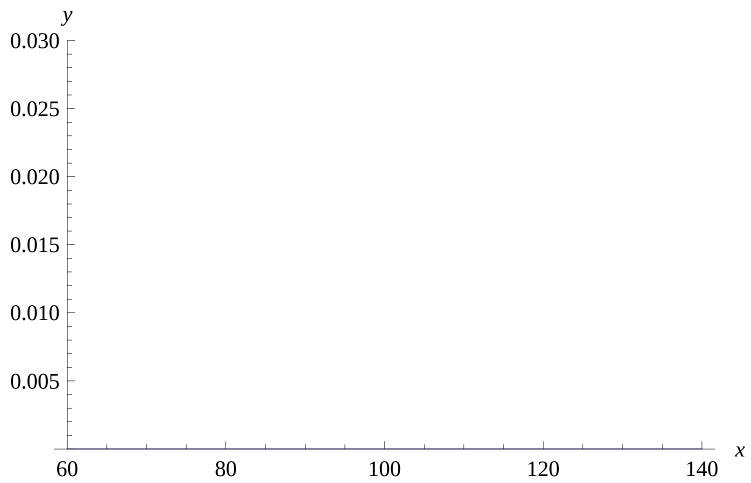
Approximating Definite Integrals

Purpose. The purpose of this lab is to investigate several different methods of approximating definite integrals. We will see how the errors for the various approximation methods compare to one another, and how they change when more subintervals are used. We will derive error bounds for the approximation techniques and see how those error bounds are used.

Preview. One of the most important functions in probability is the normal density function given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ and σ are constants called the mean and standard deviation, respectively. Let's assume that $\mu = 100$ and $\sigma = 15$. Graph this function on your calculator. Use an x-range between 60 and 140 and a y-range between 0 and 0.03. Sketch the graph below:



Notice the distinctive bell-curve shape. This particular normal distribution, with mean $\mu = 100$ and standard deviation $\sigma = 15$, is often used as a model for the distribution of IQs. This means that to find the fraction of individuals whose IQ is between a and b, we merely have to evaluate the following definite integral:

$$\int_a^b p(x)\,dx = \int_a^b \frac{1}{15\sqrt{2\pi}} e^{-\frac{(x-100)^2}{2(15^2)}}\,dx.$$

(a) Write down a definite integral that represents the fraction of the population whose IQs fall between 90 and 110.

(b) The Mensa organization is for people with very high IQs. The only criterion for membership is to have an IQ within the top 2 percent of the population. Suppose that

A is the lowest IQ that you could have and still be a member of Mensa. Write down the equation that A must satisfy using our current IQ distribution function.

(c) Solving these types of problems amounts to integrating the IQ density function p. Find an antiderivative for

$$\frac{1}{15\sqrt{2\pi}}e^{-\frac{(x-100)^2}{2(15^2)}}.$$

Check your answer by differentiation. Don't spend more than five minutes on this step.

You should have had significant trouble with the last part. In fact, it has been shown that you cannot find an antiderivative of p using compositions of power functions, exponential functions, trigonometric functions, logarithmic functions, or any combination of these involving $+ - \times \div$. If we can't find an explicit antiderivative for a normal density function, then it seems that we are severely limited in the types of questions that we can answer using this framework. Fortunately, in practice we only need to *approximate* these definite integrals. Let's discuss various methods of approximating definite integrals.

Part I: Right-Hand Sum, Left-Hand Sum

Write down the definition of the definite integral:

$$\int_a^b f(x)\,dx =$$

where $\Delta x = \frac{b-a}{n}$ and x_k^* is any point in the kth subinterval. If we did not care about computing the definite integral exactly, but only wanted to *approximate* the integral, we might just choose a particular value of n and compute the sum in the definition, without worrying about the limit. We will argue that if n is large, our result will be very close to the actual limit.

