementary calculus, one of the first things proven about integrals is that the integral is the limit of the Riemann sum, and the Riemann sum is the sum of rectangles that approximate the area under the curve.

For example, the area $S_0 = f(x_L)(x_R - x_L)$ is the area of a rectangle of height $f(x_L)$ and width $x_R - x_L$. If the function is a constant $y = f(x_L)$, then we have that $f(x_L) = f(x_R)$, the height of the function is a constant, and the area under the curve is exactly S_0 .

Most functions are not constants, though, so we have to continue our approximation. Let's define $x_0 = x_L$, $x_1 = (x_R - x_L)/2$, $x_2 = x_R$, so we have broken the interval up into two intervals with left endpoints x_0 and x_1 . Consider the sum $S_2 = f(x_0)(x_1 - x_0) + f(x_1)(x_2 - x_1)$. Actually, it's simpler to notice that $\Delta_2 = x_1 - x_0 = x_2 - x_1$ and thus that our approximation is

$$S_2 = \Delta_2 \cdot (f(x_0) + f(x_1)).$$

We would hope that this is a better approximation to the area under the curve, because now we are using two points instead of just one to approximate the function. (Draw the picture, which I would do except that's much more work.)

Going further, let $\Delta_4 = (x_R - x_L)/4$, define

$$\begin{aligned}x_0 &= x_L \\x_1 &= x_L + \Delta_4 \\x_2 &= x_L + 2 \cdot \Delta_4 \\x_3 &= x_L + 3 \cdot \Delta_4 \\x_4 &= x_L + 4 \cdot \Delta_4 = x_R\end{aligned}$$

and let's use the four rectangles with left endpoints x_0, x_1, x_2, x_3 as our approximations, so the integral will be approximated by

$$\begin{aligned}S_4 &= (f(x_0) + f(x_1) + f(x_2) + f(x_3)) \cdot \Delta_4 \\&= (f(x_0) + f(x_1) + f(x_2) + f(x_3)) \cdot \left(\frac{x_R - x_L}{4}\right).\end{aligned}$$

By now you are getting the picture. Namely, we let

$$\Delta_N = (x_R - x_L)/N,$$

compute the sum

$$S_N = \left(\sum_{i=0}^{N-1} f(x_L + i \cdot \Delta_N)\right) \Delta_N$$

for increasing values of N .

Note that we could just as easily have chosen to use the right endpoints of the subintervals instead of the left endpoints. As the subintervals get shorter and shorter, the difference between the two choices will disappear, so we lose no generality by simply sticking with the left endpoints.

In truth, what we are doing with this approach numerically is exactly what is done in proving that the integral exists and is equal to the area under the curve for positive functions.

1.2 Example

Let's start with a simple example of a linear function. Let $f(x) = 4.0x + 5.0$, and let's find the area under the line from $x = 1.0$ to $x = 5.0$. Since this is a trapezoid, we can compute the area exactly as half the height times the sum of the bases. In this case the trapezoid is on its side compared to the way it is usually presented in algebra and geometry books, so the "height" is actually the x -distance, from 1.0 to 5.0, and half that is 2.0. The "bases" are $f(1.0) = 9.0$ and $f(5.0) = 25.0$. So the area is $S = 2.0 \cdot (9.0 + 25.0) = 68.0$. Our approximation with one interval is height ($f(1.0) = 9.0$) times width (4.0) to get 36.0. With two intervals, we have

$$S_2 = (f(1.0) + f(3.0)) \cdot 2.0 = (9.0 + 17.0) \cdot 2.0 = 52.0$$

With four intervals, we have

$$S_4 = (f(1.0) + f(2.0) + f(3.0) + f(4.0)) \cdot 1.0 = (9.0 + 13.0 + 17.0 + 21.0) \cdot 1.0 = 60.0$$

and in general we get the following table

N	Δ_N	S_N
1	4.000000	36.000000
2	2.000000	52.000000
4	1.000000	60.000000
8	0.500000	64.000000
16	0.250000	66.000000
32	0.125000	67.000000
64	0.062500	67.500000
128	0.031250	67.750000
256	0.015625	67.875000
512	0.007813	67.937500

1.3 The Trapezoid Rule

Our previous numerical approach was to use rectangles to approximate the area. We can do better than that with only a little more effort. Instead of using the rectangle of width Δ_N and height equal to the functional value $f(x_i)$ at the left endpoint of each subinterval, we can use the trapezoid with one "base" of height $f(x_i)$ and the other base of height $f(x_{i+1})$ at the left and right endpoints, respectively, of each subintervals.

