Integration.

Alan Parks, Lawrence University, 1989, revised 1992, 96, 99, 2000.

1. Variation sums.

You have been introduced to the integral through two ideas, combined in the Fundamental Theorem of Calculus: if $f(x)$ is continuous, then

$$(A) \qquad \int_a^b f(x)dx = F(b) - F(a) \qquad \text{where} \qquad F'(x) = f(x)$$

and

$$(B) \qquad \int_a^b f(x)dx = \lim \sum_i f(x_i)\Delta x_i$$

where the x_i are chosen inside subintervals of length Δx_i which divide up the interval $[a, b]$. Equation A is used primarily to compute the values of integrals, and equation B is used to recognize applications – that the integral computes some desired quantity which is approximated by the *Riemann sums* on the right hand side.

Up until the early nineteenth century, the connection between equations A and B was treated informally, with a certain element of vagueness concerning the precise nature of the limit in equation B. This vagueness did not pose a problem; on the contrary, Newton, Leibnitz, Euler and many others used the integral to obtain concrete formulas and theoretical implications that solved longstanding problems of physical and celestial mechanics, geometry and number theory. Equation A seemed to be the main fact, and it was used even (in fact especially!) when the integrand $f(x)$ was a power series. Everything seemed to work perfectly, despite the marginal discomfort felt by some regarding the informal nature of equation B.

But through the work of Euler, (Daniel) Bernoulli, and Fourier on infinite series of trigonometric functions, and that of Cauchy on power series, the defects of the informal approach became more and more apparent. There arose calculations which seemed, on the one hand, to be of the sort that had worked before, but which, on the other hand, yielded results which could not be true. It was clear what was needed: a more formal definition of the integral and a rigorous verification of its properties. Of the various definitions that were suggested during the nineteenth century, that of Riemann seemed to be best, therefore the integral that one uses in calculus is called the *Riemann integral.* Of course this definition involved the sums named after him. It may seem strange to you that there would be any latitude at all in defining the integral, and in a sense there isn't: we certainly want it to be the case that equation A holds. What is more subtle is that equation A does not tell the whole story, and that it is not a good starting place. We will certainly *end up* there!

Darboux discovered a definition of the integral which is equivalent to that of Riemann, but which is easier to explain and to work with. We will now exposit Darboux's definition of the Riemann integral. The key to this approach is to separate the problem of the existence of the integral from the problem of computing its value.

We assume that the reader is familiar with the Completeness Axiom for the reals and with the properties of continuous functions, including the important theorem that a continuous function on a closed interval is uniformly continuous there. Also as usual, although we may use the sine and cosine functions in the construction of examples (and in doing so we will assume their "familiar" properties), we do not assume they are rigorously understood, and so we will not use them in proofs.

We begin, confusingly enough, by defining two terms for the same thing: subdivision and partition. Both of these terms refer to chopping up a closed interval into a finite number of sub-intervals. A *subdivision* of $[a, b]$ is a finite set of points in the interval, which includes the endpoints. The elements of such a set can always be written in order $a = x_0 < x_1 < \ldots < x_n = b$. The *partition* P associated with this subdivision is the finite set of intervals $P = \{ [x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n] \}$. Usually it will be more convenient to work with the partition than the subdivision, but whenever we say, "let P be a partition," we understand that the partition comes from a subdivision. It is easy to see that the sum of the lengths of the intervals in a partition of $[a, b]$ is the length $b - a$ of the whole interval. (In general, we will denote the length of a closed interval I by $|I|$.)

Let $f(x)$ be defined and bounded on $[a, b]$. For a closed interval $I \subseteq [a, b]$, since f is bounded, the set of $|f(x) - f(y)|$ for all $x, y \in I$ is bounded. Define the *variation on I of f*:

$$\text{var}(I, f) = \sup\{ |f(x) - f(y)| \quad \text{such that} \quad x, y \in I\}$$

We will make use of the following line of reasoning. If $I \subseteq J$, then $\text{var}(J, f)$ is an upper bound for the $|f(x) - f(y)|$ for $x, y \in I$. Therefore, the least upper bound $\text{var}(I, f)$ is less than or equal to the upper bound $\text{var}(J, f)$. To repeat: if $I \subseteq J$, then $\text{var}(I, f) \leq \text{var}(J, f)$.

Let $f : [a, b] \to \mathbb{R}$ be bounded, and let P be a partition of $[a, b]$. Define the *variation sum*

$$\Sigma_V(P, f) = \sum_{I \in P} \text{var}(I, f) \cdot |I|$$

In class we will draw a picture of this sum – it looks like the sum of the areas of rectangles that "follow the curve $y = f(x)$ along;" one rectangle for each sub-interval I in the partition P.

We say that f is *integrable on* $[a, b]$ if the variation sums can be made arbitrarily small. In other words, if $\epsilon > 0$, there is a partition P of $[a, b]$, such that $\Sigma_V(P, f) < \epsilon$. Later, we will see that the integral $\int_a^b f$ is defined when f is an integrable function; as we mentioned before, it turns out to be a good idea not to consider the value of the integral at first. We will keep our focus on the variation sums and their properties for now.

To take a simple example, let $f(x) = x^2$ on $[0, 1]$ and consider the partition P_n of $[0, 1]$ into n equal sub-intervals. The sub-division consists of points j/n for $0 \leq j \leq n$, and the intervals are

$$\left[\frac{j-1}{n}, \frac{j}{n}\right] \qquad \text{for} \qquad 1 \leq j \leq n$$

On a particular sub-interval $[c, d]$, the function x^2 has a minimum at $x = c$ and a maximum at $x = d$. It follows that

$$\text{var}\left(\left[\frac{j-1}{n}, \frac{j}{n}\right], x^2\right) = \left(\frac{j}{n}\right)^2 - \left(\frac{j-1}{n}\right)^2$$

The width of each sub-interval is $1/n$. Now we can compute the variation sum:

$$\Sigma_V(P_n, x^2) = \sum_{j=1}^{n} \text{var}\left(\left[\frac{j-1}{n}, \frac{j}{n}\right], x^2\right) \cdot \frac{1}{n} = \sum_{j=1}^{n} \left[\left(\frac{j}{n}\right)^2 - \left(\frac{j-1}{n}\right)^2\right] \cdot \frac{1}{n}$$

When the sum on the right is expanded out, all the terms cancel except for two of them:

$$\Sigma_V(P_n, x^2) = \frac{1}{n}$$

We see that this can be made arbitrarily small by making n large. We conclude that x^2 is integrable on $[0, 1]$.

In the next section, we will prove that a wide variety of functions are integrable. We will need to understand the relationship between variation sums for two different partitions. We say that the partition R *refines* the partition P if each subdivision in P is a subdivision in R. In other words, suppose that the subdivision of P is

$$a = x_0 < x_1 < \ldots < x_n = b$$

Then the x_i are part of (maybe all of) the subdivision of R.

Note that Proposition 1.1 does not assume that $f(x)$ is integrable, only that it is bounded.