In a previous lab, we encountered the Right and Left-Hand Sums on n subintervals for $\int_a^b f(x)\,dx$. For this method of approximation, we choose x_k^* to be the right or left-hand endpoints of each subinterval. Fill in the rest of the definition for the Right and Left-Hand Sums for the integral $\int_a^b f(x)\,dx$ with n subintervals.

$$\text{RHS}(n) = \sum_{k=}$$

$$\text{LHS}(n) = \sum_{k=}$$

We will define the error for a particular approximation method to be

$$E(n) = \text{Actual value} - \text{Estimated value}.$$

We will test these approximation methods on the integral $\int_1^2 \frac{1}{x}\,dx$. In practice, we would not need to approximate this integral because we can compute it exactly. Do so now.

$$\int_1^2 \frac{1}{x}\,dx =$$

Since we know the exact value of this integral, we will be able to find the error exactly. (We're "pretending" that the calculator gives exact values, which is not true.) Compute LHS(n), RHS(n), and the errors for the integral $\int_1^2 \frac{1}{x}\,dx$ and fill in the appropriate boxes in the following tables. You may round your answers to seven decimal places. Leave the last column blank for now.

n	RHS(n)	$E(n)$	Ratio
2			X
4			
8			
16			

n	LHS(n)	$E(n)$	Ratio
2			X
4			
8			
16			

You should have obtained an error of 0.1098138 for RHS(2) and an error of 0.0586234 for RHS(4). Notice that $0.0586234 \div 0.1098138 = 0.5338437$. In the "Ratio" column of the Right-Hand Sum table, fill in 0.5338437 for the ratio in the $n = 4$ row. Fill in the last column for the two tables above. In each row, divide the current error by the error in the previous row and fill that number in the "Ratio" column.

(a) Approximately what happens to the error in RHS for this particular definite integral as we double n, the number of subintervals?

(b) Approximately what happens to the error in LHS for this particular definite integral as we double n, the number of subintervals?

Part II: Midpoint Sum

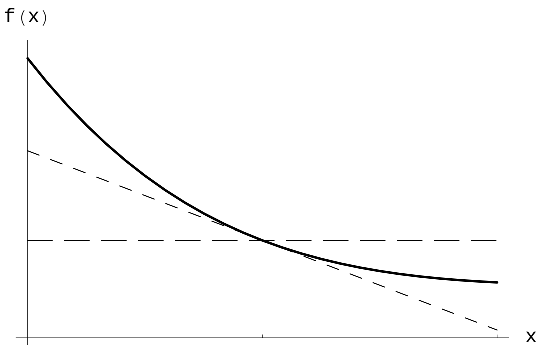
For the Midpoint Sum, we choose x_k^* to be the midpoint of the kth subinterval. To find a formula for the midpoint of the kth subinterval, take the average of the left-hand endpoint and the right-hand endpoint.

(a) Show that the midpoint of the kth subinterval is $a + (k + \frac{1}{2})\Delta x$.

(b) Now fill in the formula for the Midpoint Sum with n subintervals:

