5

Elementary Scalar Calculus

5.1 Objectives

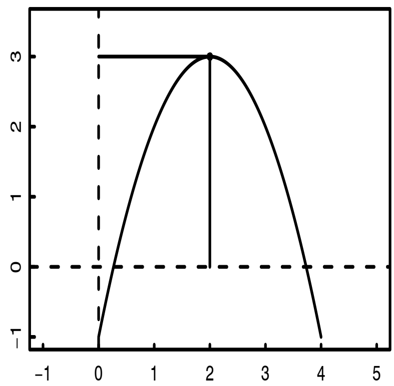
This chapter introduces the basics of calculus operating on scalar quantities, and most of these principles can be understood quite readily. Many find that the language and imagery of calculus are a lot more intimidating than the actual material (once they commit themselves to studying it). There are two primary tools of calculus, differentiation and integration, both of which are introduced here. A further chapter gives additional details and nuances, as well as an explanation of calculus on nonscalar values like vectors and matrices.

5.2 Limits and Lines

The first important idea on the way to understanding calculus is that of a limit. The key point is to see how functions behave when some value is made: arbitrarily small or arbitrarily large on some measure, or arbitrarily close to some finite value. That is, we are specifically interested in how a function tends to or converges to some point or line in the limit.

Consider the function $f(x) = 3 - (x-2)^2$ over the domain (support) $[0:4]$. This function is **unimodal**, possessing a single maximum point, and it is **symmetric**, meaning that the shape is mirrored on either side of the line through the middle (which is the mode here). This function is graphed in the figure at the right.

Fig. 5.1. $f(x) = 3 - (x-2)^2$



What happens as the function approaches this mode from either direction? Consider "creeping up" on the top of the hill from either direction, as tabulated:

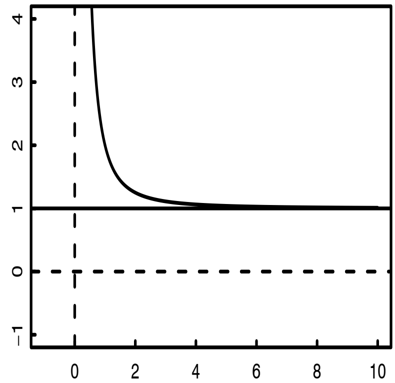
Left	x	1.8000	1.9000	1.9500	1.9900
	$f(x)$	2.9600	2.9900	2.9975	2.9999
Right	x	2.2000	2.1000	2.0500	2.0100
	$f(x)$	2.9600	2.9900	2.9975	2.9999

It should be obvious from the graph as well as these listed values that the **limit of the function** as $x \to 2$ is 3, approached from either direction. This is denoted $\lim_{x\to 2} f(x) = 3$ for the general case, as well as $\lim_{x\to 2-} f(x) = 3$ for approaching from the left and $\lim_{x\to 2+} f(x) = 3$ for approaching from the right. The reason that the right-hand limit and the left-hand limit are equal is that the function is continuous at the point of interest. If the function is smooth (continuous at all points; no gaps and no "corners" that cause problems here), then the left-hand limit and the right-hand limit are always identical except for at infinity.

Let us consider a more interesting case. Can a function have a finite limiting value in $f(x)$ as x goes to infinity? The answer is absolutely yes, and this turns out to be an important principle in understanding some functions.

An interesting new function is $f(x) = 1 + 1/x^2$ over the domain $(0{:}\infty^+)$. Note that this function's range is over the positive real numbers greater than or equal to one because of the square placed here on x. Again, the function is graphed in the figure at right showing the line at $y = 1$. What happens as this function approaches infinity from from the left? Obviously it does not make sense to approach infinity from the right! Consider again tabulated the values:

Fig. 5.2. $f(x) = 1 + 1/x^2$



x	1	2	3	6	12	24	100
$f(x)$	2.0000	1.2500	1.1111	1.0278	1.0069	1.0017	1.0001

Once again the effect is not subtle. As x gets arbitrarily large, $f(x)$ gets progressively closer to 1. The curve approaches but never seems to reach $f(x) = 1$ on the graph above. What occurs at exactly ∞ though? Plug ∞ into the function and see what results: $f(x) = 1+1/\infty = 1$ ($1/\infty$ is defined as zero because 1 divided by progressively larger numbers gets progressively smaller and infinity is the largest number). So ***in the limit*** (and only in the limit) the function reaches 1, and for every finite value the curve is above the horizontal line at one. We say here that the value 1 is the **asymptotic value** of the function $f(x)$ as $x \to \infty$ and that the line $y = 1$ is the **asymptote**: $\lim_{x\to\infty} f(x) = 1$.

There is another limit of interest for this function. What happens at $x = 0$? Plugging this value into the function gives $f(x) = 1 + 1/0$. This produces a result that we cannot use because dividing by zero is not defined, so the function has no allowable value for $x = 0$ but does have allowable values for every positive x. Therefore the asymptotic value of $f(x)$ with x approaching zero from the right is infinity, which makes the vertical line $y = 0$ an asymptote of a different kind for this function: $\lim_{x\to 0^+} f(x) = \infty$.

There are specific properties of interest for limits (tabulated here for the variable x going to some arbitrary value X).

Properties of Limits, $\exists \lim_{x\to X} f(x), \lim_{x\to X} g(x)$**, Constant** k

$\rightarrowtail$ Addition and Subtraction

$$\lim_{x\to X}[f(x)+g(x)] = \lim_{x\to X} f(x) + \lim_{x\to X} g(x)$$

$$\lim_{x\to X}[f(x)-g(x)] = \lim_{x\to X} f(x) - \lim_{x\to X} g(x)$$

$\rightarrowtail$ Multiplication $\quad \lim_{x\to X}[f(x)g(x)] = \lim_{x\to X} f(x) \lim_{x\to X} g(x)$

$\rightarrowtail$ Scalar Multiplication $\quad \lim_{x\to X}[kg(x)] = k\lim_{x\to X} g(x)$

$\rightarrowtail$ Division $\left(\lim_{x\to X} g(x) \neq 0\right)$ $\quad \lim_{x\to X}[f(x)/g(x)] = \frac{\lim_{x\to X} f(x)}{\lim_{x\to X} g(x)}$

$\rightarrowtail$ Constants $\quad \lim_{x\to X} k = k$

$\rightarrowtail$ Natural Exponent $\quad \lim_{x\to\infty}\left[1+\frac{k}{x}\right]^x = e^k$

Armed with these rules we can analyze the asymptotic properties of more complex functions in the following examples.

★ **Example 5.1: Quadratic Expression.**

$$\lim_{x\to 2}\left[\frac{x^2+5}{x-3}\right] = \frac{\lim_{x\to 2} x^2+5}{\lim_{x\to 2} x-3} = \frac{2^2+5}{2-3} = -9.$$

★ **Example 5.2: Polynomial Ratio.**

$$\lim_{x\to 1}\left[\frac{x^3-1}{x-1}\right] = \lim_{x\to 1}\left[\frac{(x-1)(x+1)(x+1)-x(x-1)}{(x-1)}\right]$$
$$= \lim_{x\to 1}\left[\frac{(x+1)^2-x}{1}\right] = \lim_{x\to 1}(x+1)^2 - \lim_{x\to 1}(x) = 3.$$

★ **Example 5.3: Fractions and Exponents.**

$$\lim_{x\to\infty}\left[\frac{\left(1+\frac{k_1}{x}\right)^x}{\left(1+\frac{k_2}{x}\right)^x}\right]=\frac{\lim\limits_{x\to\infty}\left(1+\frac{k_1}{x}\right)^x}{\lim\limits_{x\to\infty}\left(1+\frac{k_2}{x}\right)^x}=\frac{e^{k_1}}{e^{k_2}}=e^{k_1-k_2}.$$

★ **Example 5.4: Mixed Polynomials.**

$$\lim_{x\to 1}\left[\frac{\sqrt{x}-1}{x-1}\right]=\lim_{x\to 1}\left[\frac{(\sqrt{x}-1)(\sqrt{x}+1)}{(x-1)(\sqrt{x}+1)}\right]=\lim_{x\to 1}\left[\frac{x-1}{(x-1)(\sqrt{x}+1)}\right]$$
$$=\lim_{x\to 1}\left[\frac{1}{(\sqrt{x}+1)}\right]=\frac{1}{\lim\limits_{x\to 1}\sqrt{x}+1}=0.5.$$