Proposition 1.1. *Let $f : [a,b] \to \mathbb{R}$ be bounded. If P and R are partitions of $[a,b]$, and if R refines P, then* var$(R,f) \leq$ var(P,f)

Proof. Given one of the intervals $I \in P$, the intervals in R that are contained in I make up a partition of I. Let R_I be this set of intervals; to repeat, R_I is the set of all $J \in R$ such that $J \subseteq I$. The sum of $|J|$ over all $J \in R_I$ is $|I|$. Furthermore, each $J \in R$ is in R_I for a unique element I of P. In light of this, we see that

$$\Sigma_V(R,f) = \sum_{J \in R} \text{var}(J,f) \cdot |J| = \sum_{I \in P} \sum_{J \in R_I} \text{var}(J,f) \cdot |J| \tag{1.1}$$

For $I \in P$ and $J \in R_I$ we have $J \subseteq I$, and so, as we remarked above, we have $\text{var}(J,f) \leq \text{var}(I,f)$. We use this to continue the calculation (1.1):

$$\sum_{I \in P} \sum_{J \in R_I} \text{var}(J,f) \cdot |J| \leq \sum_{I \in P} \sum_{J \in R_I} \text{var}(I,f) \cdot |J| \tag{1.2}$$

Since $\text{var}(I,f)$ does not depend on J, we can factor it out:

$$\sum_{I \in P} \sum_{J \in R_I} \text{var}(I,f) \cdot |J| = \sum_{I \in P} \text{var}(I,f) \cdot \sum_{J \in R_I} |J| \tag{1.3}$$

The sum of $|J|$ over $J \in R_I$ is $|I|$, and so

$$\sum_{I \in P} \text{var}(I,f) \cdot \sum_{J \in R_I} |J| = \sum_{I \in P} \text{var}(I,f) \cdot |I| = \Sigma_V(P,f)$$

Putting (1.1) and (1.2) and (1.3) together with this last equality, we obtain $\Sigma_V(R,f) \leq \Sigma_V(P,f)$. QED

We will need that fact that if P and Q are given partitions, then there is a partition R that refines both of them. This is obvious: for instance, let the sub-division for R be the union of the sub-divisions for P and Q.

2. Integrable functions.

We want to identify a large class of integrable functions. We will stop short of a full characterization of such functions.

Theorem 2.1. *Let $f(x)$ be continuous on $[a,b]$. Then $f(x)$ is integrable on $[a,b]$.*

Proof. Continuous functions are bounded, and so $f(x)$ is bounded on $[a,b]$. Let $\epsilon > 0$. Since $f(x)$ is uniformly continuous, so that there is $\delta > 0$ such that $|x-y| < \delta$ for $x, y \in [a,b]$ implies that $|f(x)-f(y)| < \epsilon$.

We will use this as follows. Let I be an interval contained in $[a,b]$ with $|I| < \delta$. If $x, y \in I$, then $|x-y| < \delta$, so that $|f(x) - f(y)| < \epsilon$. This shows that $\text{var}(I,f) < \epsilon$.

Let P be a partition of $[a,b]$ such that each $I \in P$ has $|I| < \delta$. Then

$$\Sigma_V(P,f) = \sum_{I \in P} \text{var}(I,f) \cdot |I| \leq \sum_{I \in P} \epsilon \cdot |I| = \epsilon \cdot (b-a)$$

The number $\epsilon \cdot (b-a)$ can be made arbitrarily small. QED

We point out some obvious consequences. Every polynomial is integrable over every closed interval. Every rational or algebraic function is integrable over every closed interval inside its domain. Every function differentiable on a closed interval is integrable there.

We can generalize Theorem 2.1 to include some discontinuities.

Theorem 2.2. *Let $f : [a, b] \to \mathbb{R}$ be bounded and continuous except possibly at finitely many points. Then $f(x)$ is integrable on $[a, b]$.*

Proof. The proof is like that of Theorem 2.1 except that we have to step around the discontinuities. Here goes. Let $a_1, a_2, \ldots, a_n$ be the discontinuity points of $f(x)$ on $[a, b]$. Choose t small enough so that the intervals $[a_i - t, a_i + t]$ do not overlap.

Define D to be the set of points which are in $[a, b]$ but not in any of the *open* intervals $(a_i - t, a_i + t)$. Then D consists of finitely many (at most $n + 1$) closed intervals. Let $\epsilon > 0$, and on each of the closed sub-intervals I of D, there is a $\delta > 0$ such that if $|x - y| < \delta$ for $x, y \in I$, then $|f(x) - f(y)| < \epsilon$. A sufficiently small δ will have this property for all of the sub-intervals.

Let P be a partition of $[a, b]$ using the intervals $[a_i - t, a_i + t]$ for $1 \leq i \leq n$, as well as other intervals, making sure that each of the other intervals has length no greater than δ. Let G name these other intervals. Thus if $I \in P$ then either $I = [a_i - t, a_i + t]$ for some i, or $I \in G$. If $I \in G$, and if $x, y \in I$, then $|f(x) - f(y)| < \epsilon$; thus, $\text{var}(I, f) < \epsilon$.

Let $B = \text{var}([a, b], f)$, and then $|f(x) - f(y)| \leq B$ for all $x, y \in [a, b]$. Compute

$$\Sigma_V(P, f) = \sum_{I \in P} \text{var}(I, f) \cdot |I| = \sum_{i=1}^{n} \text{var}([a_i - t, a_i + t], f) \cdot 2t \quad + \quad \sum_{I \in G} \text{var}(I, f) \cdot |I|$$

For each i the quantity $\text{var}([a_i - t, a_i + t], f)$ is no greater than B. In the other sum, the $\text{var}(I, f)$ are no greater than ϵ. Making these replacements we estimate

$$\Sigma_V(P, f) \leq \sum_{i=1}^{n} B \cdot 2t + \sum_{I \in G} \epsilon \cdot |I| \leq B \cdot 2t \cdot n + \epsilon \cdot (b - a)$$

Because a, b, B, and n are constants, this last expression can be made arbitrarily small by making t and ϵ small.
QED

Theorem 2.2 applies to such functions as

$$f(x) = \begin{cases} \sin(x^{-1}) & \text{when } x \neq 0 \\ 0 & \text{when } x = 0 \end{cases}$$

From a closed interval $[a, b]$, choose a sequence c_k for $k \geq 1$, with all the c_k different from each other. Now let d_k be a sequence which converges to 0 (for example, we could have $d_k = 1/k$). Define

$$f(x) = \begin{cases} d_k & \text{if } x = c_k \\ 0 & \text{otherwise} \end{cases}$$

In class we will show that f is integrable, but it has infinitely many discontinuities. On the other hand, the function

$$g(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

has infinitely many discontinuities and is not integrable. Thus, the case of infinitely many discontinuities is ambiguous.

In our next result continuity is not mentioned at all. It features the same collapsing sum that occurred when we computed the variation sum for x^2 on $[0, 1]$.

Theorem 2.3. *Let $f(x)$ be defined and monotonic on $[a, b]$. Then $f(x)$ is integrable on $[a, b]$.*

Proof. That $f(x)$ is monotonic shows that $f(a)$ and $f(b)$ are bounds on its values, thus it is bounded. For the rest of the proof assume that $f(x)$ is non-decreasing (the case of non-increasing will be seen to be similar). Given a sub-interval $[c, d]$ contained in $[a, b]$, observe that $x, y \in [c, d]$ implies that $f(x) \le f(d)$ and $f(y) \ge f(c)$, so that $f(x) - f(y) \le f(d) - f(c)$. This shows that $f(d) - f(c) = \text{var}([c, d], f)$.

Let P be the partition $[a, b]$ coming from a sub-division x_j (for $0 \le j \le n$) with the x_j equally spaced. Each interval in P has width $(b - a)/n$. In light of all this, the variation sum collapses, as happened for x^2 in the example above:

$$\Sigma_V(P, f) = \sum_{j=1}^{n} (f(x_j) - f(x_{j-1}))\frac{b-a}{n} = (f(b) - f(a))\frac{b-a}{n}$$

This sum can be made arbitrarily small by making n large. QED

Next we show that the sum of integrable functions is integrable. You also know that the sum of integrals is the integral of the sum; this result will surface as an easy consequence of our work in the next section.