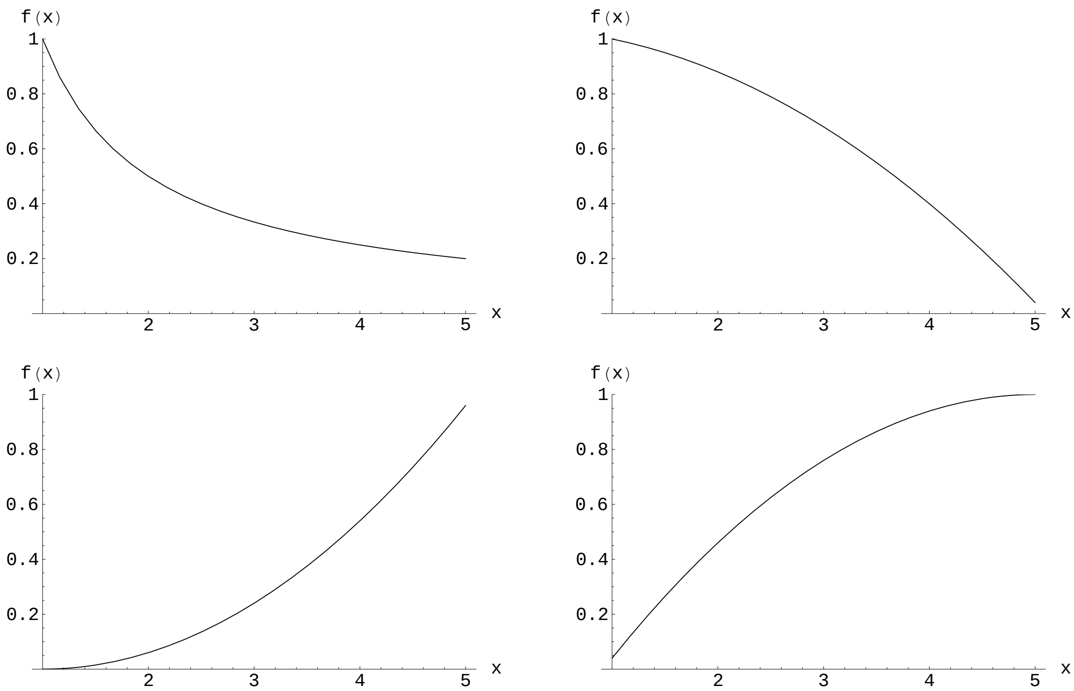
$$\mathrm{MID}(n) = \sum_{k=}$$

(c) To understand when the Midpoint Sum is an underestimate and when it is an overestimate, consider the following picture:



The sloped line is the tangent line to the graph at the midpoint of the interval. Explain why the area underneath the tangent line is equal to the midpoint approximation of this function over one interval.

(d) Draw sloped lines corresponding to Midpoint Sums with one subinterval on each of the following graphs, then use them to answer the questions below.



Which types of functions do Midpoint Sums overestimate?
Which types of functions do Midpoint Sums underestimate?

(e) Now use your calculator to fill in the following table, rounding your answer to seven decimal places, just as before. Estimate the same integral, $\int_1^2 \frac{1}{x}\,dx$. This time, you may fill out the ratio of errors, just as you did for LHS and RHS.

n	MID(n)	$E(n)$	Ratio
2			X
4			
8			
16			

Approximately what happens to the error in MID as we double n, the number of subintervals?

Part III: Trapezoid Sum

Look back at the error tables for LHS and RHS. You may have noticed that for $n = 2$, the error from the LHS approximation and the error from the RHS approximation add together to give something very close to zero. The same is true for the other values of n that you computed. To get a better approximation, we can average the LHS and RHS. We call this new approximation the Trapezoid Sum:

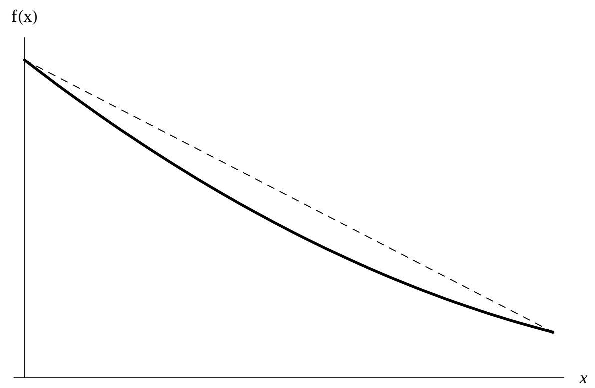
$$\text{TRAP}(n) = \frac{1}{2}(\text{LHS}(n) + \text{RHS}(n)).$$

(a) In order to get the errors to approximately cancel, why didn't we just add LHS(n) and RHS(n)? Why did we have to take their average?