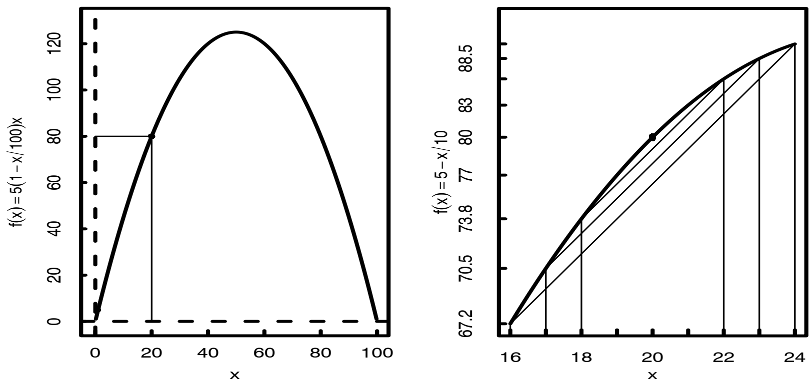
5.3 Understanding Rates, Changes, and Derivatives

So why is it important to spend all that time on limits? We now turn to the definition of a derivative, which is based on a limit. To illustrate the discussion we will use a formal model from sociology that seeks to explain thresholds in voluntary racial segregation. Granovetter and Soong (1988) built on the foundational work of Thomas Schelling by mathematizing the idea that members of a racial group are progressively less likely to remain in a neighborhood as the proportion of another racial group rises. Assuming just blacks and whites, we can define the following terms: x is the percentage of whites, R is the "tolerance" of whites for the ratio of whites to blacks, and N_w is the total number of whites living in the neighborhood. In Granovetter and Soong's model, the function $f(x)$ defines a mobility frontier whereby an absolute number of blacks above the frontier causes whites to move out and an absolute number of blacks below the frontier causes whites to move in (or stay). They then developed and justified the function:

$$f(x)=R\left[1-\frac{x}{N_w}\right]x,$$

which is depicted in the first panel of Figure 5.3 for $N_w=100$ and $R=5$ (so $f(x)=5x-\frac{1}{20}x^2$). We can see that the number of blacks tolerated by whites increases sharply moving right from zero, hits a maximum at 125, and then decreases back to zero. This means that the tolerated level was **monotonically increasing** (constantly increasing or staying the same, i.e., nondecreasing) until the maxima and then **monotonically decreasing** (constantly decreasing or staying the same, i.e., nonincreasing) until the tolerated level reaches zero.

Fig. 5.3. Describing the Rate of Change



We are actually interested here in the **rate of change** of the tolerated number as opposed to the tolerated number itself: The rate of increase steadily declines from an earnest starting point until it reaches zero; then the rate of decrease starts slowly and gradually picks up until the velocity is zero. This can be summarized by the following table (recall that $\in$ means "an element of").

Region	Speed	Rate of Change
$x \in (0{:}50]$	increasing	decreasing
$x = 50$	maximum	zero
$x \in [50{:}100)$	decreasing	increasing

Say that we are interested in the rate of change *at exactly time* $x = 20$, which is the point designated at coordinates $(20, 80)$ in the first panel of Figure 5.3. How would we calculate this? A reasonable approximation can be made with line segments. Specifically, starting 4 units away from 20 in either direction, go 1 unit along the x-axis toward the point at 20 and construct line segments connecting the points along the curve at these x levels. The slope of the line segment (easily calculated from Section 1.5.1) is therefore an approximation to the instantaneous rate at $x = 20$, "rise-over-run," given by the segment

$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

So the first line segment of interest has values $x_1 = 16$ and $x_2 = 24$. If we call the width of the interval $h = x_2 - x_1$, then the point of interest, x, is at the center of this interval and we say

$$\begin{aligned} m &= \frac{f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)}{h} \\ &= \frac{f(x+h) - f(x)}{h} \end{aligned}$$

because $f(h/2)$ can move between functions in the numerator. This segment is shown as the lowest (longest) line segment in the second panel of Figure 5.3 and has slope 2.6625.

In fact, this estimate is not quite right, but it is an average of a slightly faster rate of change (below) and a slightly slower rate of change (above). Because this is an estimate, it is reasonable to ask how we can improve it. The obvious idea is to decrease the width of the interval around the point of interest. First go to 17–23 and then 18–22, and construct new line segments and therefore new estimates as shown in the second panel of Figure 5.3. At each reduction in interval width we are improving the estimate of the instantaneous rate of change at $x = 20$. Notice the nonlinear scale on the y-axis produced by the curvature of the function.

When should we stop? The answer to this question is found back in the previous discussion of limits. Define h again as the length of the intervals created as just described and call the expression for the slope of the line segment $m(x)$, to distinguish the slope form from the function itself. The point where $\lim\limits_{h \to 0}$ occurs is the point where we get ***exactly*** the instantaneous rate of change at $x = 20$ since the width of the interval is now zero, yet it is still "centered" around $(20, 80)$. This instantaneous rate is equal to the slope of the **tangent line** (not to be confused with the tangent trigonometric function from Chapter 2) to the curve at the point x: the line that touches the curve ***only at this one point***. It can be shown that there exists a unique tangent line for every point on a smooth curve.

So let us apply this logic to our function and perform the algebra very mechanically:

$$\begin{aligned}
\lim_{h\to 0} m(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{\left[5(x+h) - \frac{1}{20}(x+h)^2\right] - \left[5x - \frac{1}{20}x^2\right]}{h} \\
&= \lim_{h\to 0} \frac{5h - \frac{1}{20}(x^2 + 2xh + h^2) + \frac{1}{20}x^2}{h} \\
&= \lim_{h\to 0} \frac{5h - \frac{2}{20}xh - \frac{1}{20}h^2}{h} \\
&= \lim_{h\to 0} \left(5 - \frac{1}{10}x - \frac{1}{20}h\right) = 5 - \frac{1}{10}x.
\end{aligned}$$

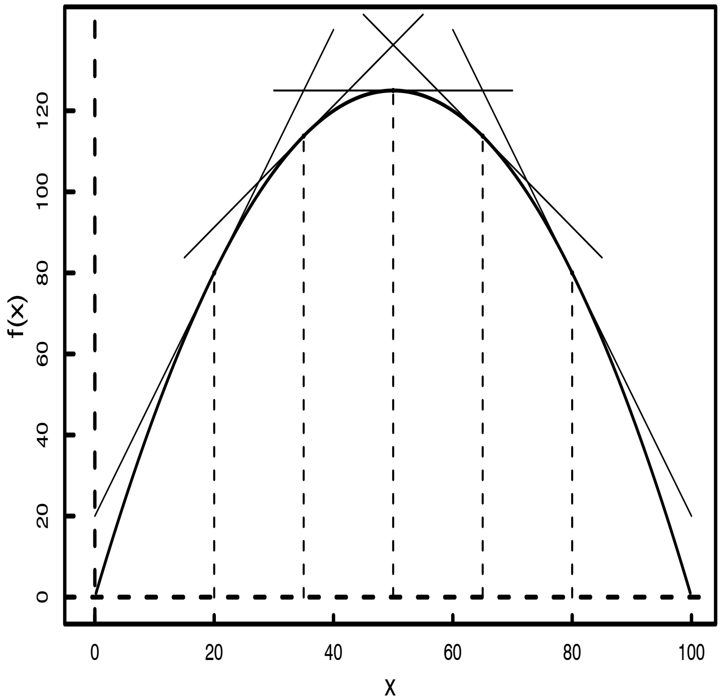
This means that for any allowable x point we now have an expression for the instantaneous slope at that point. Label this with a prime to clarify that it is a different, but related, function: $f'(x) = 5 - \frac{1}{10}x$. Our point of interest is 20, so $f'(20) = 3$. Figure 5.4 shows tangent lines plotted at various points on $f(x)$. Note that the tangent line at the maxima is “flat,” having slope zero. This is an important principle that we will make extensive use of later.