Proposition 2.4. *Let $f(x)$ and $g(x)$ be integrable on $[a, b]$. Then $f(x) + g(x)$ is integrable there.*

Proof. We begin with a line of reasoning that we will use repeatedly. Choose $\epsilon > 0$. Since $f(x)$ is integrable, there is a partition P such that $\Sigma_V(P, f) < \epsilon$, and similarly that g is integrable produces a partition Q such that $\Sigma_V(Q, g) < \epsilon$. Let R be a partition that refines both P and Q. By Proposition 1.1, we have $\Sigma_V(R, f) \le \Sigma_V(P, f)$ and that $\Sigma_V(R, g) \le \Sigma_V(Q, g)$. Thus $\Sigma_V(R, f) < \epsilon$ and $\Sigma_V(R, g) < \epsilon$. We can forget P and Q and work solely with R.

Let $I \in R$. For $x, y \in I$, we have

$$\begin{aligned} |(f(x) + g(x)) - (f(y) + g(y))| &= |f(x) - f(y) + g(x) - g(y)| \le |f(x) - f(y)| + |g(x) - g(y)| \\ &\le \text{var}(I, f) + \text{var}(I, g) \end{aligned}$$

It follows that

$$\text{var}(I, f + g) \le \text{var}(I, f) + \text{var}(I, g)$$

Then

$$\begin{aligned} \Sigma_V(R, f + g) &= \sum_{I \in R} \text{var}(I, f + g) \cdot |I| \le \sum_{I \in R} (\text{var}(I, f) + \text{var}(I, g)) \cdot |I| \\ &= \sum_{I \in R} \text{var}(I, f) \cdot |I| + \sum_{I \in R} \text{var}(I, g) \cdot |I| \\ &= \Sigma_V(R, f) + \Sigma_V(R, g) < \epsilon + \epsilon \end{aligned}$$

We see that $\Sigma_V(P, f + g)$ can be made arbitrarily small. QED

Our next result is a little deeper, having to do with composite functions. First of all, we mention that it is possible to have $f(x)$ integrable and $g(x)$ integrable without $f(g(x))$ being integrable (in fact, g can even be continuous!). However, when f is continuous and g is integrable, everything is fine.

Proposition 2.5. *Let $g : [a, b] \to [c, d]$ be integrable, and suppose that $f : [c, d] \to \mathbb{R}$ is continuous. Then $f(g(x))$ is integrable on $[a, b]$.*

Proof. Since f is continuous, it is bounded on $[c, d]$. If $x \in [a, b]$, then $g(x) \in [c, d]$, and so $f(g(x))$ is between the bounds of f on $[c, d]$. This proves that $f(g(x))$ is bounded on $[a, b]$. Let $A = \text{var}([a, b], f(g))$.

Let $\epsilon > 0$. Since f is uniformly continuous on $[c, d]$, there is $\delta > 0$ such that if $r, s \in [c, d]$ and $|r - s| < \delta$, then $|f(r) - f(s)| < \epsilon$.

Now we need a partition P of $[a, b]$ on which to estimate $\Sigma_V(P, f(g))$. For reasons that you will observe as we do the estimate, we want a partition P such that $\Sigma_V(P, g) < \delta\epsilon$ (observe that this involves g, not $f(g)$). This partition exists because g is integrable.

We will estimate $\Sigma_V(P, f(g))$ by splitting P up into two pieces. Let

$$Q = \{I \in P \mid \text{var}(I, g) \geq \delta\}$$

For $I \in Q$, we have $\text{var}(I, g) \geq \delta$, so that, in particular $\text{var}(I, g) \neq 0$. Then also,

$$\text{var}(I, f(g)) = \frac{\text{var}(I, f(g))}{\text{var}(I, g)}\text{var}(I, g)$$

We will need to estimate this. We have

$$\text{var}(I, f(g)) \leq A \quad \text{and} \quad \text{var}(I, g) \geq \delta$$

so that, since $\text{var}(I, g) \geq 0$, we have

$$\frac{\text{var}(I, f(g))}{\text{var}(I, g)}\text{var}(I, g) \leq \frac{A}{\delta}\text{var}(I, g)$$

Now consider $I \in P$ with $I \notin Q$. Then $\text{var}(I, g) < \delta$. If $r, s \in I$, then we have $|g(r) - g(s)| < \delta$. Recall how δ was defined from ϵ and the continuity of f. We have

$$|f(g(r)) - f(g(s))| < \epsilon \quad \text{for all} \quad r, s \in I$$

It follows that $\text{var}(I, f(g)) < \epsilon$ for all $I \in P$ with $I \notin Q$.

We are ready to estimate:

$$\begin{aligned}
\Sigma_V(P, f(g)) &= \sum_{I \in Q} \text{var}(I, f(g)) \cdot |I| + \sum_{I \notin Q} \text{var}(I, f(g)) \cdot |I| \\
&\leq \sum_{I \in Q} \frac{A}{\delta}\text{var}(I, g) \cdot |I| + \sum_{I \notin Q} \epsilon \cdot |I| \\
&\leq \frac{A}{\delta}\Sigma_V(P, g) + \epsilon \sum_{I \notin Q} |I| < \frac{A}{\delta}\delta\epsilon + \epsilon(b - a) = (A + b - a)\epsilon
\end{aligned}$$

The numbers A, a, b are constants. Thus, the variation sums for $f(g)$ can be made arbitrarily small, so that $f(g)$ is integrable. QED

Proposition 2.5 can be used cleverly to establish some other expected results.

Proposition 2.6.
a) *If $f(x)$ is integrable on $[a, b]$ and if k is a constant, then $k \cdot f(x)$ is integrable on $[a, b]$.*
b) *If $f(x)$ is integrable on $[a, b]$, then so is $|f(x)|$.*
c) *If $f(x)$ and $g(x)$ are integrable on $[a, b]$, then so is $f(x) \cdot g(x)$.*

Proof. The function $k \cdot f(x)$ is the composite of the integrable function $f(x)$ with the continuous function $k \cdot x$.

The identify $|f(x)| = \sqrt{f(x)^2}$ shows that $|f(x)|$ is the composite of the function f with the continuous square and square root functions.

The identity $2 \cdot f(x) \cdot g(x) = (f(x) + g(x))^2 - f(x)^2 - g(x)^2$ can be used with Proposition 2.4, 2.5, and statement (a) of the present proposition to show that $f(x) \cdot g(x)$ is integrable if f and g are. QED

Next we show how being integrable on a closed interval is related to being integrable on sub-intervals. You will expect an equation relating the three integrals in the following; that identity will be proved in the next section.

Proposition 2.7. *Let $a \le b \le c$. Then the function $f(x)$ is integrable on $[a, c]$ if and only if it is integrable on $[a, b]$ and integrable on $[b, c]$.*

Proof. Let $f(x)$ be integrable on $[a, c]$ (the whole interval). Choose $\epsilon > 0$ and let P be a partition of $[a, c]$ such that $\Sigma_V(P, f) < \epsilon$. Let R be a partition of $[a, c]$ that refines P and the partition $\{[a, b], [b, c]\}$. In particular, R can be broken into two subsets Q and T such that Q is a partition of $[a, b]$ and T is a partition of $[b, c]$. Observe that

$$\Sigma_V(R, f) = \Sigma_V(Q, f) + \Sigma_V(T, f)$$

and since the three numbers in this equation are non-negative, we also have

$$\Sigma_V(R, f) \ge \Sigma_V(Q, f) \quad \text{and} \quad \Sigma_V(R, f) \ge \Sigma_V(T, f) \tag{2.1}$$

Since R refines P, Proposition 1.1 shows that $\Sigma_V(P, f) \ge \Sigma_V(R, f)$ so since $\Sigma_V(P, f) < \epsilon$ we see that $\Sigma_V(R, f) > \epsilon$. Then the inequalities (2.1) then conclude that $\Sigma_V(Q, f) < \epsilon$ and $\Sigma_V(T, f) < \epsilon$. That ϵ is arbitrary now proves that $f(x)$ is integrable on $[a, b]$ and on $[b, c]$.