(b) Use the formulas for RHS(n) and LHS(n) to show that TRAP(n) can be computed using the following formula:

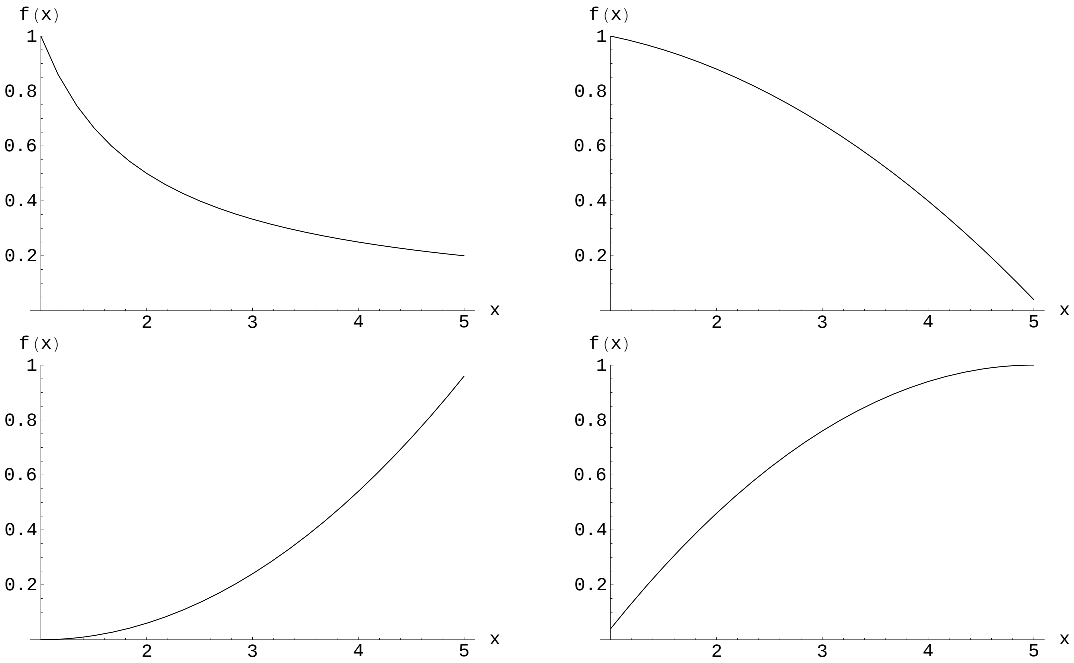
$$\text{TRAP}(n) = \left(\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{1}{2}f(x_n)\right)\Delta x$$

(c) To see why we call this new sum a Trapezoid Sum, look at the following graph.



Draw rectangles representing a Left-Hand Sum and a Right-Hand Sum on the previous graph. Explain why the area underneath the sloped line is equal to the average of the areas underneath the LHS rectangle and the RHS rectangle. Why is this called a Trapezoid Sum?

(d) Draw sloped lines corresponding to Trapezoid Sums with one subinterval on each of the following graphs.



Which types of functions do Trapezoid Sums overestimate?
Which types of functions do Trapezoid Sums underestimate?

(e) Now use your calculator to fill in the following table, rounding your answer to seven decimal places, just as before. Estimate the same integral, $\int_1^2 \frac{1}{x}\,dx$. This time, you may fill out the ratio of the errors, just as you did for LHS, RHS, and MID.

n	TRAP(n)	$E(n)$	Ratio
2			X
4			
8			
16			

Approximately what happens to the error in TRAP as we double n, the number of subintervals?

Part IV: Simpson's Rule

Notice that the error for MID(16) is about 0.0001220, while the error for TRAP(16) is about −0.0002440. We see that the error for the Trapezoid Sum is about −2 times the error for the Midpoint Sum. In fact, this is true for all the n that you computed. In order to get a much smaller error we will define a new approximation technique called Simpson's Rule:

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

(a) Explain why this is a good choice based on our observation. Why did we have to divide by 3?

(b) Now use your calculator to fill in the following table, rounding your answer to seven decimal places, just as before. Estimate the same integral, $\int_1^2 \frac{1}{x}\,dx$. This time, you may fill out the ratio of the errors, just as you did for LHS, RHS, MID and TRAP.

n	SIMP(n)	$E(n)$	Ratio
2			X
4			
8			
16			

Approximately what happens to the error in SIMP as we double n, the number of subintervals?

