7.7) Approximate Integration:

-So if there is no way to integrate e^{x^2} by hand, how would we determine $\int_0^1 e^{x^2} dx$? While we have seen that there are ways of taking definite integrals of expressions without integrating them, like $\int_{-1}^{1} \sqrt{1 - x^2} dx$, we have no geometric knowledge to help us with an expression like this one.

-If we can't take definite integrals, we can always go back to the basics: approximation with Riemann sums. Don't forget, we still have that in general:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_i^*)\Delta x$$

-You can still approximate the area under a curve from one end to another by creating a series of rectangles formed by left-endpoint rectangles or right-endpoint rectangles. We can still create equal-width subintervals (width $\Delta x = \frac{b-a}{n}$) on $x_0 = a,\ x_1, \ldots, x_{n-1}, x_n = b$ and form a **left endpoint approximation** or **right endpoint approximation** using respectively:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_{i-1})\Delta x = L_n \qquad \int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_i)\Delta x = R_n$$

-Of course, you probably remember that the best approximations of area under a curve are still found with **midpoint approximations**, that use midpoints:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f\left(\frac{x_i - x_{i-1}}{2}\right)\Delta x = \sum_{i=1}^{n} f(\overline{x}_i)\Delta x = M_n$$

-However, a new rule that can be used to approximate definite integrals is using something called the **Trapezoidal Rule**, which is that if you want to approximate, you could always take an average of the left and right endpoint approximations:

$$\int_a^b f(x)dx \approx \frac{1}{2}\left(\sum_{i=1}^{n} f(x_i)\Delta x + \sum_{i=1}^{n} f(x_{i-1})\Delta x\right) = T_n$$

-From this definition, you can even combine terms of the summation together, since left-endpoints of one subinterval are right-endpoints of the previous subinterval:

$$\int_a^b f(x)dx \approx \frac{1}{2}\Big((f(x_1)\Delta x + f(x_2)\Delta x + \ldots + f(x_n)\Delta x) + (f(x_0)\Delta x + f(x_1)\Delta x + \ldots + f(x_{n-1})\Delta x)\Big)$$

$$\int_a^b f(x)dx \approx \frac{1}{2}\Big(f(x_0)\Delta x + 2f(x_1)\Delta x + 2f(x_2)\Delta x + \ldots + 2f(x_{n-1})\Delta x + f(x_n)\Delta x\Big)$$

$$\int_a^b f(x)dx \approx \frac{\Delta x}{2}\left(f(x_0) + 2f(x_1) + 2f(x_2) + ... + 2f(x_{n-1}) + f(x_n)\right)$$

-Why call it the Trapezoidal rule? It has to do with how the shapes used in this approximation summation have more in common with trapezoids than rectangles. On top of that, a rewrite of some of the terms used in the trapezoidal summation can turn the sum of areas of rectangles into a sum of trapezoids.

-Either way, whether you use the midpoint rule or the trapezoidal rule to approximate definite integrals, you will need to know what your endpoints and midpoints are, and your width in order to compute areas: $\Delta x = \frac{b-a}{n}$,

$$ith\, left - endpoint = a + (i - 1)\Delta x, \qquad ith\, right - endpoint = a + (i)\Delta x,$$

$$ith\, midpoint = a + \frac{(2i-1)}{2}\Delta x$$

Example: Use the midpoint rule and the trapezoidal rule to approximate $\int_1^2 \frac{1}{x}dx$ with n = 5 subintervals.

Solution: We know from definite integrals that this equals $ln(2) \approx 0.6931471806$, but let's see how close an approximation we can get from 5 subintervals.

First we'll do the midpoint rule, which requires 5 midpoints with width $\Delta x = \frac{2-1}{5} = 0.2$:

$$\overline{x}_1 = 1 + \frac{(2(1)-1)}{2} * 0.2 = 1.1, \qquad \overline{x}_2 = 1.3, \qquad \overline{x}_3 = 1.5, \qquad \overline{x}_4 = 1.7,$$

$$\overline{x}_5 = 1.9$$

We have the midpoints, we have the width, so let's approximate:

$$\int_1^2 \frac{1}{x}dx \approx f(1.1)(0.2) + f(1.3)(0.2) + f(1.5)(0.2) + f(1.7)(0.2) + f(1.9)(0.2)$$

$$\int_1^2 \frac{1}{x}dx \approx (0.2)\left(\frac{1}{1.1} + \frac{1}{1.3} + \frac{1}{1.5} + \frac{1}{1.7} + \frac{1}{1.9}\right) \approx 0.691908$$