Conversely, if $f(x)$ is integrable on $[a, b]$ and on $[b, c]$, then there are partitions Q of $[a, b]$ and T of $[b, c]$ such that $\Sigma_V(Q, f) < \epsilon$ and $\Sigma_V(T, f) < \epsilon$ where ϵ is a given positive number. Let $R = Q \cup T$ and then R is a partition of $[a, c]$ for which $\Sigma_V(R, f) = \Sigma_V(Q, f) + \Sigma_V(T, f) < 2\epsilon$. QED

We close this section by showing that if an integrable function is "messed up" at finitely many points, the resulting function is still integrable. Later, we will discover the surprising fact that the messing up does not change the value of the integral.

Proposition 2.8. *Let $f(x)$ be integrable on $[a, b]$, and let $g(x)$ be defined on $[a, b]$ with $f(x) = g(x)$ except at finitely many points. Then $g(x)$ is integrable on $[a, b]$.*

Proof. Let $a_1, a_2, \dots, a_n$ be the points at which f and g disagree. Put $h(x) = g(x) - f(x)$ so that $h(x)$ is zero except at the a_i. In particular, $h(x)$ is obviously bounded, and it is continuous except at finitely many points. By Theorem 2.2 the function h is integrable. Since $f(x)$ and $h(x)$ are integrable, Proposition 2.4 shows that $g(x) = h(x) + f(x)$ is integrable. QED

3. Riemann sums and the definition of the integral.

Suppose that $f : [a, b] \to \mathbb{R}$ and that P is a partition of $[a, b]$. If for each $I \in P$, we choose $x_I \in I$, then a *Riemann sum for f on $[a, b]$ using P* results, having the form

$$\sum_{I \in P} f(x_I) \cdot |I| \tag{3.1}$$

Riemann sums link integration to its many applications, and we will use them to give a definition of the integral.

We need to introduce two sums that look somewhat like the sum in (3.1). Let $f : [a, b] \to \mathbb{R}$ be bounded, and let P be a partition of $[a, b]$. For $I \in P$, define

$$\sup(I, f) = \sup\{f(x) \mid x \in I\} \qquad \text{and} \qquad \Sigma^u(P, f) = \sum_{I \in P} \sup(I, f) \cdot |I|$$

This is called an *upper Riemann sum for f*. As with the variation sum, we will draw a picture suggesting what this looks like.

Similarly, define

$$\inf(I,f) = \inf\{f(x) \mid x \in I\} \qquad \text{and} \qquad \Sigma_l(P,f) = \sum_{I\in P} \inf(I,f)\cdot |I|$$

This is called a *lower Riemann sum for f*.

If, as in (3.1), we choose $x_I \in I$ for each $I \in P$, then we see that $\inf(I,f) \le f(x_I) \le \sup(I,f)$, and it follows that

$$\Sigma_l(P,f) \le \sum_{I\in P} f(x_I)\cdot |I| \le \Sigma^u(P,f) \tag{3.2}$$

The inequalities in (3.2) say that every Riemann sum for f lies between the lower and upper Riemann sum.

We need to link the upper and lower Riemann sums to the variation sums considered before.

Proposition 3.1. *Let $f : [a,b] \to \mathbb{R}$ be bounded and let P be a partition of $[a,b]$. Then $\Sigma^u(P,f) - \Sigma_l(P,f) \le \Sigma_V(P,f)$.*

Proof. Let $I \in P$ and let $\epsilon > 0$. Then there is $x \in I$ with $\sup(I,f) - \epsilon < f(x)$. Also, there is $y \in I$ with $f(y) < \inf(I,f) + \epsilon$. Putting this together:

$$f(x) - f(y) > \sup(I,f) - \epsilon - (\inf(I,f) + \epsilon) = \sup(I,f) - \inf(I,f) - 2\epsilon$$

Recalling the definition of $\mathrm{var}(I,f)$, we see that

$$\mathrm{var}(I,f) \ge |f(x) - f(y)| > \sup(I,f) - \inf(I,f) - 2\epsilon$$

Since ϵ is arbitrary, this proves that

$$\mathrm{var}(I,f) \ge \sup(I,f) - \inf(I,f) \tag{3.3}$$

Employ (3.3) in the following:

$$\begin{aligned}\Sigma_V(P,f) &= \sum_{I\in P} \mathrm{var}(I,f)\cdot|I| \ge \sum_{I\in P}(\sup(I,f) - \inf(I,f))\cdot|I| \\ &= \sum_{I\in P}\sup(I,f)\cdot|I| - \sum_{I\in P}\inf(I,f)\cdot|I| = \Sigma^u(P,f) - \Sigma_l(P,f)\end{aligned}$$

as needed. QED

We cannot help mentioning that it is in fact true that the variation sum is *equal* to the difference of the upper and lower sums; we will not need this fact, however. We need one more auxiliary fact, explaining what happens to lower and upper sums when a partition is refined.

Proposition 3.2. *Let $f : [a,b] \to \mathbb{R}$ be bounded and suppose that P and R are partitions of $[a,b]$ with R refining P. Then*

$$\Sigma_l(P,f) \le \Sigma_l(R,f) \le \Sigma^u(R,f) \le \Sigma^u(P,f)$$

Proof. For each $I \in P$, define R_I to be the set of $J \in R$ with $J \subseteq I$. In this setting, if $x \in J$, then $x \in I$ and so $f(x) \leq \sup(I, f)$. In other words, $\sup(I, f)$ is an upper bound for $f(x)$ with $x \in J$. Therefore, $\sup(J, f) \leq \sup(I, f)$. Now compute as in the proof of Proposition 1.1,

$$\begin{aligned}\Sigma^u(R, f) &= \sum_{J \in R} \sup(J, f) \cdot |J| = \sum_{I \in P} \sum_{J \in R_I} \sup(J, f) \cdot |J| \\ &\leq \sum_{I \in P} \sum_{J \in R_I} \sup(I, f) \cdot |J| \\ &= \sum_{I \in P} \sup(I, f) \cdot \sum_{J \in R_I} |J| \\ &= \sum_{I \in P} \sup(I, f) \cdot |I| = \Sigma^u(P, f)\end{aligned}$$

and this proves that $\Sigma^u(R, f) \leq \Sigma^u(P, f)$.

The proof that $\Sigma_l(P, f) \leq \Sigma_l(R, f)$ is similar and will be done either in class or as a homework problem.

It is obvious that $\Sigma_l(R, f) \leq \Sigma^u(R, f)$. QED

We are ready for the definition of the integral. The following is the main theorem of these notes.

Theorem 3.3. *Let $f : [a, b] \to \mathbb{R}$ be integrable. Then there is a unique number $\int_a^b f$ such that*

$$\Sigma_l(P, f) \leq \int_a^b f \leq \Sigma^u(P, f)$$

for all partitions P of $[a, b]$.