What we have done here is produce the **derivative** of the function $f(x)$, denoted $f'(x)$, also called **differentiating** $f(x)$. This derivative process is fundamental and has the definition

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

The derivative expression $f'(x)$ is Euler’s version of Newton’s notation, but it is often better to use Leibniz’s notation $\frac{d}{dx}f(x)$, which resembles the limit derivation we just performed, substituting $\Delta x = h$. The change (delta) in x is

Fig. 5.4. Tangent Lines on $f(x) = 5x - \frac{1}{20}x^2$



therefore

$$\frac{d}{dx}f(x) = \frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{\Delta f(x)}{\Delta x}.$$

This latter notation for the derivative is generally preferred because it better reflects the change in the function $f(x)$ for an infinitesimal change in x, and it is easier to manipulate in more complex problems. Also, note that the fractional form of Leibniz's notation is given in two different ways, which are absolutely equivalent:

$$\frac{d}{dx}u = \frac{du}{dx},$$

for some function $u = f(x)$. Having said all that, Newton's form is more compact and looks nicer in simple problems, so it is important to know each form because they are both useful.

To summarize what we have done so far:

Summary of Derivative Theory

↣	Existence	$f'(x)$ at x exists iff $f(x)$ is continuous at x, and there is no point where the right-hand derivative and the left-hand derivative are different
↣	Definition	$f'(x) = \lim\limits_{h \to 0} \frac{f(x+h)-f(x)}{h}$
↣	Tangent Line	$f'(x)$ is the slope of the line tangent to $f()$ at x; this is the limit of the enclosed secant lines

The second existence condition needs further explanation. This is sometimes call the "no corners" condition because these points are geometric corners of the function and have the condition that

$$\lim_{\Delta x \to 0^-} \frac{\Delta f(x)}{\Delta x} \neq \lim_{\Delta x \to 0^+} \frac{\Delta f(x)}{\Delta x}.$$

That is, taking these limits to the left and to the right of the point produces different results. The classic example is the function $f(x) = |x|$, which looks like a "V" centered at the origin. So infinitesimally approaching $(0, 0)$ from the left, $\Delta x \to 0^-$, is different from infinitesimally approaching $(0, 0)$ from the right, $\Delta x \to 0^+$. Another way to think about this is related to Figure 5.4. Each of the illustrated tangent lines is uniquely determined by the selected point along the function. At a corner point the respective line would be allowed to "swing around" to an infinite number of places because it is resting on a single point (atom). Thus no unique derivative can be specified.

★ **Example 5.5: Derivatives for Analyzing Legislative Committee Size.** Francis (1982) wanted to find criteria for determining "optimal" committee sizes in Congress or state-level legislatures. This is an important question because committees are key organizational and procedural components of American legislatures. A great number of scholars of American politics have observed the central role that committee government plays, but not nearly as many have sought to understand committee size and subsequent efficiency. Efficiency is defined by Francis as minimizing two criteria for committee members:

- ***Decision Costs:*** (Y_d) the time and energy required for obtaining policy information, bargaining with colleagues, and actual meeting time.
- ***External Costs:*** (Y_e) the electoral and institutional costs of producing nonconsensual committee decisions (i.e., conflict).

Francis modeled these costs as a function of committee size in the following partly-linear specifications:

$$Y_d = a_d + b_d X^g, \qquad g > 0$$

$$Y_e = a_e + b_e X^k, \qquad k < 0,$$

where the a terms are intercepts (i.e., the costs for nonexistence or non-membership), the b terms are slopes (the relative importance of the multiplied term X) in the linear sense discussed in Section 1.5.1, and X is the size of the committee. The interesting terms here are the exponents on X. Since g is necessarily positive, increasing committee size increases decision costs, which makes logical sense. The term k is restricted to be negative, implying that the larger the committee, the closer the representation on the committee is to the full chamber or the electorate and therefore the lower such outside costs of committee decisions are likely to be to a member.

The key point of the Francis model is that committee size is a trade-off between Y_d and Y_e since they move in opposite directions for increasing or decreasing numbers of members. This can be expressed in one single equation by asserting that the two costs are treated equally for members (an assumption that could easily be generalized by weighting the two functions differently) and adding the two cost equations:

$$Y = Y_d + Y_e = a_d + b_d X^g + a_e + b_e X^k.$$

So now we have a single equation that expresses the full identified costs to members as a function of X. How would we understand the effect of changing the committee size? Taking the derivative of this expression with respect to X gives the instantaneous rate of change in these political costs at levels of X, and understanding changes in this rate helps to identify better or worse committee sizes for known or estimated values of a_d, b_d, g, a_e b_e, and also k.

The derivative operating on a polynomial does the following: It multiplies the expression by the exponent, it decreases the exponent by one, and it gets rid of any isolated constant terms. These rules are further explained and illustrated in the next section. The consequence in this case is that the first derivative of the Francis model is

$$\frac{d}{dx}Y = gb_dX^{g-1} + kb_eX^{k-1},$$

which allowed him to see the instantaneous effect of changes in committee size and to see where important regions are located. Francis thus found a minimum point by looking for an X value where $\frac{d}{dx}Y$ is equal to zero (i.e., the tangent line is flat). This value of X (subject to one more qualification to make sure that it is not a maximum) minimizes costs as a function of committee size given the known parameters.

5.4 Derivative Rules for Common Functions

It would be pretty annoying if every time we wanted to obtain the derivative of a function we had to calculate it through the limit as was done in the last section. Fortunately there are a number of basic rules such that taking derivatives on polynomials and other functional forms is relatively easy.

5.4.1 Basic Algebraic Rules for Derivatives

We will provide a number of basic rules here without proof. Most of them are fairly intuitive.

The **power rule** (already introduced in the last example) defines how to treat exponents in calculating derivatives. If n is any real number (including fractions and negative values), then

$$\frac{d}{dx}x^n = nx^{n-1}.$$

Straightforward examples include

$$\frac{d}{dx}x^2 = 2x \qquad \frac{d}{dx}3x^3 = 9x^2 \qquad \frac{d}{dx}x^{1000} = 1000x^{999}$$

$$\frac{d}{dx}x = 1 \qquad \frac{d}{dx}\sqrt{x} = \frac{1}{2}x^{-\frac{1}{2}} \qquad \frac{d}{dx}\left(-\frac{5}{8}x^{-\frac{8}{5}}\right) = x^{-\frac{13}{5}}.$$

This is a good point in the discussion to pause and make sure that the six examples above are well understood by the reader as the power rule is perhaps the most fundamental derivative operation.

The **derivative of a constant** is always zero:

$$\frac{d}{dx}k = 0, \quad \forall k.$$

This makes sense because a derivative is a rate of change and constants do not change. For example,

$$\frac{d}{dx}2 = 0$$

(i.e., there is no change to account for in 2). However, when a constant is multiplying some function of x, it is immaterial to the derivative operation, but it has to be accounted for later:

$$\frac{d}{dx}kf(x) = k\frac{d}{dx}f(x).$$

As an example, the derivative of $f(x) = 3x$ is simply 3 since $\frac{d}{dx}f(x)$ is 1.

The **derivative of a sum** is just the sum of the derivatives, provided that each component of the sum has a defined derivative:

$$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x),$$

and of course this rule is not limited to just two components in the sum:

$$\frac{d}{dx}\sum_{i=1}^{k} f_i(x) = \frac{d}{dx}f_1(x) + \frac{d}{dx}f_2(x) + \cdots + \frac{d}{dx}f_k(x).$$

For example,

$$\frac{d}{dx}\left(x^3 - 11x^2 + 3x - 5\right) = \frac{d}{dx}x^3 - \frac{d}{dx}11x^2 + \frac{d}{dx}3x - \frac{d}{dx}5$$
$$= 3x^2 - 22x + 3.$$

Unfortunately, the **product rule** is a bit more intricate than these simple methods. The product rule is the sum of two pieces where in each piece one of the two multiplied functions is left alone and the other is differentiated:

$$\frac{d}{dx}\left[f(x)g(x)\right] = f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x).$$