Not a bad approximation. What about the trapezoidal rule? We will need endpoints for this, from x_0 to x_5:

$$x_0 = 1 \qquad x_1 = 1.2 \qquad x_2 = 1.4 \qquad x_3 = 1.6 \qquad x_4 = 1.8 \qquad x_5 = 2$$

Remember, you plug x_0 and x_5 into $\frac{1}{x}$, but you plug the rest into $2\left(\frac{1}{x}\right)$. You use the output of the function for the first left endpoint and last right endpoint, but you use twice the output for the rest. The trapezoidal rule approximation equals:

$$\int_1^2 \frac{1}{x}dx \approx \frac{1}{2}(f(1)(0.2) + 2f(1.2)(0.2) + 2f(1.4)(0.2) + 2f(1.6)(0.2) + 2f(1.8)(0.2) + f(2)(0.2))$$

$$\int_1^2 \frac{1}{x}dx \approx \frac{0.2}{2}\left(\frac{1}{1} + \frac{2}{1.2} + \frac{2}{1.4} + \frac{2}{1.6} + \frac{2}{1.8} + \frac{1}{2}\right) \approx 0.695635$$

-They are both fairly good approximations, though they are both a little off. The **error** in using an approximation is defined as the amount that needs to be added to that approximation to make it exact. In other words:

$$error = Actual - Approximation,$$ or

$$\int_a^b f(x)dx = Approximation + error$$

-A positive error implies the approximation is an underestimate, while a negative error means the approximation is an overestimate. So if we found the errors for our approximations from the last example, we would have:

$E_M = ln(2) - 0.691908 = 0.001239,$ $\quad E_T = ln(2) - 0.695635 = -0.002488$

-So in general, the error from using the midpoint rule and the trapezoidal rule respectively are:

$$E_M = \int_a^b f(x)dx - M_n \qquad E_T = \int_a^b f(x)dx - T_n$$

Exercise: Use the midpoint rule and the trapezoidal rule to approximate $\int_1^2 e^x dx$ with n = 5 subintervals, and find the error of each compared to the actual area of $e^2 - e^1 \approx 4.67077427$.

-As has been the case with all Riemann sums and approximations we have seen this semester, the more shapes that are used, the larger n gets, and the more accurate the approximation will be, whether you use the midpoint rule or the Trapezoidal rule. That also means the errors from these two rules will get smaller as n gets larger.

-However, here are some other observations worth noticing for most uses of the various approximation rules:

-The errors in left and right endpoint approximations are opposite in sign and decrease by a factor of about 2 when n is doubled.

-The Trapezoidal and Midpoint Rules are both more accurate than the endpoint approximations.

-The Trapezoidal and Midpoint Rules are opposite in sign and decrease by a factor of about 4 when n is doubled.

-The size of the error in the Midpoint Rule is about half the size of the error in the Trapezoidal Rule.

-That raises the interesting question of: if you want to approximate the area under a curve to within a certain error, how many rectangles/trapezoids will you need? How big does n need to be for an approximation to be "close enough" to the actual definite integral value?

-We don't have time to prove it in this class (see numerical analysis for more info), but estimates and approximations actually depend on the curvature of the function more than anything, which is found by the second derivative of the function. As long as the curvature of the function f(x) does not exceed a constant value K on the interval [a,b], then the errors of approximations using the midpoint or trapezoidal rule won't exceed a particular value either.

Error Bounds:

Suppose $|f''(x)| \leq K$ for $a \leq x \leq b$, and E_T and E_M are the Trapezoidal and Midpoint rule errors for an approximation using n subintervals. Then:

$$\left|E_T\right| \leq \frac{K(b-a)^3}{12n^2} \qquad\qquad \left|E_M\right| \leq \frac{K(b-a)^3}{24n^2}$$

-For example, take $f(x) = \frac{1}{x}$, our case from earlier. $f''(x) = -\frac{2}{x^3}$, which has a largest absolute value at x = 1, where $f''(1) = -2$. So if we let K = 2, then we can be certain that the approximations found using the trapezoidal rule will be no larger than:

$$\left|E_T\right| \leq \frac{2(2-1)^3}{12(5)^2} = \frac{2}{150} = 0.006667$$

-It should be no surprise our approximation from earlier, -0.002488, was smaller than this, and if we took the time to find the error bound for the Midpoint rule, we'd get that it was no larger than 0.003333.