Several remarks are in order. Of course the notation $\int_a^b f$ is the *integral of f on* $[a, b]$. You may wonder where the dx is; we will talk about this later, for now we can safely live without it. Moreover, there is no necessity of using a variable name, writing $f(x)$ instead of just f, since the integral depends on f as a function rather than on the name of its variable.

The integral notation goes back to Leibnitz, who imagined the area bounded by a closed interval $[a, b]$ on the x-axis, the lines $x = a$ and $x = b$, and a curve $y = f(x)$ above the x-axis. Imagine dividing the area up into *infinitely many, infinitely thin* vertical rectangles, each of width dx (where the differential is idealized as being an "infinitely thin" width). The foregoing is not formal, but it has proved to be very helpful in pointing the way to formal ideas. The "height" of an infinitely thin rectangle at the point x would be $f(x)$, and so the total area would be the "infinite sum" of all rectangle-areas $f(x) \cdot dx$. Leibnitz used $\int_a^b$ to denote the sum "over all x in $[a, b]$." The integral sign is a German "s," the first letter in the German word summe (sum). Thus, the notation $\int_a^b f(x) \cdot dx$ was originally meant to denote the sum of "all" products $f(x)$ with infinitely thin widths dx as x ranges between a and b.

In the nineteenth century, when many mathematicians were engaged in putting the facts of analysis on a firm footing, Riemann saw a way to replace Leibnitz' "infinite sum" by the limit of ordinary sums. Riemann's approach involves the "Riemann sums" named for him. Riemann sums make it fairly easy to establish the elementary properties of the integral, and that is why we use them. In calculus courses, it is customary to speak of Riemann sums "approaching the integral in the limit" as the partition gets finer and finer. We will see that Theorem 3.3 is actually more convenient, especially for applications.

While we are on the subject, we mention that Riemann was not just interested in giving a definition of the ordinary calculus integral of one variable, but in generalizing the integral to other settings. Of the many attempts at defining integration in general, that of Lebesgue seems to have worked the best. At the advanced level in analysis, one studies *Lebesgue integration.*

Proof of Theorem 3.3: Let P be a partition of $[a, b]$, fixed for the moment. Consider an arbitrary partition Q of $[a, b]$. There is a partition R that refines both P and Q, and we intend to apply Proposition 3.2 to R and P and Q in turn. Since R refines Q, we have $\Sigma_l(Q, f) \le \Sigma_l(R, f) \le \Sigma^u(R, f)$. Since R refines P, we have $\Sigma^u(R, f) \le \Sigma^u(P, f)$. Putting these together:

$$\Sigma_l(Q, f) \le \Sigma_l(R, f) \le \Sigma^u(R, f) \le \Sigma^u(P, f)$$

and we conclude that $\Sigma_l(Q, f) \le \Sigma^u(P, f)$. Since Q is arbitrary, this shows that $\Sigma^u(P, f)$ is an upper bound for all lower Riemann sums. Therefore, the set of lower Riemann sums is bounded above, and thus it has a sup q. (We will see momentarily that q is the integral!) Because $\Sigma^u(P, f)$ is an upper bound for the lower sums, we have $q \le \Sigma^u(P, f)$.

The quantity q was defined independent of P. Thus, the inequality $q \le \Sigma^u(P, f)$ holds for all partitions P. Since we also have $\Sigma_l(P, f) \le q$, we see that the number q is between the lower and upper sums of every partition.

To finish the proof, it remains to prove that there is only one number between all the lower and upper Riemann sums. Let q, r be between these sums and choose $\epsilon > 0$. Because f is integrable, there is a partition P of $[a, b]$ such that $\Sigma_V(P, f) < \epsilon$. By Proposition 3.1 we have $\Sigma^u(P, f) - \Sigma_l(P, f) \le \Sigma_V(P, f)$, and so $\Sigma^u(P, f) - \Sigma_l(P, f) < \epsilon$. Since q and r are both between $\Sigma_l(P, f)$ and $\Sigma^u(P, f)$, we see that q and r are less than ϵ apart. Since ϵ is arbitrary, this proves that $q = r$. QED

Theorem 3.3 can be used to estimate the difference between a Riemann sum and the integral. The following corollary contains two useful versions of such estimates, the second of which shows how to compute the integral as a limit.

Corollary 3.4. *Let f be integrable on $[a, b]$.*
a) *If $\epsilon > 0$, and if P is a partition of $[a, b]$ with $\Sigma_V(P, f) < \epsilon$, and if R is a Riemann sum for f on $[a, b]$ using P, then R and $\int_a^b f$ are less than ϵ apart.*
b) *If P_n is a sequence of partitions of $[a, b]$ such that $\Sigma_V(P_n, f) \to 0$ as $n \to \infty$, and if R_n is a Riemann sums for f on $[a, b]$ using P_n, then $R_n \to \int_a^b f$ as $n \to \infty$.*

Proof. For (a), note that the Riemann sum R is between the upper and lower Riemann sums for the partition P, as is the integral. Thus, the integral and R are no more than $\Sigma^u(P, f) - \Sigma_l(P, f)$ apart. By Proposition 3.1, this is less than or equal to $\Sigma_V(P, f) < \epsilon$.

Statement (b) follows easily. QED

Armed with Theorem 3.3 and Corollary 3.4, we can go back to some of our results in section 2 and get the expected integral formulas.

Theorem 3.5.
a) *If f, g are integrable on $[a, b]$, then $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.*
b) *If f is integrable on $[a, b]$ and k is a constant, then $\int_a^b (kf) = k \int_a^b f$.*
c) *If f is integrable on $[a, c]$ and if $a \le b \le c$, then $\int_a^c f = \int_a^b f + \int_b^c f$.*
d) *If f is integrable on $[a, b]$ and if $g : [a, b] \to \mathbb{R}$ is equal to f except at finitely many points, then $\int_a^b f = \int_a^b g$.*

Proof. These arguments are all similar, each using Riemann sums to approximate the various integrals. For (a), let $\epsilon > 0$. There is a partition P of $[a, b]$ such that $\Sigma_V(P, f) < \epsilon$ and $\Sigma_V(P, g) < \epsilon$ and $\Sigma_V(P, f + g) < \epsilon$. Choose $x_I \in I$ for each $I \in P$. By Corollary 3.4, we have that

$$\left| \int_a^b f - \sum_{I \in P} f(x_I) \cdot |I| \right| < \epsilon \quad \text{and} \quad \left| \int_a^b g - \sum_{I \in P} g(x_I) \cdot |I| \right| < \epsilon$$

and

$$\left|\sum_{I\in P}(f(x_I)+g(x_I))\cdot|I|-\int_a^b(f+g)\right|<\epsilon$$

Also observe that

$$\sum_{I\in P}f(x_I)\cdot|I|+\sum_{I\in P}g(x_I)\cdot|I|=\sum_{I\in P}(f(x_I)+g(x_I))\cdot|I|$$

It follows that

$$\left|\int_a^b f+\int_a^b g-\int_a^b(f+g)\right|<3\epsilon$$

and since ϵ is arbitrary, this establishes the desired equality.

We will be more sketchy about the rest of the conclusions, discussing the details in class. For (b), let $\epsilon>0$ and get a partition P such that $\Sigma_V(P,f)<\epsilon$ and $\Sigma_V(P,kf)<\epsilon$. A Riemann sum for kf is k times a Riemann sum for f, and (b) follows.