(c) Use the formulas for TRAP and MID to derive the following formula for Simpson's Rule:

$$\text{SIMP}(n) = \frac{1}{6}\Big(f(x_0) + 4f(x_{\frac{1}{2}}) + 2f(x_1) + 4f(x_{\frac{3}{2}}) + \cdots + +4f(x_{\frac{n-1}{2}}) + f(x_n)\Big)\,\Delta x.$$

Note that if k is an integer, then x_k is from the Trapezoid Sum. Otherwise x_k is from the Midpoint Sum.

(d) Suppose you were given the following values for a function f:

x	3	4	5	6	7	8
$f(x)$	3.4	4.1	4.5	4.3	3.0	2.9

Explain why you cannot use this data to calculate a Simpson's Rule approximation for $\int_3^8 f(x)\,dx$. Instead, use the data to calculate a Simpson's Rule approximation for $\int_3^7 f(x)\,dx$. In general, Simpson's rule requires an _________ number of points.

Part V: Error Bounds

There are some serious gaps in the development thus far. We investigated the errors in the approximation techniques for a single integral that we can solve exactly. Do the patterns that we observed hold even for integrals that we cannot evaluate exactly? Is it possible to know something about the error without being able to find the error exactly? It would seem that these approximation techniques are useless unless we can somehow find a way to quantify the error.

Fortunately, there is a very precise way of getting an upper bound on the error for each approximation technique. We state the theorems here. Students interested in the proof of the theorems should consult the supplemental section of this lab.

Theorem: Let f be a function that is differentiable on (a, b) and continuous on $[a, b]$. Also, suppose that $|f'(x)| \leq M_1$ for all x in the interval $[a, b]$. Then E_n, the error used in approximating $\int_a^b f(x)\,dx$ with a Left-Hand Sum with n subintervals, satisfies the following inequality:

$$|E_n| \leq \frac{M_1(b-a)^2}{2n}.$$

(The theorem remains true if "Left-Hand Sum" is replaced with "Right-Hand Sum" in the statement of the theorem.)

Let us see how to apply this theorem. Recall that earlier we were interested in integrals of the form

$$\int_a^b \frac{1}{15\sqrt{2\pi}} e^{-\frac{(x-100)^2}{2(15^2)}}\,dx.$$

We can simplify matters by instead considering the normal density function with mean $\mu = 0$ and standard deviation $\sigma = 1$. This function is the "simplest" normal density function, in some sense, and shares many properties with all the other normal density functions.

We will apply the theorem to the integral

$$\int_{-1}^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

(a) Find the derivative of the function defined by $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

(b) Find the maximum and minimum values of f' over the interval $[-1, 1]$. Use them to define a number M_1 that fits the hypotheses of the theorem. In order to get the strictest error bound, try to use the smallest possible value for M_1.

(c) Approximate $\displaystyle\int_{-1}^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$ with a Left-Hand Sum with 20 subintervals.

The answer you got should be close to 0.68. This means that about 68% of the population lies within one standard deviation of the mean. You will revisit this idea later in the course.

(d) Apply the theorem to discover how large the absolute value of the error could be for this particular approximation.

(e) How large would n have to be in order to guarantee that a Left-Hand Sum with n subintervals would approximate the actual integral with an error no larger than 0.0005? (Such an approximation is said to be valid to three decimal places.)

Now we have discovered the power of the theorem. We cannot calculate

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

exactly, but we can approximate this integral to any degree of accuracy that we please by using a Left-Hand Sum with a large enough number of subintervals. What is more, we can deduce from the theorem what "large enough" means for a particular definite integral.

Here are the error bounds for the other methods of approximation.