This is actually not a very difficult formula to remember due to its symmetry. As an example, we now differentiate the following:

$$\frac{d}{dx}(3x^2 - 3)(4x^3 - 2x)$$
$$= (3x^2 - 3)\frac{d}{dx}(4x^3 - 2x) + (4x^3 - 2x)\frac{d}{dx}(3x^2 - 3)$$
$$= (3x^2 - 3)(12x^2 - 2) + (4x^3 - 2x)(6x)$$
$$= (36x^4 - 6x^2 - 36x^2 + 6) + (24x^4 - 12x^2)$$
$$= 60x^4 - 54x^2 + 6$$

(where we also used the sum property). It is not difficult to check this answer by multiplying the functions first and then taking the derivative (sometimes this will be hard, thus motivating the product rule):

$$\frac{d}{dx}(3x^2 - 3)(4x^3 - 2x) = \frac{d}{dx}(12x^5 - 6x^3 - 12x^3 + 6x)$$
$$= 60x^4 - 54x^2 + 6.$$

Unlike the product rule, the **quotient rule** is not very intuitive nor easy to remember:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}, \quad g(x) \neq 0.$$

Here is an example using the same component terms:

$$\frac{d}{dx}\left(\frac{3x^2-3}{4x^3-2x}\right) = \frac{(4x^3-2x)\frac{d}{dx}(3x^2-3)-(3x^2-3)\frac{d}{dx}(4x^3-2x)}{(4x^3-2x)(4x^3-2x)}$$

$$= \frac{(4x^3-2x)(6x)-(3x^2-3)(12x^2-2)}{16x^6-16x^4+4x^2}$$

$$= \frac{-6x^4+15x^2-3}{8x^6-8x^4+2x^2}.$$

But since quotients are products where one of the terms is raised to the -1 power, it is generally easier to remember and easier to execute the product rule with this adjustment:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{d}{dx}\left[f(x)g(x)^{-1}\right]$$

$$= f(x)\frac{d}{dx}g(x)^{-1} + g(x)^{-1}\frac{d}{dx}f(x).$$

This would be fine, but we do not yet know how to calculate $\frac{d}{dx}g(x)^{-1}$ in general since there are nested components that are functions of x: $g(x)^{-1} = (4x^3-2x)^{-1}$. Such a calculation requires the ***chain rule.***

The **chain rule** provides a means of differentiating nested functions. In Chapter 1, on page 20, we saw nested functions of the form $f \circ g = f(g(x))$. The case of $g(x)^{-1} = (4x^3-2x)^{-1}$ fits this categorization because the inner function is $g(x) = 4x^3-2x$ and the outer function is $f(u) = u^{-1}$. Typically u is used as a placeholder here to make the point that there is a distinct subfunction. To correctly differentiate such a nested function, we have to account for the actual ***order*** of the nesting relationship. This is done by the chain rule, which is given by

$$\frac{d}{dx}f(g(x)) = f'(g(x))g'(x),$$

provided of course that $f(x)$ and $g(x)$ are both differentiable functions. We can also express this in the other standard notation. If $y = f(u)$ and $u = g(x)$ are both differentiable functions, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx},$$

which may better show the point of the operation. If we think about this in purely fractional terms, it is clear that du cancels out of the right-hand side, making the equality obvious. Let us use this new tool to calculate the function $g(x)^{-1}$ from above ($g(x) = 4x^3 - 2x$):

$$\begin{aligned}\frac{d}{dx}g(x)^{-1} &= (-1)(4x^3 - 2x)^{-2} \times \frac{d}{dx}(4x^3 - 2x)\\ &= \frac{-(12x^2 - 2)}{(4x^3 - 2x)^2}\\ &= \frac{-6x^2 + 1}{8x^6 - 8x^4 + 2x^2}.\end{aligned}$$

So we see that applying the chain rule ends up being quite mechanical.

In many cases one is applying the chain rule along with other derivative rules in the context of the same problem. For example, suppose we want the derivative

$$\frac{d}{dy}\left(\frac{y^2 - 1}{y^2 + 1}\right)^3,$$

which clearly requires the chain rule. Now define u such that

$$u = \frac{y^2 - 1}{y^2 + 1} \qquad \text{and} \qquad y = u^3,$$

which redefines the problem as

$$\frac{d}{dy}\left(\frac{y^2 - 1}{y^2 + 1}\right)^3 = \frac{dy}{du}\frac{du}{dy}.$$

We proceed mechanically simply by keeping track of the two components:

$$\begin{aligned}
\frac{dy}{du}\frac{du}{dy} &= 3\left(\frac{y^2-1}{y^2+1}\right)^2 \frac{d}{dy}\left[(y^2-1)(y^2+1)^{-1}\right] \\
&= 3\left(\frac{y^2-1}{y^2+1}\right)^2 \left[2y(y^2+1)^{-1} + (y^2-1)(-1)(y^2+1)^{-2}2y\right] \\
&= 3\left(\frac{y^2-1}{y^2+1}\right)^2 \left[2y(y^2+1)^{-1} - 2y(y^2-1)(y^2+1)^{-2}\right] \\
&= 6y\frac{(y^2-1)^2(y^2+1)-(y^2-1)^3}{(y^2+1)^4} \\
&= 6y\frac{(y^2-1)^2\left[(y^2+1)-(y^2-1)\right]}{(y^2+1)^4} \\
&= 12y\frac{(y^2-1)^2}{(y^2+1)^4}.
\end{aligned}$$

Note that another application of the chain rule was required with the inner step because of the term $(y^2+1)^{-1}$. While this is modestly inconvenient, it comes from the more efficient use of the product rule with the second term raised to the -1 power rather than the use of the quotient rule.

★ Example 5.6: Productivity and Interdependence in Work Groups.

We might ask, how does the productivity of workers affect total organizational productivity differently in environments where there is a great amount of interdependence of tasks relative to an environment where tasks are processed independently? Stinchcombe and Harris (1969) developed a mathematical model to explain such productivity differences and the subsequent effect of greater supervision. Not surprisingly, they found that the effect of additional supervision is greater for work groups that are more interdependent in processing tasks. Define T_1 as the total production when each person's task is independent and T_2 as the total production when *every* task is interdependent. Admittedly, these are extreme cases, but the point is to show differences, so they are likely to be maximally revealing.

In the independent case, the total organizational productivity is

$$T_1 = \sum_{j=1}^{n} bp_j = nb\bar{p},$$

where p_j is the jth worker's performance, b is an efficiency constant for the entire organization that measures how individual performance contributes to total performance, and there are n workers. The notation $\bar{p}$ indicates the average (mean) of all the p_j workers. In the interdependent case, we get instead

$$K\prod_{j=1}^{n} p_j,$$

where K is the total rate of production when everyone is productive. What the product notation (explained on page 12) shows here is that if even one worker is totally unproductive, $p_j = 0$, then the entire organization is unproductive.

Also, productivity is a function of willingness to use ability at each worker's level, so we can define a function $p_j = f(x_j)$ that relates productivity to ability. The premise is that supervision affects this function by motivating people to use their abilities to the greatest extent possible. Therefore we are interested in comparing

$$\frac{\partial T_1}{\partial x_j} \quad \text{and} \quad \frac{\partial T_2}{\partial x_j}.$$

But we cannot take this derivative directly since T is a function of p and p is a function of x. Fortunately the chain rule sorts this out for us:

$$\frac{\partial T_1}{\partial p_j}\frac{\partial p_j}{\partial x_j} = \frac{T_1}{n\bar{p}}\frac{\partial p_j}{\partial x_j}$$

$$\frac{\partial T_2}{\partial p_j}\frac{\partial p_j}{\partial x_j} = \frac{T_2}{p_j}\frac{\partial p_j}{\partial x_j}.$$

The interdependent case is simple because $b = T_1/n\bar{p}$ is how much any one of the individuals contributes through performance, and therefore how much the jth worker contributes. The interdependent case comes from dividing out the p_jth productivity from the total product. The key for comparison was that the second fraction in both expressions ($\partial p_j/\partial x_j$) is the same, so the effect of different organizations was completely testable as the first fraction only. Stinchcombe and Harris' subsequent claim was that "the marginal productivity of the jth worker's performance is about n times as great in the interdependent case but varies according to his absolute level of performance" since the marginal for the interdependent case wass dependent on a single worker and the marginal for the independent case was dependent only on the average worker. So those with very low performance were more harmful in the interdependent case than in the independent case, and these are the cases that should be addressed first.