For (c), get a partition R for $[a,b]$ such that $\Sigma_V(R,f)<\epsilon$, and get a partition S for $[b,c]$ such that $\Sigma_V(S,f)<\epsilon$. Then $R\cup S$ is a partition of $[a,c]$ and $\Sigma_V(R\cup S,f)<2\epsilon$. The sum of a Riemann sum using R and a Riemann sum using S is a Riemann sum using $R\cup S$.

In the setting of (d), get a partition of $[a,b]$ such that $\Sigma_V(P,f)<\epsilon$ and $\Sigma_V(P,g)<\epsilon$. Because f and g disagree at only finitely many points, we can choose a Riemann sum R for f using P, such that $f(x_I)=g(x_I)$ for all chosen x_I. Then R is a Riemann sum for g as well. QED

We pause to call attention to Theorem 3.5d, which says that the integral is not affected by changing finitely many points. This seems to say that the integral computes some sort of "cumulative" or "average" behavior rather than a point by point quantity.

Our next result says that integration preserves inequalities. The inequality 3.6c is called the *triangle inequality.*

Proposition 3.6.

a) *Let h be integrable and non-negative on $[a,b]$. Then $\int_a^b h\ge 0$.*

b) *Let f,g be integrable on $[a,b]$ and suppose that $f(x)\le g(x)$ for all $x\in[a,b]$. Then $\int_a^b f\le\int_a^b g$.*

c) *Let f be integrable on $[a,b]$. Then $|\int_a^b f|\le\int_a^b|f|$.*

Proof. For (a), observe that all Riemann sums for h on $[a,b]$ are non-negative.

For (b), the function $g(x)-f(x)$ is non-negative, and so its integral is non-negative by (a). Since

$$0\le\int_a^b(g-f)=\int_a^b g-\int_a^b f$$

(by Theorem 3.5a,b), we see that $\int_a^b g\ge\int_a^b f$.

For (c), let $s=\pm1$ be chosen so that $s\cdot\int_a^b f=|\int_a^b f|$. Then $s\cdot f\le|f|$, and so by (b) we have

$$\int_a^b s\cdot f\le\int_a^b|f|$$

QED

Books on analysis and calculus books tend to give a formulation of the integral different from that of these notes. They involve not the variation sum of a partition but its *mesh* – the maximum width of one of its sub-intervals. We now prove that if the mesh of a partition is small enough, then its variation sum is also small. Although we will not need this result, it might help you understand the approach of other treatises on integration.

Proposition 3.7. *Let f be integrable on $[a,b]$. For all $\epsilon > 0$, there is $\delta > 0$ such that if P is a partition of $[a,b]$ having mesh less than δ, then $\Sigma_V(P,f) < \epsilon$.*

Proof. Because f is integrable, there is a partition Q of $[a,b]$ such that $\Sigma_V(Q,f) < \epsilon/2$. (The use of $\epsilon/2$ rather than ϵ will be elucidated shortly.) Suppose that the sub-division that defines Q has n elements. Let $A = \mathrm{var}([a,b],f)$, and let $\delta < \epsilon/(2nA)$.

Let P be a partition with mesh less than δ. We divide P into two sets: P_1 consists of $I \in P$ such that $I \subseteq J$ for some $J \in Q$, and P_2 consists of the remaining elements of P, if any. It is easy to see that

$$\sum_{I\in P_1} \mathrm{var}(I,f)\cdot|I| \le \sum_{J\in Q} \mathrm{var}(J,f)\cdot|J| = \Sigma_V(Q,f) < \epsilon/2 \tag{3.4}$$

If $I \in P_2$, then there is an endpoint of Q internal to I; it follows that P_2 contains at most n elements, and we can estimate

$$\sum_{I\in P_2} \mathrm{var}(I,f)\cdot|I| \le nA\delta < \epsilon/2 \tag{3.5}$$

Putting (3.4) and (3.5) together, we obtain

$$\Sigma_V(P,f) = \sum_{I\in P} \mathrm{var}(I,f)\cdot|I| = \sum_{I\in P_1} \mathrm{var}(I,f)\cdot|I| + \sum_{I\in P_2} \mathrm{var}(I,f)\cdot|I| < \epsilon/2 + \epsilon/2 = \epsilon$$

as needed. QED

4. The fundamental theorem of calculus.

We are ready to consider the link between integrals and antiderivatives. The precise nature of this link was discovered by both Newton and Leibnitz – in two rather different ways. We have already hinted at Leibnitz' notion of the integral as an infinite sum; Newton's idea may be understood as follows.

If $f(x)$ is integrable on $[a,b]$ and if $t \in [a,b]$, then by Proposition 2.7 the function $f(x)$ is integrable on $[a,t]$. Newton thought up the definition

$$F(t) = \int_a^t f(x) \qquad \text{for} \quad t \in [a,b]$$

We call F *the integral of $f(x)$ on $[a,b]$*. One way to think of $F(t)$ is as an "area function", since if $f(x) > 0$ on $[a,b]$, then $F(t)$ looks to represent the area under the curve $y = f(x)$ over the interval $[a,t]$.

If $c \le d$ in $[a,b]$, then Proposition 3.3 shows that

$$\int_a^d f(x) = \int_a^c f(x) + \int_c^d f(x)$$

so that

$$F(d) - F(c) = \int_c^d f(x) \tag{4.1}$$

If we define $\int_d^c f(x)$ to be $-\int_c^d f(x)$ then observe that $\int_d^c f(x) = F(c) - F(d)$ so that equation (4.1) holds for *all* values $c, d \in [a,b]$ regardless which of c or d is larger. Another technical point: the variable t for the function $F(t)$ is a place-holder; we could equally well write $F(x)$ as $F(t)$. In the definition of F, we have to distinguish between the variable of integration and the variable for F (the upper limit of integration). We will usually use x as the variable for F.

In general $F(x)$ does not have to have a derivative, but it does have to be continuous.

Proposition 4.1. *Let $F(x)$ be the integral of $f(x)$ on $[a, b]$. Then $F(x)$ is continuous.*

Proof. Choose $c < d$ in $[a, b]$. The definition in Theorem 3.3 of the integral $\int_c^d f(x)$ shows that it is less than or equal to any particular upper sum. The upper sum for the partition $\{[c, d]\}$ is $\sup([c, d], f)(d - c)$. With this in mind, compute

$$F(d) - F(c) = \int_c^d f(x) \leq \sup([c, d], f)(d - c) \leq \sup([a, b], f)(d - c)$$

Also

$$F(c) - F(d) = -\int_c^d f(x) \leq -\inf([a, b], f)(d - c)$$

It follows that $|F(d) - F(c)|$ is bounded above by a constant times $|d - c|$. (The constant can be the larger of the two numbers $|\sup([a, b], f)|$ and $|\inf([a, b], f)|$.) This proves that F is continuous. QED

Our next two facts are absolutely essential.

Integral Mean Value Theorem. *Let $f(x)$ be continuous on $[a, b]$, and let $c, d \in [a, b]$. Then there is r between c and d such that*

$$\int_c^d f(x) = f(r)(d - c)$$

Proof. First assume that $c \leq d$. Let P be the partition $\{[c, d]\}$ (of the interval $[c, d]$!), and use the inequality

$$\inf([c, d], f)(d - c) = \Sigma_l([c, d], f) \leq \int_c^d f(x) \leq \Sigma^u([c, d], f) = \sup([c, d], f)(d - c)$$

Because f is continuous, the numbers $\inf([c, d], f)$ and $\sup([c, d], f)$ are values of $f(x)$. By the Intermediate Value Theorem, there is $r \in [c, d]$ such that

$$f(r)(d - c) = \int_c^d f(x)$$

The remaining case is $c > d$, and then by the previous paragraph, there is $r \in [d, c]$ such that $\int_d^c f(x) = f(r)(c - d)$. Multiplying this equation by -1 gives the desired conclusion. QED

We call attention to the hypothesis that f is continuous in the following.