-You might also be wondering, did we have to let K = 2? Can't we use any value K that is bigger than the maximum value of $f''(x)$ on the interval? Yes, we can, but why would we want to? After all, a small error bound is more useful than a larger one, so if we have the option of choosing any K bigger than $f''(x)$, why wouldn't we pick the smallest K that is bigger than or equal to $f''(x)$?

-Still, the real benefit of the error bound is it gives us a means of being able to find out how large n must be to get an approximation from the midpoint or trapezoidal rule that is accurate enough for whatever purpose we may need.

Example: How large should n be in order to guarantee that the Trapezoidal and Midpoint Rule approximations for $\int_1^2 \frac{1}{x}dx$ are accurate to within 0.0001?

Solution: To say that the approximations are within a certain value is another way of saying the error bound is less than or equal to that certain value. So we need to find out what n should be so that:

$$\left|E_T\right| < 0.0001$$

We can find this out for the Trapezoidal Rule alone, since the error bound for the Midpoint Rule will be about half of the error bound for the Trapezoidal Rule, so if n is large enough to make the Trapezoidal rule accurate to within 0.0001, then it will be large enough to make the Midpoint Rule accurate to within 0.0001 as well (probably less). This is the same as determining what n must be for the following to be true:

$$\frac{K(b-a)^3}{12n^2} < 0.0001$$

We know that the curvature constant K is 2, we know that (b-a) = (2-1)=1, so we need to solve for n in:

$$\frac{2}{12n^2} < 0.0001$$
$$\frac{12n^2}{2} > 10000$$
$$n^2 > \frac{10000}{6}$$
$$n > \sqrt{\frac{10000}{6}} \approx 40.8$$

In order for the Trapezoidal Rule approximation to be 0.0001 or less from the actual value of $\int_1^2 \frac{1}{x} dx$, you need to use 41 or more rectangles. For the record, if you did this for the midpoint rule, you'd need 29 rectangles to be within 0.0001 of the actual value, but again 41 rectangles ensures it for both rules.

-If the error bound value is given to you, you can use it to find n that makes the approximations accurate to within the error bound. However, if the phrasing "within k decimal places" is used (which is quite common) then the value you can use for your error bound can be $0.5 * 10^{-k}$.

Exercise: How large should n be in order to guarantee that the Trapezoidal and Midpoint Rule approximations for $\int_1^2 e^x dx$ are accurate to within 0.0001?

-Now let's tackle that pesky example that has haunted us for two sections now:

Example: a) Use the Midpoint Rule with n = 10 to approximate the integral $\int_0^1 e^{x^2} dx$.

b) Give an upper bound for the error involved in this approximation.

Solution: a) We need 10 midpoints, separated by a width of $\frac{b-a}{n} = \frac{1-0}{10} = 0.1$,

$\bar{x}_1 = 0 + 0.05 = 0.05, \bar{x}_2 = 0.15, \bar{x}_3 = 0.25, \bar{x}_4 = 0.35, \bar{x}_5 = 0.45, \bar{x}_6 = 0.55, \bar{x}_7 = 0.65,$
$\bar{x}_8 = 0.75, \bar{x}_9 = 0.85, \bar{x}_{10} = 0.95$

We have the midpoints, so we have to plug them all in, add the outputs up, and multiply by 0.1:

$\int_0^1 e^{x^2} dx \approx 0.1\left(e^{0.0025} + e^{0.0225} + e^{0.0625} + e^{0.1225} + e^{0.2025} + e^{0.3025} + e^{0.4225} + e^{0.5625} + e^{0.7225} + e^{0.9025}\right)$
≈ 1.460393

b) How accurate is this approximation? Let's find the error bound:

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

We will need K, which is the maximum value of the second derivative of e^{x^2} on $[0,1]$. This appears to be an increasing function on [0,1] (e^x increases everywhere, and x^2 increases over all non-negative intervals), so the maximum value of $\frac{d^2}{dx^2}\left[e^{x^2}\right]$ occurs at x = 1. What is the second derivative of e^{x^2}? Let's find out:

$$\frac{d^2}{dx^2}\left[e^{x^2}\right] = \frac{d}{dx}\left[2xe^{x^2}\right] = \left[2e^{x^2} + 4x^2e^{x^2}\right]$$