5.4.2 Derivatives of Logarithms and Exponents

We have already seen the use of real numbers as exponents in the derivative process. What if the variable itself is in the exponent? For this form we have

$$\frac{d}{dx}n^x = \log(n)n^x \qquad \text{and} \qquad \frac{d}{dx}n^{f(x)} = \log(n)n^x\frac{df(x)}{dx},$$

where the log function is the natural log (denoted $\log_n()$ or $\ln()$). So the derivative "peels off" a value of n that is "logged." This is especially handy when n is the e value, since

$$\frac{d}{dx}e^x = \log(e)e^x = e^x,$$

meaning that e^x is ***invariant*** to the derivative operation. But for compound functions with $u = f(x)$, we need to account for the chain rule:

$$\frac{d}{dx}e^u = e^u\frac{du}{dx},$$

for u a function of the x.

Relatedly, the derivatives of the logarithm are given by

$$\frac{d}{dx}\log(x) = \frac{1}{x},$$

for the natural log (recall that the default in the social sciences for logarithms where the base is not identified is a natural log, $\log_n$). The chain rule is

$$\frac{d}{dx}\log(u) = \frac{1}{u}\frac{d}{dx}u.$$

More generally, for other bases we have

$$\frac{d}{dx}\log_b(x) = \frac{1}{x\text{ln(b)}}, \text{ because } \log_b(x) = \frac{\text{ln(x)}}{\text{ln(b)}}$$

with the associated chain rule version:

$$\frac{d}{dx}\log_b(u) = \frac{1}{u\text{ln(b)}}\frac{d}{dx}u.$$

As an additional illustrative example, we find the derivative of the logarithmic function $y = \log(2x^2 - 1)$:

$$\begin{aligned}\frac{d}{dx}y &= \frac{d}{dx}\log(2x^2 - 1)\\ &= \frac{1}{2x^2 - 1}\frac{d}{dx}(2x^2 - 1)\\ &= \frac{4x}{2x^2 - 1}.\end{aligned}$$

That was pretty simple actually, so now perform a more complicated calculation. The procedure **logarithmic differentiation** uses the log function to make the process of differentiating a difficult function easier. The basic idea is to log the function, take the derivative of this version, and then compensate back at the last step. Start with the function:

$$y = \frac{(3x^2 - 3)^{\frac{1}{3}}(4x - 2)^{\frac{1}{4}}}{(x^2 + 1)^{\frac{1}{2}}},$$

which would be quite involved using the product rule, the quotient rule, and the chain rule. So instead, let us take the derivative of

$$\log(y) = \frac{1}{3}\log(3x^2 - 3) + \frac{1}{4}\log(4x - 2) - \frac{1}{2}\log(x^2 + 1)$$

(note the minus sign in front of the last term because it was in the denominator). Now we take the derivative and solve on this easier metric using the additive property of derivation rather than the product rule and the quotient rule. The left-hand side becomes $\frac{du}{dy}\frac{dy}{dx} = \frac{1}{y}\frac{d}{dx}y$ since it denotes a function of x. We then proceed as

$$\begin{aligned}
\frac{1}{y}\frac{d}{dx}y &= \frac{1}{3}\left(\frac{1}{3x^2-3}\right)(6x) + \frac{1}{4}\left(\frac{1}{4x-2}\right)(4) - \frac{1}{2}\left(\frac{1}{x^2+1}\right)(2x) \\
&= \frac{2x(4x-2)(x^2+1) + (3x^2-3)(x^2+1) - x(3x^2-3)(4x-2)}{(3x^2-3)(4x-2)(x^2+1)} \\
&= \frac{-x^4+2x^3+20x-10x-3}{(3x^2-3)(4x-2)(x^2+1)}.
\end{aligned}$$

Now multiply both sides by y and simplify:

$$\frac{d}{dx}y = \frac{-x^4+2x^3+20x-10x-3}{(3x^2-3)(4x-2)(x^2+1)} \frac{(3x^2-3)^{\frac{1}{3}}(4x-2)^{\frac{1}{4}}}{(x^2+1)^{\frac{1}{2}}}$$

$$\frac{d}{dx}y = \frac{-x^4+2x^3+20x-10x-3}{(3x^2-3)^{\frac{2}{3}}(4x-2)^{\frac{3}{4}}(x^2+1)^{\frac{3}{2}}}.$$

So this calculation is made much easier due to the logarithm.

★ **Example 5.7: Security Trade-Offs for Arms Versus Alliances**. Sorokin (1994) evaluated the decisions that nations make in seeking security through building their armed forces and seeking alliances with other nations. A standard theory in the international relations literature asserts that nations (states) form alliances predominantly to protect themselves from threatening states (Walt 1987, 1988). Thus they rely on their own armed services as well as the armed services of other allied nations as a deterrence from war. However, as Sorokin pointed out, both arms and alliances are costly, and so states will seek a balance that maximizes the security benefit from necessarily limited resources.

How can this be modeled? Consider a state labeled i and its erstwhile ally labeled j. They each have military capability labeled M_i and M_j, correspondingly. This is a convenient simplification that helps to construct an illustrative model, and it includes such factors as the numbers of soldiers, quantity and quality of military hardware, as well as geographic constraints.

It would be unreasonable to say that just because i had an alliance with j it could automatically count on receiving the full level of M_j support if attacked. Sorokin thus introduced the term $T \in [0:1]$, which indicates the "tightness" of the alliance, where higher values imply a higher probability of country j providing M_j military support or the proportion of M_j to be provided. So $T = 0$ indicates no military alliances whatsoever, and values very close to 1 indicate a very tight military alliance such as the heyday of NATO and the Warsaw Pact.

The variable of primary interest is the amount of security that nation i receives from the combination of their military capability and the ally's capability weighted by the tightness of the alliance. This term is labeled S_i and is defined as

$$S_i = \log(M_i + 1) + T \log(M_j + 1).$$

The logarithm is specified because increasing levels of military capability are assumed to give diminishing levels of security as capabilities rise at higher levels, and the 1 term gives a baseline.

So if $T = 0.5$, then one unit of M_i is equivalent to two units of M_j in security terms. But rather than simply list out hypothetical levels for substantive analysis, it would be more revealing to obtain the **marginal effects** of each variable, which are the ***individual*** contributions of each term. There are three quantities of interest, and we can obtain marginal effect equations for each by taking three individual first derivatives that provide the instantaneous rate of change in security at chosen levels.

Because we have three variables to keep track of, we will use slightly different notation in taking first derivatives. The **partial derivative notation** replaces "d" with "∂" but performs exactly the same operation. The replacement is just to remind us that there are other random quantities in the equation and we have picked just one of them to differentiate with this particular expression (more on this in Chapter 6). The three marginal effects from the security equation are given by

$$\text{marginal effect of } M_i : \quad \frac{\partial S_i}{\partial M_i} = \frac{1}{1+M_i} > 0$$

$$\text{marginal effect of } M_j : \quad \frac{\partial S_i}{\partial M_j} = \frac{T}{1+M_j} > 0$$

$$\text{marginal effect of } T : \quad \frac{\partial S_i}{\partial T} = \log(1+M_j) \geq 0.$$

What can we learn from this? The marginal effects of M_i and M_j are declining with increases in level, meaning that the rate of increase in security decreases. This shows that adding more men and arms has a diminishing effect, but this is exactly the motivation for seeking a mixture of arms under national command and arms from an ally since limited resources will then necessarily leverage more security. Note also that the marginal effect of M_j includes the term T. This means that this marginal effect is defined only at levels of tightness, which makes intuitive sense as well. Of course the reverse is also true since the marginal effect of T depends as well on the military capability of the ally.

5.4.3 L'Hospital's Rule

The early Greeks were wary of zero and the Pythagoreans outlawed it. Zero causes problems. In fact, there have been times when zero was considered an "evil" number (and ironically other times when it was considered proof of the existence of god). One problem, already mentioned, caused by zero is when it ends up in the denominator of a fraction. In this case we say that the fraction is "undefined," which sounds like a nonanswer or some kind of a dodge. A certain conundrum in particular is the case of $0/0$. The seventh-century Indian mathematician Brahmagupta claimed it was zero, but his mathematical heirs, such as Bhaskara in the twelfth century, believed that $1/0$ must be infinite and yet it would be only one unit away from $0/0 = 0$, thus producing a paradox.

