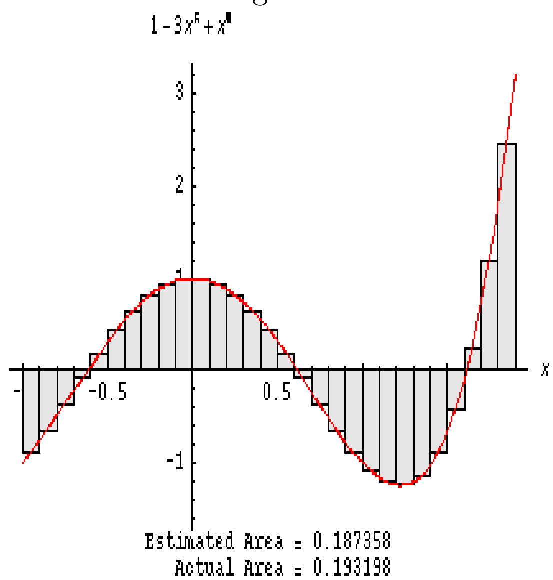
Section 7.7: Approximate Integration

1. The Definite Integrals

Before we start, we recall the definition of a Riemann sum. Suppose that $f(x)$ is a continuous function on some interval $[a, b]$. Break $[a, b]$ up into n subintervals $[x_0, x_1]$, $[x_1, x_2]$, $\ldots$, $[x_{n-1}, x_n]$ where the interval finishing at x_i has length Δx_i (though we usually do, there is no reason why we should insist that the intervals have equal length). For each interval $[x_i, x_{i+1}]$, we choose a point x_{i+1}^* in that interval. We then define the definite integral of $f(x)$ over $[a, b]$ to be

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x_i^*$$

where the limit means we are taking more and more subdivisions. Graphically we have the following:

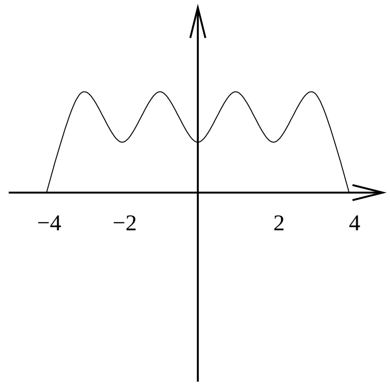


When we initially introduced the idea of a definite integral, we usually insisted that the intervals have equal length. In addition, though in principle x_i^* could be anything, we only really considered two possible choices.

(i) Left Hand Sum: We take the values x_i^* to be the left hand end point of each interval.

(ii) Right Hand Sum: We take the values x_i^* to be the right hand end point of each interval.

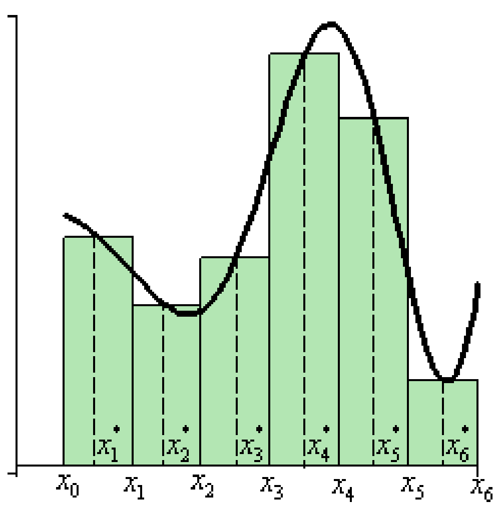
Each of these sums were good under certain circumstances, but there are lots of examples where none of them give a very accurate answer. For example, in the figure below, if we took 4 subdivision and took left and right hand sums, they would both be very inaccurate because the endpoints are precisely where the minimum values are taking place.



With this example in mind, it seems like a good to have as many other types of sum as possible to obtain the best approximation. In this section we shall consider three different types of sum, one called the midpoint rule, one the Trapezoid Rule and the other Simpsons Rule.

2. The Midpoint and Trapezoid Rules

The midpoint rule is the easiest to understand (and you may have actually considered it too!). Instead of taking left or right hand endpoints in the interval, we use the midpoints in the Riemann sum. The basic idea is that by using the midpoint, we will reduce the error because we will "lose" some area and "gain" some area, hopefully cancelling out to give a better approximation (see illustration below).