Theorem: Let f be a function whose second derivative f'' is continuous on (a, b). Also, suppose that $|f''(x)| \leq M_2$ for all x in the interval $[a, b]$. Then E_n, the error used in approximating $\int_a^b f(x)\, dx$ with a Trapezoid Sum with n subintervals, satisfies the following inequality:

$$|E_n| \leq \frac{M_2(b-a)^3}{12n^2}.$$

Theorem: Let f be a function whose second derivative f'' is continuous on (a, b). Also, suppose that $|f''(x)| \leq M_2$ for all x in the interval $[a, b]$. Then E_n, the error used in approximating $\int_a^b f(x)\, dx$ with a Midpoint Sum with n subintervals, satisfies the following inequality:

$$|E_n| \leq \frac{M_2(b-a)^3}{24n^2}.$$

Theorem: Let f be a function whose fourth derivative $f^{(4)}$ is continuous on (a, b). Also, suppose that $|f^{(4)}(x)| \leq M_4$ for all x in the interval $[a, b]$. Then E_n, the error used in approximating $\int_a^b f(x)\, dx$ with Simpson's Rule with n subintervals, satisfies the following inequality:

$$|E_n| \leq \frac{M_4(b-a)^3}{180n^4}.$$

(f) Use the theorem to determine how many subintervals we should use to approximate

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

with a Trapezoid Sum in order to guarantee that the error is at most 0.0005. Do the same for the Midpoint Sum.

(g) Compare these numbers with the number of subintervals you would need to get the same accuracy with Right-Hand and Left-Hand Sums. Arrange the approximation techniques in order from least efficient to most efficient.

(h) We often say that Left-Hand Sums and Right-Hand Sums are *first-order* approximation methods, while Trapezoid Sums and Midpoint Sums are *second-order* approximation methods. Explain how this terminology connects to the error bounds from the theorems. What order approximation technique would you say that Simpson's Rule is?

Problems

1. Approximate

$$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

to within 0.0005 using LHS, RHS, MID, and TRAP. Explain each step carefully, using words and avoiding calculator notation. How many subintervals did you need to choose for each method?

2. Farmer Al wants to plant a field with grass for his horses. He knows that it will take about 30 pounds of fescue seed to plant one acre of land. (One acre is 43,560 square feet.) The seed comes in one pound increments. Farmer Al is very frugal and does not want to spend a nickel more than he has to. He measures how wide the field is at regular intervals. Here is a chart of his measurements. All measurements are in feet.

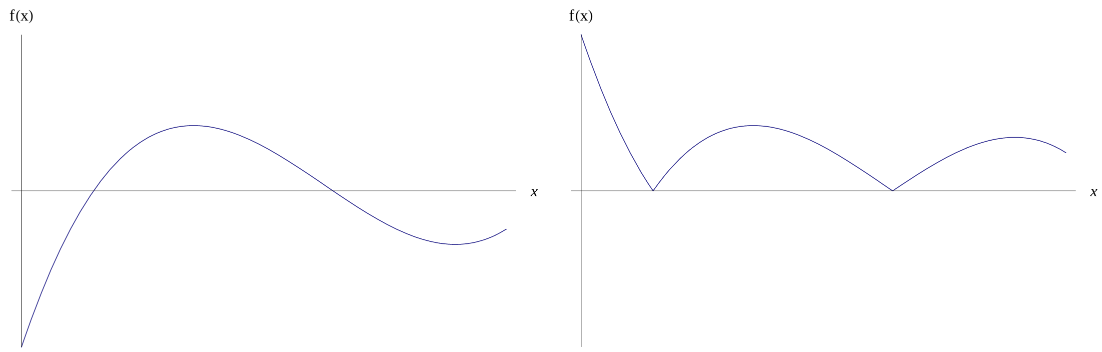
Dist. from barn	0	100	200	300	400	500	600	700	800	900
Width of field	856	873	910	957	1068	1072	1127	1168	1255	1278

Give Farmer Al an upper bound and a lower bound on how many pounds of fescue seed that he needs to buy. What assumption or assumptions should you make about the field in order to guarantee that the values you found actually are upper and lower bounds? Do you think these assumptions are reasonable? Why or why not? If Farmer Al tells you that the range that you gave him is just too large, and that he would do *anything* to obtain a smaller range, what course of action would you recommend to Farmer Al?