The second derivative equals 6e when x = 1, so this is what we will let K equal. Now we can find the upper bound of the midpoint rule approximation for $\int_0^1 e^{x^2}\,dx$:

$$\left|E_M\right| \le \frac{6e(1)^3}{24(10)^2} = \frac{e}{400} \approx 0.007$$

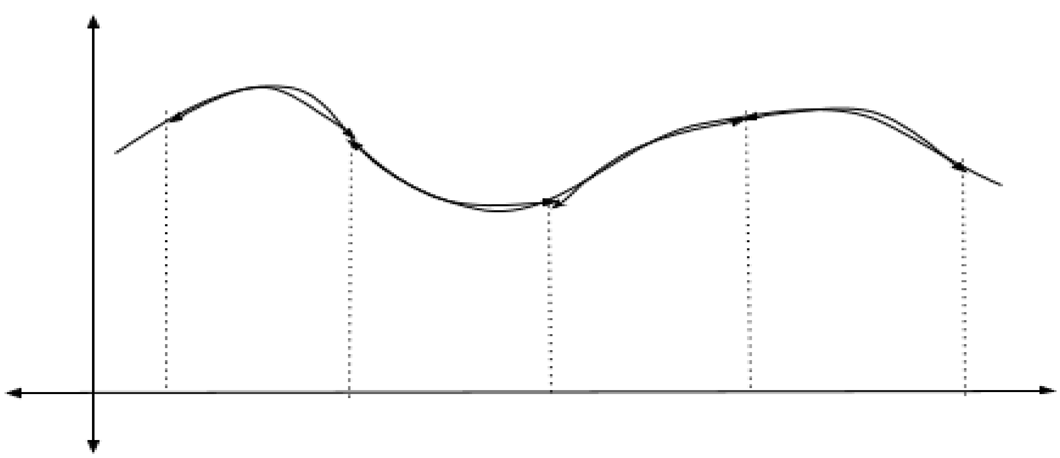
Our midpoint rule estimate of $1.460393$is no more than about 0.007 units from the actual value of $\int_0^1 e^{x^2}\,dx$.

It could be anywhere from about 1.453393 to 1.467393.

Exercise: a) Use the Midpoint Rule with n = 10 to approximate the integral $\int_0^1 sin(x^2)dx$.

b) Give an upper bound for the error involved in this approximation.

-We've used rectangles and trapezoids to approximate definite integrals, but how about parabolas? The area under parabolas does require integrals, but parabolas are not particularly difficult to integrate, so they can be used to approximate areas under curves too.



-Divide [a,b] into n subintervals of equal length, $\Delta x = \frac{b-a}{n}$. Then starting with $[x_0, x_1]$ and $[x_1, x_2]$,draw a parabola that passes through $P_0 = f(x_0)$, $P_1 = f(x_1)$, and $P_2 = f(x_2)$. Then do the same for the next consecutive three endpoints and consecutive two subintervals. We use lots of parabolas, such that any given parabola intersects the curve in three places, call them P_i, P_{i+1}, $and\ P_{i+2}$.

-We will need the equations of these parabolas we create which can be rather difficult to find. So to simplify, let's take a look at the first integral that would be formed to find area under the first parabola:

-Suppose the first three points are $x_0 = -h$, $x_1 = 0$, and $x_2 = h$, and so the area under the generic parabola that intersects the curve at x = -h, x = 0 and x = h could be expressed by:

$$\int_{-h}^{h}(Ax^2 + Bx + C)dx = 2\int_{0}^{h}(Ax^2 + C)dx = \frac{2Ax^3}{3} + 2Cx\Big|_0^h = \frac{h}{3}(2Ah^2 + 6C)$$

-Now presume that the points this first parabolas passes through are: $(-h, y_0),\ (0, y_1),\ and\ (h, y_2)$. Plugging these numbers into $Ax^2 + Bx + C = y$ gives us:

$$Ah^2 - Bh + C = y_0$$
$$C = y_1$$
$$Ah^2 + Bh + C = y_2$$

-Given our earlier area under the curve was $\frac{h}{3}(2Ah^2 + 6C)$, we can actually see from the above system of equations that this area under the curve can also be written in terms of y_0, y_1, and y_2:

$$\frac{h}{3}(2Ah^2 + 6C) = \frac{h}{3}(y_0 + 4y_1 + y_2)$$

-This expression, $\frac{h}{3}(y_0 + 4y_1 + y_2)$ is our standard expression that we can use to approximate the area under any given parabola, accounting for the fact that different parabolas will use different endpoints and thus have different y_i values. The first parabola will have an area under the curve of about $\frac{h}{3}(y_0 + 4y_1 + y_2)$, the second parabolas will have an area of about $\frac{h}{3}(y_2 + 4y_3 + y_4)$, and so on.