Fortunately for us calculus provides a means of evaluating the special case of $0/0$.

Assume that $f(x)$ and $g(x)$ are differentiable functions at a where $f(a) = 0$ and $g(a) = 0$. L'Hospital's rule states that

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)},$$

provided that $g'(x) \neq 0$. In plainer words, the limit of the ratio of their two functions is equal to the limit of the ratio of the two derivatives. Thus, even if the original ratio is not interpretable, we can often get a result from the ratio of the derivatives. Guillaume L'Hospital was a wealthy French aristocrat who studied under Johann Bernoulli and subsequently wrote the world's first calculus textbook using correspondence from Bernoulli. L'Hospital's rule is thus misnamed for its disseminator rather than its creator.

As an example, we can evaluate the following ratio, which produces $0/0$ at the point 0:

$$\lim_{x \to 0} \frac{x}{\log(1-x)} = \lim_{x \to 0} \frac{\frac{d}{dx} x}{\frac{d}{dx} \log(1-x)}$$

$$= \lim_{x \to 0} \frac{1}{\frac{1}{-(1-x)}} = -1.$$

L'Hospital's rule can also be applied for the form ∞/∞: Assume that $f(x)$ and $g(x)$ are differentiable functions at a where $f(a) = \infty$ and $g(a) = \infty$; then again $\lim\limits_{x \to a} f(x)/g(x) = \lim\limits_{x \to a} f'(x)/g'(x)$.

Here is an example where this is handy. Note the repeated use of the product rule and the chain rule in this calculation:

$$\lim_{x \to \infty} \frac{(\log(x))^2}{x^2 \log(x)} = \lim_{x \to \infty} \frac{\frac{d}{dx}(\log(x))^2}{\frac{d}{dx} x^2 \log(x)}$$

$$= \lim_{x \to \infty} \frac{2\log(x)\frac{1}{x}}{2x\log(x) + x^2\frac{1}{x}}$$

$$= \lim_{x \to \infty} \frac{\log(x)}{x^2\log(x) + \frac{1}{2}x^2}.$$

It seems like we are stuck here, but we can actually apply L'Hospital's rule again, so after the derivatives we have

$$= \lim_{x\to\infty} \frac{\frac{1}{x}}{2x\log(x) + x^2\frac{1}{x} + x}$$

$$= \lim_{x\to\infty} \frac{1}{2x^2(\log(x)+1)} = 0.$$

★ **Example 5.8: Analyzing an Infinite Series for Sociology Data**. Peterson (1991) wrote critically about sources of bias in models that describe durations: how long some observed phenomena lasts [also called hazard models or event-history models; see Box-Steffensmeier and Jones (2004) for a review]. In his appendix he claimed that the series defined by

$$a_{j,i} = j_i \times \exp(-\alpha j_i), \qquad \alpha > 0, j_i = 1, 2, 3, \ldots,$$

goes to zero in the limit as j_i continues counting to infinity. His evidence is the application of L'Hospital's rule twice:

$$\lim_{j_i\to\infty} \frac{j_i}{\exp(\alpha j_i)} = \lim_{j_i\to\infty} \frac{1}{\alpha\exp(\alpha j_i)} = \lim_{j_i\to\infty} \frac{0}{\alpha^2\exp(\alpha j_i)}.$$

Did we need the second application of L'Hospital's rule? It appears not, because after the first iteration we have a constant in the numerator and positive values of the increasing term in the denominator. Nonetheless, it is no less true and pretty obvious after the second iteration.

5.4.4 Applications: Rolle's Theorem and the Mean Value Theorem

There are some interesting consequences for considering derivatives of functions that are over bounded regions of the x-axis. These are stated and explained here without proof because they are standard results.

Rolle's Theorem:

- Assume a function $f(x)$ that is continuous on the closed interval $[a\!:\!b]$ and differentiable on the open interval $(a\!:\!b)$. Note that it would be unreasonable to require differentiability at the endpoints.
- $f(a) = 0$ and $f(b) = 0$.
- Then there is guaranteed to be at least one point $\hat{x}$ in $(a\!:\!b)$ such that $f'(\hat{x}) = 0$.

Think about what this theorem is saying. A point with a zero derivative is a minima or a maxima (the tangent line is flat), so the theorem is saying that if the endpoints of the interval are both on the x-axis, then there must be one or more points that are modes or anti-modes. Is this logical? Start at the point $[a, 0]$. Suppose from there the function increased. To get back to the required endpoint at $[b, 0]$ it would have to "turn around" somewhere above the x-axis, thus guaranteeing a maximum in the interval. Suppose instead that the function left $[a, 0]$ and decreased. Also, to get back to $[b, 0]$ it would have to also turn around somewhere below the x-axis, now guaranteeing a minimum. There is one more case that is pathological (mathematicians love reminding people about these). Suppose that the function was just a flat line from $[a, 0]$ to $[b, 0]$. Then every point is a maxima and Rolle's Theorem is still true. Now we have exhausted the possibilities since the function leaving either endpoint has to either increase, decrease, or stay the same. So we have just provided an informal proof! Also, we have stated this theorem for $f(a) = 0$ and $f(b) = 0$, but it is really more general and can be restated for $f(a) = f(b) = k$, with any constant k

Mean Value Theorem:

- Assume a function $f(x)$ that is continuous on the closed interval $[a\!:\!b]$ and differentiable on the open interval $(a\!:\!b)$.
- There is now guaranteed to be at least one point $\hat{x}$ in $(a\!:\!b)$ such that $f(b) - f(a) = f'(\hat{x})(b - a)$.

This theorem just says that between the function values at the start and finish of the interval there will be an "average" point. Another way to think about this is to rearrange the result as

$$\frac{f(b) - f(a)}{b - a} = f'(\hat{x})$$

so that the left-hand side gives a slope equation, rise-over-run. This says that the line that connects the endpoints of the function has a slope that is equal to the derivative somewhere inbetween. When stated this way, we can see that it comes from Rolle's Theorem where $f(a) = f(b) = 0$.

Both of these theorems show that the derivative is a fundamental procedure for understanding polynomials and other functions. Remarkably, derivative calculus is a relatively recent development in the history of mathematics, which is of course a very long history. While there were glimmers of differentiation and integration prior to the seventeenth century, it was not until Newton, and independently Leibniz, codified and integrated these ideas that calculus was born. This event represents a dramatic turning point in mathematics, and perhaps in human civilization as well, as it lead to an explosion of knowledge and understanding. In fact, much of the mathematics of the eighteenth and early nineteenth centuries was devoted to understanding the details and implications of this new and exciting tool. We have thus far visited one-half of the world of calculus by looking at derivatives; we now turn our attention to the other half, which is integral calculus.

5.5 Understanding Areas, Slices, and Integrals

One of the fundamental mathematics problems is to find the area "under" a curve, designated by R. By this we mean the area below the curve given by a smooth, bounded function, $f(x)$, and above the x-axis (i.e., $f(x) \geq 0,\ \forall x \in [a\!:\!b]$). This is illustrated in Figure 5.5. Actually, this characterization is a bit too restrictive because other areas in the coordinate axis can also be measured and we will want to treat unbounded or discontinuous areas as well, but we will stick with this setup for now. **Integration** is a calculus procedure for measuring areas and is as fundamental a process as differentiation.

5.5.1 Riemann Integrals

So how would we go about measuring such an area? Here is a really mechanical and fundamental way. First "slice up" the area under the curve with a set of bars that are approximately as high as the curve at different places. This would then be somewhat like a histogram approximation of R where we simply sum up the sizes of the set of rectangles (a very easy task). This method is sometimes referred to as the **rectangle rule** but is formally called **Riemann integration**. It is the simplest but least accurate method for numerical integration. More formally, define n disjoint intervals along the x-axis of width $h = (b-a)/n$ so that the lowest edge is $x_0 = a$, the highest edge is $x_n = b$, and for $i = 2, \ldots, n-1$, $x_i = a + ih$, produces a histogram-like approximation of R. The key point is that for the ith bar the approximation of $f(x)$ over h is $f(a+ih)$.