Result 2.1. The midpoint rule:

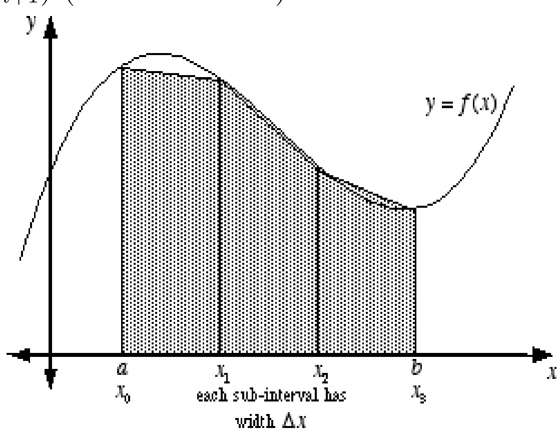
$$\int_a^b = \lim_{n\to\infty} \Delta x \left[f(x_1^*) + f(x_2^*)) + \cdots + f(x_{n-1}^*) + f(x_n^*) \right]$$

where x_i^* is the midpoint of the ith interval.

Example 2.2. Excluding linear functions, there are four basic shapes for graphs depending upon concavity and whether it is increasing or decreasing. Discuss the accuracy of the midpoint rule over the four different types of shapes and in particular whether or not it gives and overestimate, underestimate or whether that is undetermined.

(i) CU/DEC - Underestimate
(ii) CU/INC - Underestimate
(iii) CD/DEC - Overestimate
(iv) CD/INC - Overestimate

Suppose that $f(x)$ is a continuous function on $[a, b]$. We can construct a Riemann sum by instead of taking a sum of rectangles, taking a sum of trapezoids which better fit the graph. In general, if we divide $[a, b]$ up into n intervals, instead of a rectangle with the height determined by a point in each interval, for the interval $[x_i, x_{i+1}]$, we take the trapezoid over the interval $[x_i, x_{i+1}]$ bounded by the x-axis and the line connecting $f(x_i)$ and $f(x_{i+1})$ (see illustration).



The area of the trapezoids (shaded) approximately equals the area bounded by $y = f(x)$.

$$\int_a^b f(x)dx = \frac{\Delta x}{2}\left[f(x_0) + 2f(x_1) + 2f(x_2) + f(x_3)\right].$$

We need to work out the area of each trapeziod. However, from elementary geometry, the area of this trapezoid will be $(f(x_i)+f(x_{i+1}))\Delta x/x$. Summing up the areas of all the trapezoids, we get the total area. Observe unless $i = 0$ or $i = n$, the value $f(x_i)$ will appear twice in the sum (it will be the left hand side of one trapezoid and the right hand side for another trapezoid). Letting $\Delta x- > 0$, we get

Result 2.3. The trapezoid rule:

$$\int_a^b = \lim_{n\to\infty} \frac{\Delta x}{2}\Big[f(x_0) + 2f(x_1)) + 2f(x_2) + \cdots + 2f(x_{n-1}) + 2f(x_n)\Big]$$

Observe that the trapezoid rule is very useful because it does not approximate using rectangles, but instead trapezoids, so we would expect the errors to be a lot less than the errors generated by the left and right hand sums. We look at an example.

Example 2.4. Excluding linear functions, there are four basic shapes for graphs depending upon concavity and whether it is increasing or decreasing. Discuss the accuracy of the trapezoid rule over the four different types of shapes and in particular whether or not it gives and overestimate, underestimate or whether that is undetermined.

(i) CU/DEC - Overestimate
(ii) CU/INC - Overestimate
(iii) CD/DEC - Underestimate
(iv) CD/INC - Underestimate

We now look at a couple of examples of explicit calculations.

Example 2.5. (i) Find the values for the Midpoint and Trapezoid of the function $f(x) = x^2$ with subinterval endpoints $\{0, 1, 3, 4\}$.

First observe that this is not a symmetrical interval, so we cannot use a calculator and we need to modify the formulas for trapezoid and midpoint. We evaluate by hand (remember, the area of a trapezoid $= (f(x_i) + f(x_{i+1}))\Delta x/2$).

Midpoint $= 1(1/2)^2+2(2)^2+1(7/2)^2 = 1/4+8+12+1/4 = 20 + 1/2$

Trapezoid $= 1(0 + 1)/2 + 2(1 + 9)/2 + 1(9 + 16)/2 = 23$

(ii) Determine whether trapezoid or midpoint gives the best approximation to the integral $\int_0^3 3x^2 - 2xdx$ for the values $n = 10$ and $n = 100$.