Fundamental Theorem of Calculus. *Let $f(x)$ be continuous on $[a, b]$. Then the integral $F(x)$ for $f(x)$ on $[a, b]$ is an antiderivative for $f(x)$. If $G(x)$ is an antiderivative of $f(x)$ on $[a, b]$, then*

$$\int_a^b f(x) = G(b) - G(a)$$

Proof. Choose $c, c + \delta \in [a, b]$ with $\delta \neq 0$. By the Integral Mean Value Theorem there is r between c and $c + \delta$ such that

$$\int_c^{c+\delta} f(x) = f(r)(c + \delta - c) = f(r)\delta$$

(the number r depends on c and $c+\delta$). We can divide by δ and use equation (4.1) to obtain

$$\frac{1}{\delta}\int_c^{c+\delta} f(x) = \frac{F(c+\delta)-F(c)}{\delta} = f(r)$$

Letting δ go to zero, r goes to c, so that $f(r)$ goes to $f(c)$ since f is continuous. Thus $F'(c) = f(c)$.

If $G(x)$ is an antiderivative of $f(x)$ then the Mean Value Theorem finds a constant C such that $G(x) = F(x) + C$ and then $G(b) - G(a) = F(b) + C - (F(a) + C) = F(b) - F(a)$ which is $\int_a^b f(x)$. QED

The Fundamental Theorem seems to say that

$$F(b) - F(a) = \int_a^b f(x) \quad \text{when} \quad F'(x) = f(x) \tag{4.2}$$

In fact, it *does* say this **if** $f(x)$ is continuous. We want to hedge this fact with two examples. If f is not continuous, but only integrable, we can still define the function $F(y) = \int_a^y f(x)$ so that $F(y)$ is continuous and the integral equation in (4.2) is valid (rather trivially). However, we do not have $F'(x) = f(x)$ except at those x where $f(x)$ is continuous. For example, changing the value of f at some point c does not change the integral (Proposition 3.5d), and so does not change F, but it would invalidate $F'(c) = f(c)$. Thus, the integral of an integrable function does not have to be its antiderivative.

There is a second thing that can go wrong with the formula (4.2). Define

$$F(x) = \begin{cases} x^2\sin(x^{-2}) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

It is not too hard to see that $F(x)$ has a derivative for every x. Indeed, $F'(x)$ can be computed from the usual rules when $x \neq 0$. And $F'(0)$ can be computed directly from the definition of the derivative. We get

$$F'(x) = \begin{cases} 2x\sin(x^{-2}) - 2x^{-1}\cos(x^{-2}) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

The presence of the x^{-1} factor in $F'(x)$ makes it fairly easy to see that $F'(x)$ is not bounded on $[0,1]$, and so it is not integrable there. Thus $\int_0^1 F'(x)$ doesn't make sense, even though F is continuous.

5. Leibnitz differential notation.

We introduced Leibnitz' approach to the integral as an infinite sum of terms $f(x)\cdot dx$ where dx is infinitely small (but not 0!). The differential notation can also be used to think about the derivative:

$$F'(x) = \frac{dF}{dx}$$

This notation makes is easy to anticipate the Fundamental Theorem: if $F'(x) = f(x)$ so that $\frac{dF}{dx} = f(x)$, then $dF = f(x)\cdot dx$ and

$$\int_a^b f(x)\cdot dx = \int_a^b dF$$

If we think of $\int_a^b dF$ as a sum of small changes in F, then this "sum" ought to be the net change in F across $[a,b]$; in other words, we ought to have

$$\int_a^b dF = F(b) - F(a)$$

The foregoing is not a proof, but it motivates the use of the differential.

From a modern point of view, the differential is a convenient notational device – a device that allows us correctly to apply standard transformations of the integral. The two most commonly used transformations are substitution and parts, usually thought of as "techniques of integration" to be used along with the power rule and partial fractions to find explicit function formulas for an antiderivative.

The two results we give can be generalized in many ways. We will stick to simple formulations, emphasizing the use of the differential.

Substitution. Let $g(x)$ have a continuous derivative on $[a, b]$ and suppose that g maps $[a, b]$ into the interval $[c, d]$. Let $f(x)$ be continuous on $[c, d]$. We consider the integral $\int_a^b f(g(x)) \cdot g'(x) \cdot dx$. Write $y = g(x)$ so that $\frac{dy}{dx} = g'(x)$, which can be written $dy = g'(x) \cdot dx$, and we substitute

$$\int_a^b f(g(x)) \cdot g'(x) \cdot dx = \int_{g(a)}^{g(b)} f(y) \cdot dy \tag{5.1}$$

To show that the formula is correct, let F be an antiderivative for f on $[c, d]$. Then the Chain Rule shows that $F(g(x))$ has derivative $f(g(x))g'(x)$, and so Fundamental Theorem shows that the integral on the left of (5.1) is $F(g(b)) - F(g(a))$. The integral on the right has the same value, since $F' = f$.

Parts. Assume that U and V are continuously differentiable on $[a, b]$. As in Substitution, we write $dU = U'(x) \cdot dx$ and $dV = V'(x) \cdot dx$. The parts formula is

$$\int_a^b U \cdot dV = U \cdot V\Big|_a^b - \int_a^b V \cdot dU$$

This is true simply because of the product rule:

$$U \cdot V\Big|_a^b = \int_a^b \frac{d}{dx}(U \cdot V) \cdot dx = \int_a^b (U' \cdot V + U \cdot V') \cdot dx = \int_a^b V \cdot dU + \int_a^b U \cdot dV$$

as needed.

6. Applications of integration – a critique.

Retrieve your calculus text and look up the section(s) on applications of integration. You will find integration formulas for many interesting quantities: areas, volumes, moments, arc length, mass, momentum, etc. Although we appreciate these applications very much, we would like to criticize the way in which most texts derive the integral formulas.

Consider, for example, the argument that purports to show the following: if $f(x)$ is positive and continuous on $[a, b]$, then $\int_a^b f(x)$ is the area between $y = f(x)$, the x-axis, and $a \leq x \leq b$. Most texts use a Riemann sum to approximate the area as the sum of the areas of rectangles, "letting the mesh go to 0" in the partition to sharpen the approximation. Our Theorem 3.1 shows that the Riemann sums get close to the integral as the mesh gets small; the crucial question is how we know that the Riemann sums approach the area. There are at least two (somewhat overlapping) ways in which this question could be answered, stemming from two possible attitudes toward the area we are trying to compute. If we assume an intuitive understanding of area (such an understanding seems implicit in most texts), then in order to know that Riemann sums converge to the area, we would need to estimate the difference between a Riemann sum and the area. But the texts do not do this. Indeed, it is hard to imagine how this might be done without having definite formulas for area in advance – and it is precisely the point of using the integral to *obtain* such formulas. Alternatively or additionally, we might take a more formal approach and presume to be *defining* area via an integral formula. Many texts give an intuitive argument that Riemann sums ought to approximate area, and then they use the intuitive argument to motivate the *definition* of area as an integral. The trouble with this approach is that it raises the question how we know that the "area" defined by the integral has all the properties that area is supposed to have. This is not a trivial issue; for example, rotating the plane about an axis should not effect the area of a given region; it is difficult to adjust the integral for the effect of a simple rotation.