3. Estimate the percentage of the population whose IQ falls between 90 and 115 by estimating the definite integral that you wrote earlier in the lab. Explain each step carefully, using words and avoiding calculator notation. What would you consider to be an acceptable amount of error for this estimate? Explain your answer.

Supplemental Section: Proof of the Theorem

Before we start, we will need to borrow an inequality that is associated to integrals. Examine the following pictures.



Using a sentence or two, explain why

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx$$

for a general function f.

As a warmup, let's prove the theorem in the case that $n = 1$. Here, the inequality that we must prove is

$$|E_1| \leq M_1 \frac{(b-a)^2}{2}.$$

(a) What is the Left-Hand Sum Approximation to $\int_a^b f(x)\, dx$ if $n = 1$?

(b) Justify each of the following steps.

Justification

$$\begin{aligned} |E_1| &= \left| \int_a^b f(x)\, dx - f(a)(b-a) \right| \\ &= \left| \int_a^b (f(x) - f(a))\, dx \right| \end{aligned}$$

(c) Now let's work with the integrand. Notice that the hypotheses of this theorem mirror the hypotheses of the Mean Value Theorem. Apply the Mean Value Theorem to f over the interval $[a, x]$. Justify each of the following steps.

		Justification
$\lvert E_1\rvert$	$= \left\lvert \int_a^b (f(x) - f(a))\, dx \right\rvert$	From before
	$= \left\lvert \int_a^b (f'(z)(b-a))\, dx \right\rvert$	
	$\leq \int_a^b \lvert f'(z)(b-a)\rvert\, dx$	
	$\leq \int_a^b M_1(b-a)\, dx$	

(d) Now do the integral that you are left with to finish verifying the result for $n = 1$.

The general case

$$|E_n| \leq M_1 \frac{(b-a)^2}{2n}.$$

is only slightly more complicated.

(a) First, notice that there was nothing special about a and b in the argument that we gave. The argument remains valid even if we replace a with t_k and b with t_{k+1}, where t_k is the left endpoint of the $(k-1)$th subinterval. Explain why in this case our inequality can be written

$$|E_1| \leq \frac{M_1(\Delta t)^2}{2}.$$

(b) For the proof, we will need the triangle inequality, which says

$$|A_1 + A_2 + A_3 + \cdots + A_n| \leq |A_1| + |A_2| + |A_3| + \cdots + |A_n|.$$

Explain why this is true.

(c) Write the triangle inequality using Σ notation.

$$\left| \sum_{k=1}^{n} A_k \right| \leq$$

(d) Notice that we can break our original integral up into little pieces, like this:

$$\int_a^b f(x)\,dx = \int_{t_0}^{t_1} f(x)\,dx + \int_{t_1}^{t_2} f(x)\,dx + \cdots + \int_{t_{n-1}}^{t_n} f(x)\,dx$$

Write the above expression using Σ notation. (Notice that k starts at 1, not 0.)

$$\int_a^b f(x)\,dx = \sum_{k=1}^{n}$$

(e) Fill in the missing Justifications or Steps:

Justification

$$\begin{aligned}
|E_n| &= \left| \int_a^b f(x)\,dx - \sum_{k=1}^{n} f(t_{k-1})\Delta t \right| \\
&= \left| \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(x)\,dx - \sum_{k=1}^{n} f(t_{k-1})\Delta t \right| \\
&= \qquad \text{Combine the summations} \\
&\leq \qquad \text{Triangle inequality} \\
&\leq \sum_{k=1}^{n} \frac{M_1(\Delta t)^2}{2} \\
&= n\frac{M_1(\Delta t)^2}{2}
\end{aligned}$$

(f) Use the formula for Δt to obtain the error bound promised by the theorem.

Similar arguments can be used to obtain error bounds for Trapezoid Sums, Midpoint Sums and Simpson's Rule.