The only wrinkle here is that one must select whether to employ "left" or "right" Riemann integration:

$$h\sum_{i=0}^{n-1} f(a+ih), \quad \text{left Riemann integral}$$

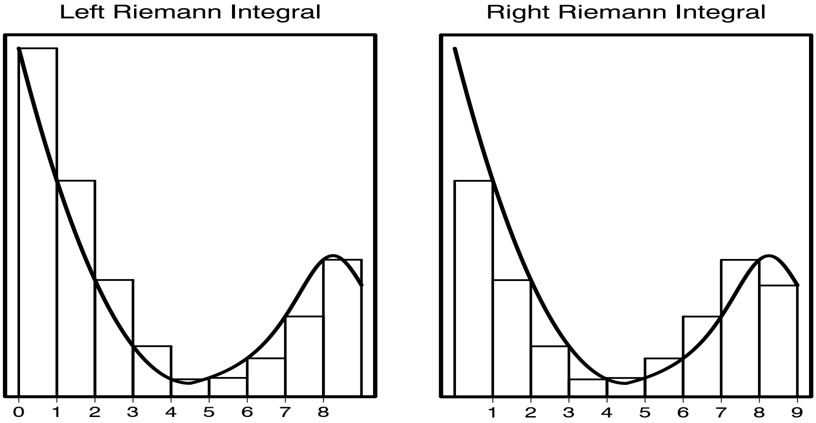
$$h\sum_{i=1}^{n} f(a+ih), \quad \text{right Riemann integral,}$$

determining which of the top corners of the bars touches the curve. Despite the obvious roughness of approximating a smooth curve with a series of rectangular bars over regular bins, Riemann integrals can be extremely useful as a crude starting point because they are easily implemented.

Figure 5.5 shows this process for both left and right types with the different indexing strategies for i along the x-axis for the function:

$$p(\theta) = \begin{cases} (6-\theta)^2/200 + 0.011 & \text{for } \theta \in [0:6) \\ \mathcal{C}(11,2)/2 & \text{for } \theta \in [6:12], \end{cases}$$

Fig. 5.5. Riemann Integration



where $\mathcal{C}(11,2)$ denotes the Cauchy (distribution) function for $\theta = 11$ and $\sigma = 2$: $\mathcal{C}(x|\theta,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}$, $-\infty < x, \theta < \infty, 0 < \sigma$.

It is evident from the two graphs that when the function is downsloping, as it is on the left-hand side, the left Riemann integral overestimates and the right Riemann integral underestimates. Conversely when the function is upsloping, as it is toward the right-hand-side, the left Riemann integral underestimates and the right Riemann integral overestimates. For the values given, the left Riemann integral is too large because there is more downsloping in the bounded region, and the right Riemann integral is too small correspondingly. There is a neat theorem that shows that the actual value of the area for one of these regions is *bounded* by the left and right Riemann integrals. Therefore, the *true* area under the curve is bounded by the two estimates given.

Obviously, because of the inaccuracies mentioned, this is not the best procedure for general use. The value of the left Riemann integral is 0.7794 and the value of the right Riemann integral is 0.6816 for this example, and such a discrepancy is disturbing. Intuitively, as the number of bars used in this process increases, the smaller the regions of curve that we are under- or overestimating. This suggests making the width of the bars (h) very small to improve accuracy.

Such a procedure is always possible since the x-axis is the real number line, and we know that there are an infinite number of places to set down the bars.

It would be very annoying if every time we wanted to measure the area under a curve defined by some function we had to create lots of these bars and sum them up. So now we can return to the idea of limit. As the number of bars increases over a bounded area, then necessarily the width of the bars decreases. So let the width of the bars go to zero in the limit, forcing an infinite number of bars. It is not technically necessary, but continue to assume that all the bars are of equal size, so this limit result holds easily. We now need to be more formal about what we are doing.

For a continuous function $f(x)$ bounded by a and b, define the following limits for left and right Riemann integrals:

$$S_{\text{left}} = \lim_{h \to 0} h \sum_{i=0}^{n-1} f(a + ih)$$

$$S_{\text{right}} = \lim_{h \to 0} h \sum_{i=1}^{n} f(a + ih),$$

where n is the number of bars, h is the width of the bars (and bins), and nh is required to cover the domain of the function, $b - a$. For every subregion the left and right Riemann integrals bound the truth, and these bounds necessarily get progressively tighter approaching the limit. So we then know that

$$S_{\text{left}} = S_{right} = R$$

because of the effect of the limit. This is a wonderful result: The limit of the Riemann process is the true area under the curve. In fact, there is specific terminology for what we have done: The **definite integral** is given by

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

where the $\int$ symbol is supposed to look somewhat like an "S" to remind us that this is really just a special kind of sum. The placement of a and b indicate the lower and upper limits of the definite integral, and $f(x)$ is now called the **integrand**. The final piece, dx, is a reminder that we are summing over infinitesimal values of x. So while the notation of integration can be intimidating to the uninitiated, it really conveys a pretty straightforward idea.

The integral here is called “definite” because the limits on the integration are defined (i.e., having specific values like a and b here). Note that this use of the word ***limit*** applies to the range of application for the integral, not a limit, in the sense of limiting functions studied in Section 5.2.

5.5.1.1 *Application: Limits of a Riemann Integral*

Suppose that we want to evaluate the function $f(x) = x^2$ over the domain $[0\!:\!1]$ using this methodology. First divide the interval up into h slices each of width $1/h$ since our interval is 1 wide. Thus the region of interest is given by the limit of a left Riemann integral:

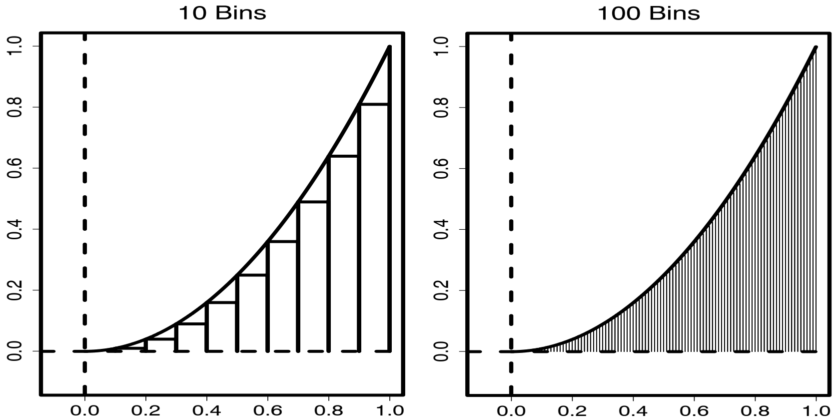
$$\begin{aligned}
R &= \lim_{h\to\infty} \sum_{i=1}^{h} \frac{1}{h} f(x)^2 = \lim_{h\to\infty} \sum_{i=1}^{h} \frac{1}{h}(i/h)^2 \\
&= \lim_{h\to\infty} \frac{1}{h^3} \sum_{i=1}^{h} i^2 = \lim_{h\to\infty} \frac{1}{h^3} \frac{h(h+1)(2h+1)}{6} \\
&= \lim_{h\to\infty} \frac{1}{6}(2 + \frac{3}{h} + \frac{1}{h^2}) = \frac{1}{3}.
\end{aligned}$$

The step out of the summation was accomplished by a well-known trick. Here it is with a relative, stated generically:

$$\sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}, \qquad \sum_{x=1}^{n} x = \frac{n(n+1)}{2}.$$

This process is shown in Figure 5.6 using left Riemann sums for 10 and 100 bins over the interval to highlight the progress that is made in going to the limit. Summing up the bin heights and dividing by the number of bins produces 0.384967 for 10 bins and 0.3383167 for 100 bins. So already at 100 bins we are fitting the curve reasonably close to the true value of one-third.

Fig. 5.6. RIEMANN SUMS FOR $f(x) = x^2$ OVER [0:1]



5.6 The Fundamental Theorem of Calculus

We start this section with some new definitions. In the last section principles of Riemann integration were explained, and here we extend these ideas. Since both the left and the right Riemann integrals produce the correct area in the limit as the number of $h_i = (x_i - x_{i-1})$ goes to infinity, it is clear that some point in between the two will also lead to convergence. Actually, it is immaterial which point we pick in the closed interval, due to the effect of the limiting operation. For slices $i = 1$ to H covering the full domain of $f(x)$, define the point $\hat{x}_i$ as an arbitrary point in the ith interval $[x_{i-1}:x_i]$. Therefore,

$$\int_a^b f(x)dx = \lim_{h\to 0} \sum_{i=1}^{H} f(\hat{x})h_i,$$

and this is now called a **Riemann sum** as opposed to a Riemann integral.