What this does is turn the previous sum

$$S_N = \left(\sum_{i=0}^{N-1} f(x_L + i \cdot \Delta_N) \right) \Delta_N$$

into a better approximation

$$S_N = \left(\sum_{i=0}^{N-1} \frac{f(x_L + i \cdot \Delta_N) + f(x_L + (i+1) \cdot \Delta_N)}{2} \right) \Delta_N$$

If we were to use this on our linear function, we would get an exact answer every time, because the trapezoids we are summing up fix the “curve” exactly. In general, the trapezoid approximation will be better than the rectangular approximation because we will be including the little triangle on top of the rectangle to our computation of area.

1.4 Quadratic Approximations

A final comment is that we can go one step further and use quadratic polynomials for approximating the area under a curve. The trapezoid method is essentially looking at the function $f(x)$ over an interval x_i to x_{i+1} and approximating the function with the straight line between $f(x_i)$ and $f(x_{i+1})$.

The theorem that we use is the one that says that $n + 1$ points determine a unique polynomial of degree n . That is, two points uniquely determine a straight line, three points uniquely determine a quadratic, four points a cubic, and so forth. This is true unless there is some degeneracy involved; for example, if the “two” points are actually the same point, then we don’t get a line, and if three points actually lie on a straight line, then we don’t get a quadratic.

But let us proceed as if we were working in the general case. Let’s assume we have three points (x_0, y_0) , (x_1, y_1) , and (x_2, y_2) . We can fit a quadratic to these three points by brute force solution of the equations for the coefficients. If our quadratic is $f(x) = ax^2 + bx + c$ for unknowns a , b , and c , then we have

$$\begin{aligned} ax_0^2 + bx_0 + c &= y_0 \\ ax_1^2 + bx_1 + c &= y_1 \\ ax_2^2 + bx_2 + c &= y_2 \end{aligned}$$

We have three linear equations in three unknowns, and with some effort we can solve these for a , b , and c .

Or we can pull the rabbit out of the hat and use *Lagrange interpolation*. The function

$$f(x) = y_0 \left(\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} \right) + y_1 \left(\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \right) + y_2 \left(\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \right)$$

is a quadratic polynomial that takes on the value y_0 at x_0 , y_1 at x_1 , and y_2 at x_2 . If we believe the theorem that three points determine a *unique* quadratic polynomial, then $f(x)$ as defined immediately above must be that quadratic. And if we look at the algebra, we find that $f(x) = ax^2 + bx + c$ for

$$\begin{aligned} a &= \frac{y_0}{(x_0-x_1)(x_0-x_2)} + \frac{y_1}{(x_1-x_0)(x_1-x_2)} + \frac{y_2}{(x_2-x_0)(x_2-x_1)} \\ b &= \frac{-y_0(x_1+x_2)}{(x_0-x_1)(x_0-x_2)} + \frac{-y_1(x_0+x_2)}{(x_1-x_0)(x_1-x_2)} + \frac{-y_2(x_0+x_1)}{(x_2-x_0)(x_2-x_1)} \\ c &= \frac{y_0(x_1x_2)}{(x_0-x_1)(x_0-x_2)} + \frac{y_1(x_0x_2)}{(x_1-x_0)(x_1-x_2)} + \frac{y_2(x_0x_1)}{(x_2-x_0)(x_2-x_1)} \end{aligned}$$

To do a quadratic approximation for the area under the curve of a function, we start with $x_0 = x_L$, $x_1 = x_0 + \Delta_N$, and $x_2 = x_0 + 2\Delta_N$ as three points that determine a quadratic. We note that $x_1 - x_0 = \Delta_N$ and $x_2 - x_0 = 2\Delta_N$, so with these equally spaced points we can simplify the formulas further.

$$\begin{aligned} a &= \frac{y_0 - 2y_1 + y_2}{2\Delta_N^2} \\ b &= -\frac{y_0(x_1+x_2) - 2y_1(x_0+x_2) + y_2(x_0+x_1)}{2\Delta_N^2} \\ c &= \frac{y_0(x_1x_2) - 2y_1(x_1x_2) + y_2(x_0x_1)}{2\Delta_N^2} \end{aligned}$$