-If we add up all the parabola areas together, we get the following approximate area under the curve f(x):

$$\int_a^b f(x)dx \approx \frac{h}{3}\Big((y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + ... + (y_{n-2} + 4y_{n-1} + y_n)\Big) =$$
$$\int_a^b f(x)dx \approx \frac{h}{3}\Big(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + ... + 2y_{n-2} + 4y_{n-1} + y_n\Big)$$

-This summation approximation is called **Simpson's Rule**, which is a means of approximating the area under the curve f(x) from x = a to x = b as long as n, the number of subintervals, is even. To write it in more familiar notation, see below:

Simpson's Rule:

If n is even and $\Delta x = \frac{b-a}{n}$, then

$$\int_a^b f(x)dx \approx \frac{\Delta x}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + ... + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\right) = S_n$$

Example: Use Simpson's Rule with n = 10 to approximate $\int_1^2 \frac{1}{x}dx$.

Solution: We have $\Delta x = 0.1$, so plugging the endpoints 1, 1.1, 1.2, ... 1.9, 2 into $\frac{1}{x}$, the approximation is about:

$$\int_1^2 \frac{1}{x}dx \approx \frac{(0.1)}{3}(f(1) + 4f(1.1) + 2f(1.2) + 4f(1.3) + 2f(1.4) + 4f(1.5) + 2f(1.6) + 4f(1.7) + 2f(1.8) + 4f(1.9)$$

$$\frac{(0.1)}{3}\left(\frac{1}{1} + \frac{4}{1.1} + \frac{2}{1.2} + \frac{4}{1.3} + \frac{2}{1.4} + \frac{4}{1.5} + \frac{2}{1.6} + \frac{4}{1.7} + \frac{2}{1.8} + \frac{4}{1.9} + \frac{1}{2}\right) \approx 0.693150$$

Exercise: Use Simpson's Rule with n = 6 to approximate $\int_1^2 e^x dx$.

-In general, the arithmetic used for Simpson's Rule shows that Simpson's Rule approximations are actually a linear combination of Trapezoidal Rule and Midpoint Rule approximations:

$$S_n = \frac{1}{3}T_n + \frac{2}{3}M_n$$

-You may recall that T_n and M_n have errors that are opposite signs, which implies that S_n is actually a better approximation in general than either of the other two. What is the error bound for Simpson's Rule? Again, we don't have room for the proof here, but not surprisingly, the error bound is similar to that of the other two rules we have seen:

-Suppose that $\left|f^{(4)}(x)\right| \leq K$ for $a \leq x \leq b$. If E_S is the error involved in using Simpson's Rule,

$$\left|E_S\right| \leq \frac{K(b-a)^5}{180n^4}$$

Example: How large should we take n in order to guarantee that the Simpson's Rule approximation for $\int_1^2 \frac{1}{x}dx$ is accurate to within 0.0001?

Solution: If f(x) = $\frac{1}{x}$, then the fourth derivative of f(x) is $f^{(4)}(x) = \frac{24}{x^5}$. $f^{(4)}(x)$ is largest at x = 1 and decreases from there, so let K = 24.

We can now solve for n in the error bound inequality for Simpson's Rule:

$$|0.0001| \leq \frac{24(2-1)^5}{180n^4}$$

$$0.0001 \leq \frac{24}{180n^4}$$

$$\frac{24}{180*0.0001} \geq n^4$$

$$n \leq \sqrt[4]{\frac{24}{180*0.0001}} = 6.04$$

Always round up to the nearest even number. Therefore you would need at least 8 subintervals to get the approximation within 0.0001 of the actual value.

-As you can see, the number of subintervals needed is much lower here than the others. There may be more work involved, but the approximation tends to be more accurate with a fixed number of subintervals for Simpson's Rule than the other rules.

Exercise: a) Use Simpson's rule with n = 10 to approximate the integral $\int_0^1 e^{x^2}\,dx$.

b) Estimate the error involved in this approximation (find the error bound).