Questions like the one just raised for area become compounded when one is calculating a quantity that has physical significance, such as the mass of an object. To show that a mass can be calculated by an integral one really does need to relate physical properties of mass to the Riemann sums considered in an "approximation of the mass," and the texts just do not do this. The texts that claim to "calculate mass," assuming mass to be understood at least intuitively, are taking liberties with mathematics, since they do not estimate the difference between a Riemann sum approximation and the mass approximated. If we purport to "define mass" as an integral, we are taking liberties with physics, since we would need to show that our definition of mass has the properties claimed by a physicist.

One way out of this mess involves Darboux's approach to the integral. Rather than view the integral as a limit, we use the uniqueness statement in Theorem 3.3 to identify the integral. This approach does not need to distinguish whether we are assuming properties of the applied quantity (e.g. area, moment, etc.) or whether we are using an integral to give a formal definition of that quantity.

To give an example, consider the area problem to which we have already referred: to calculate, or perhaps define, the area A between $y = f(x)$ and the interval $[a, b]$ on the x-axis, over which $f(x)$ is continuous. Let P be a partition of $[a, b]$; for $I \in P$, consider the minimum and maximum of $f(x)$ on I; these numbers are $\inf(I, f)$ and $\sup(I, f)$, respectively. The rectangle of width $|I|$ and height $\inf(I, f)$ fits inside the region $R(I)$ bounded by the interval I on the x-axis and $y = f(x)$ for $x \in I$. If we claim an intuitive understanding of area, we would claim that the area $\inf(I, f) \cdot |I|$ of the rectangle would be less than or equal to the area $A(I)$ of $R(I)$. If we claim to be defining (or discovering) area, we would still want $\inf(I, f) \cdot |I| \le A(I)$ to be true. Thus, we obtain the inequality whatever our prior view of area. Similarly, we obtain $A(I) \le \sup(I, f) \cdot |I|$. Furthermore, the areas of all the $R(I)$ add up to the area A of the entire region; again, either because we "know" that area has this property or because we "want to discover a definition of area that has this property." In any case, we are led to the inequalities:

$$\Sigma_l(P, f) = \sum_{I \in P} \inf(I, f) \cdot |I| \le A \le \sum_{I \in P} \inf(I, f) \cdot |I| = \Sigma^u(P, f)$$

Theorem 3.3 tells that the integral $\int_a^b f(x)$ is the only number that can fit, as A does, between all lower and upper sums. If we understand A intuitively, then we have just discovered its integral formula; if we are trying to define A, we see that it must be defined as an integral.

A discussion like the one of the previous paragraph can be given for almost all the applications of integration normally encountered in calculus or in undergraduate science. There are a couple such quantities that require a slightly different approach in that these quantities do not fit between lower and upper sums. Arc length is one such quantity.

Suppose we want to calculate the "length" of the curve $y = f(x)$ for $a \le x \le b$. To keep the technical details to a minimum, assume that $f(x)$ has an integrable derivative on $[a, b]$. For a partition P of $[a, b]$, we use the endpoints of the intervals in P to approximate the length of $f(x)$. Write the sub-division for P:

$$a = x_0 < x_1 < \cdots < x_n = b \quad \text{and define} \quad L(P) = \sum_{j=1}^{n} \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2}$$

The quantity $L(P)$ is the sum of lengths of line segment along $y = f(x)$ stretching from $(a, f(a))$ to $(b, f(b))$. The usual definition of *arc length* stipulates that the *length* of $y = f(x)$ for $a \le x \le b$ is the supremum of all the $L(P)$.

Arc Length Formula. *Let $f(x)$ have an integrable derivative on $[a, b]$. Then the length of $y = f(x)$ for $a \le x \le b$ is*

$$\int_a^b \sqrt{1 + f'(x)^2} \cdot dx$$

Proof. The triangle inequality in the plane shows that if the partition Q refines the partition P, then $L(P) \leq L(Q)$. We will use this below.

Let P be a partition of $[a,b]$, and let $I \in P$, writing $I = [x_{j-1}, x_j]$. Compute that

$$\sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \quad = \quad (x_j - x_{j-1}) \cdot \sqrt{1 + \left[\frac{f(x_j) - f(x_{j-1})}{x_j - x_{j-1}}\right]^2} \tag{6.1}$$

By the Mean Value Theorem, there is $x_I \in I$ such that

$$f(x_j) - f(x_{j-1}) = f'(x_I) \cdot (x_j - x_{j-1})$$

and when this is substituted into (6.1), remembering that $x_j - x_{j-1} = |I|$, we obtain

$$\sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \quad = \quad |I| \cdot \sqrt{1 + f'(x_I)^2} \tag{6.2}$$

Applying (6.2) to each $I \in P$, we obtain the following:

$$L(P) = \sum_{I \in P} |I| \cdot \sqrt{1 + f'(x_I)^2}$$

and this is a Riemann sum for $\sqrt{1 + f'(x)^2}$ on $[a,b]$ using P.

Let L be the length of $y = f(x)$ on $[a,b]$, and let $\epsilon > 0$. Because L is defined as a sup, there is a partition P such that $L - \epsilon < L(P) \leq L$. We can refine P to the partition Q where $\Sigma_V(Q, \sqrt{1 + f'^2}) < \epsilon$. Corollary 3.4 then shows that every Riemann sum for $\sqrt{1 + f'^2}$ using Q is within ϵ of the integral $\int_a^b \sqrt{1 + f'(x)^2}$. As we remarked at the beginning of the proof, since Q refines P, we have $L(P) \leq L(Q)$. The definition of L shows that $L(Q) \leq L$, and so $L(Q)$ is within ϵ of L. It follows that $\int_a^b \sqrt{1 + f'(x)^2}$ is within 2ϵ of L. That this is true for all ϵ shows that the integral is equal to the arc length L. QED

7. Uniform convergence.

For this section, we assume that the reader knows the definition of "uniform convergence" of a sequence $f_n(x)$ of functions over their common domain. We also suppose that preliminary examples of uniform and non-uniform convergence have been given, so that the reader is ready for the following:

Theorem 7.1. *Suppose, for each $n \geq 0$, that $f_n(x)$ is integrable on $[a,b]$. Suppose that the $f_n(x)$ converge uniformly to $f(x)$ on $[a,b]$. Then $f(x)$ is integrable on $[a,b]$ and*

$$\lim_{n \to \infty} \int_a^b f_n(x) = \int_a^b f(x)$$

Since the integral on the right can be written

$$\int_a^b \left(\lim_{n \to \infty} f_n(x) \right)$$

This theorem shows that the limit and integral can be interchanged when the convergence is uniform.

Proof of Theorem 7.1. Choose $\epsilon > 0$, and the uniformity of convergence finds N such that if $n \geq N$ and $x \in [a,b]$, then $|f_n(x) - f(x)| < \epsilon$. If $x, y \in [a,b]$, then

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(x)| \leq 2\epsilon + |f_n(x) - f_n(y)|$$

If I is a sub-interval of $[a, b]$, it follows that $\text{var}(I, f) \leq \text{var}(I, f_n) + 2\epsilon$. Then if P is a partition of $[a, b]$, we see that

$$\Sigma_V(P, f) \leq \Sigma_V(P, f_n) + 2\epsilon \cdot (b - a)$$

Therefore, f is integrable.

Furthermore, if x_I is chosen in I for every $I \in P$, we estimate

$$\left| \sum_{I \in P} f(x_I) \cdot |I| - \sum_{I \in P} f_n(x_I) \cdot |I| \right| \leq \epsilon \cdot (b - a)$$

This establishes the limit of the integrals. QED