First we need to calculate he actual value of this integral, and for this we use the fundamental theorem of calculus:

$$\int_0^3 3x^2 - 2xdx = \left[x^3 - x^2\right]_0^3 = 27 - 9 = 18.$$

Using the calculator, for $n = 10$, we get $TRAP = 18.36$, so the error is 0.36, and $MID = 17.93$, so the error is 0.07. For $n = 100$, we get $TRAP = 18.001$, so the error is .001 and $MID = 17.9993$, so the error is 0.0007. Observe that in both cases, the midpoint rule gives us the better approximation.

In both cases, we saw that the error term of the midpoint rule was less than the trapezoid rule, so the midpoint rule is generally the better to use when trying to get the best approximation. In fact, you can use mathematics to work out an explicit bound for an error term which depends upon the second derivative of a function and the number of subintervals taken.

Result 2.6. Suppose $|f''(x)| \leqslant K$ for $a \leqslant x \leqslant b$. If E_T and E_M are the errors in the Trapezoid and Midpoint Rules, then

$$|E_T| \leqslant \frac{K(b-a)^3}{12n^2}$$

and

$$|E_M| \leqslant \frac{K(b-a)^3}{24n^2}$$

(so in fact the midpoint is generally twice as good as the trapezoid).

These error bounds are very useful when trying to determine the number of subdivisions required to get the trapezoid or midpoint rules within a certain distance of the actual answer. We illustatrate with an example.

Example 2.7. Determine the number of subdivisions required to guarantee the trapezoid and midpoint rules will be within 0.001 of the actual value of $\int_0^4 e^x dx$.
First, we observe that $f''(x) = e^x$, so $f''(x) \leqslant e^4 \leqslant 54.6$ on $[0, 4]$. We want $|E_T| \leqslant 0.001$ and $|E_M| \leqslant 0.001$, so we should set

$$\frac{K(b-a)^3}{12n^2} \leqslant \frac{54.6 * 64}{12n^2} = \frac{291}{n^2} \leqslant 0.001$$

, so $291/0.001 \leqslant n^2$, or $n \geqslant 540$ and

$$\frac{K(b-a)^3}{24n^2} \leqslant \frac{145}{n^2} \leqslant 0.001$$

so $145/0.001 \leqslant n^2$ or $n \geqslant 380$.

3. Simpsons Rule

We have discovered the following:

(*i*) The Midpoint rule as a whole is a better approximation by a factor of 2 than the Trapezoid rule.
(*ii*) For the different shapes of graphs, whenever MID is an over/underestimate, TRAP is an under/overestimate.

We put these two facts together to develop a third rule which is more accurate than any of those we have seen before.

Result 3.1. We define Simpsons rule as

$$SIMP(2n) = \frac{2MID(n) + TRAP(n)}{3}$$

So Simpsons rule takes the average of two midpoints and one trapeziod (which reflects precisely what we observed above).

Warning. There are two common definitions for Simpsons rule: what we have defined for $SIMP(2n)$, some books and computer programs means $SIMP(n)$ (so you may see examples on webwork where instead of 100 intervals, in order to get the correct answer, you need to use 50 intervals). If in doubt, try both.

Provided we can use MID and $TRAP$, $SIMP$ is also very easy (we just need to be careful that when calculating $SIMP(100)$ for example, we need to use $TRAP(50)$ and $MID(50)$. Just as with $TRAP$ and MID, an error bound can be calculated which depends only upon the number of subintervals used and this time the fourth derivative.

Result 3.2. Suppose $|f^4(x)| \leqslant K$ for $a \leqslant x \leqslant b$. If $|E_S|$ is the error involved using Simpsons rule, then $|E_S| \leqslant \frac{K(b-a)^5}{180n^4}$.

We finish with an example of how to use the error bound.

Example 3.3. How many subdivisions should be used to guarantee that $SIMP$ will be within 0.001 of the actual value of $\int_0^4 e^x dx$?
This is the same as the other examples:

$$|E_S| \leqslant \frac{54.6 * 1024}{180n^4} = \frac{311}{n^4} \leqslant 0.001$$

so $n^4 \geqslant 31100$, or $n \geqslant 24$. This means when using the formula, we would have $SIMP(24) = SIMP(2*12) = (2MID(12)+TRAP(12))/3$. Checking, we get $SIMP(24) = (2 * 53.35 + 54.09)/(3) = 53.5967$.