We need one more definition before proceeding. The process of taking a derivative has an opposite, the **antiderivative**. The antiderivative corresponding to a specific derivative takes the equation form back to its previous state. So, for example, if $f(x) = \frac{1}{3}x^3$ and the derivative is $f'(x) = x^2$, then the antiderivative of the function $g(x) = x^2$ is $G(x) = \frac{1}{3}x^3$. Usually antiderivatives are designated with a capital letter. Note that the derivative of the antiderivative returns the original form: $F'(x) = f(x)$.

The antiderivative is a function in the regular sense, so we can treat it as such and apply the Mean Value Theorem discussed on page 202 for a single bin within the interval:

$$\begin{aligned} F(x_i) - F(x_{i-1}) &= F'(\hat{x}_i)(x_i - x_{i-1}) \\ &= f(\hat{x}_i)(x_i - x_{i-1}). \end{aligned}$$

The second step comes from the fact that the derivative of the antiderivative returns the function back. Now let us do this for ***every*** bin in the interval, assuming H bins:

$$\begin{aligned} F(x_1) - F(a) &= f(\hat{x}_1)(x_1 - a) \\ F(x_2) - F(x_1) &= f(\hat{x}_2)(x_2 - x_1) \\ F(x_3) - F(x_2) &= f(\hat{x}_3)(x_3 - x_2) \\ &\vdots \\ F(x_{H-1}) - F(x_{H-2}) &= f(\hat{x}_{H-1})(x_{H-1} - x_{H-2}) \\ F(x_b) - F(x_{H-1}) &= f(\hat{x}_b)(x_b - x_{H-1}). \end{aligned}$$

In adding this series of H equations, something very interesting happens on the left-hand side:

$$(F(x_1) - F(a)) + (F(x_2) - F(x_1)) + (F(x_3) - F(x_2)) + \cdots$$

can be rewritten by collecting terms:

$$-F(a) + (F(x_1) - F(x_1)) + (F(x_2) - F(x_2)) + (F(x_3) - F(x_3)) + \cdots$$

It "accordions" in the sense that there is a term in each consecutive parenthetical quantity from an individual equation that cancels out part of a previous parenthetical quantity: $F(x_1) - F(x_1)$, $F(x_2) - F(x_2)$,... $F(x_{H-1}) - F(x_{H-1})$. Therefore the only two parts left are those corresponding to the endpoints, $F(a)$ and $F(b)$, which is a great simplification! The right-hand side addition looks like

$$f(\hat{x}_1)(x_1 - a) + f(\hat{x}_2)(x_2 - x_1) + f(\hat{x}_3)(x_3 - x_2) + \\ \cdots + f(\hat{x}_{H-1})(x_{H-1} - x_{H-2}) + f(\hat{x}_b)(x_b - x_{H-1}),$$

which is just $\int_a^b f(x)dx$ from above. So we have now stumbled onto the **Fundamental Theorem of Calculus**:

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

which simply says that integration and differentiation are opposite procedures: an integral of $f(x)$ from a to b is just the antiderivative at b minus the antiderivative at a. This is really important theoretically, but it is also really important computationally because it shows that we can integrate functions by using antiderivatives rather than having to worry about the more laborious limit operations.

5.6.1 Integrating Polynomials with Antiderivatives

The use of antiderivatives for solving definite integrals is especially helpful with polynomial functions. For example, let us calculate the following definite integral:

$$\int_1^2 (15y^4 + 8y^3 - 9y^2 + y - 3)dy.$$

The antiderivative is

$$F(y) = 3y^5 + 2y^4 - 3y^3 + \frac{1}{2}y^2 - 3y,$$

since

$$\frac{d}{dy}F(y) = \frac{d}{dy}(3y^5 + 2y^4 - 3y^3 + \frac{1}{2}y^2 - 3y) = 15y^4 + 8y^3 - 9y^2 + y - 3.$$

Therefore,

$$\int_1^2 (15y^4 + 8y^3 - 9y^2 + y - 3)dy$$

$$= 3y^5 + 2y^4 - 3y^3 + \frac{1}{2}y^2 - 3y\Big|_{y=1}^{y=2}$$

$$= (3(2)^5 + 2(2)^4 - 3(2)^3 + \frac{1}{2}(2)^2 - 3(2))$$

$$- (3(1)^5 + 2(1)^4 - 3(1)^3 + \frac{1}{2}(1)^2 - 3(1))$$

$$= (96 + 32 - 24 + 2 - 6) - (3 + 2 - 3 + \frac{1}{2} - 3) = 100.5.$$

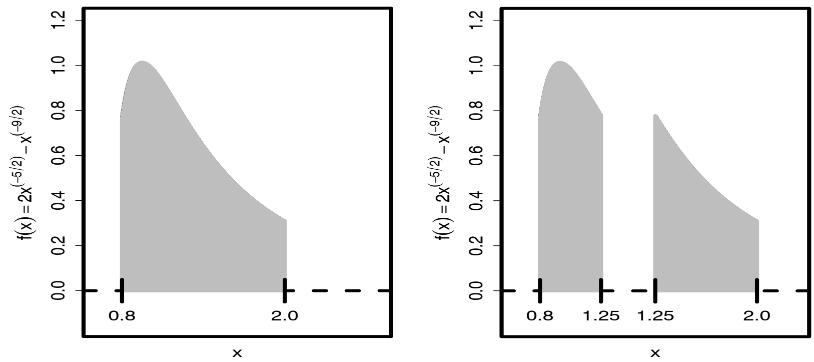
The notation for substituting in limit values, $\Big|_{y=a}^{y=b}$, is shortened here to $\Big|_a^b$, since the meaning is obvious from the dy term (the distinction is more important in the next chapter, when we study integrals of more than one variable).

Now we will summarize the basic properties of definite integrals.

Properties of Definite Integrals

$\rightarrowtail$	Constants	$\int_a^b kf(x)dx = k\int_a^b f(x)dx$
$\rightarrowtail$	Additive Property	$\int_a^b (f(x) + g(x))dx$ $= \int_a^b f(x)dx + \int_a^b g(x)dx$
$\rightarrowtail$	Linear Functions	$\int_a^b (k_1 f(x) + k_2 g(x))dx$ $= k_1\int_a^b f(x)dx + k_2\int_a^b g(x)dx$
$\rightarrowtail$	Intermediate Values	for $a \leq b \leq c$: $\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$
$\rightarrowtail$	Limit Reversibility	$\int_a^b f(x)dx = -\int_b^a f(x)dx$

Fig. 5.7. Integrating by Pieces



The first two properties are obvious by now and the third is just a combination of the first two. The fourth property is much more interesting. It says that we can split up the definite integral into two pieces based on some intermediate value between the endpoints and do the integration separately. Let us now do this with the function $f(x) = 2x^{-5/2} - x^{-9/2}dx$ integrated over $[0.8:2.0]$ with an intermediate point at 1.25:

$$\int_{0.8}^{2.0} 2x^{-\frac{5}{2}} - x^{-\frac{9}{2}}dx = \int_{0.8}^{1.25} 2x^{-\frac{5}{2}} - x^{-\frac{9}{2}}dx + \int_{1.25}^{2.0} 2x^{-\frac{5}{2}} - x^{-\frac{9}{2}}dx$$

$$= \left[\left(-\frac{2}{3}\right)2x^{-\frac{3}{2}} - \left(-\frac{2}{7}\right)x^{-\frac{7}{2}}\right]\Bigg|_{0.8}^{1.25}$$

$$+ \left[\left(-\frac{2}{3}\right)2x^{-\frac{3}{2}} - \left(-\frac{2}{7}\right)x^{-\frac{7}{2}}\right]\Bigg|_{1.25}^{2.0}$$

$$= [-0.82321 - (-1.23948)]$$

$$+ [-0.44615 - (-0.82321)] = 0.79333.$$

This is illustrated in Figure 5.7. This technique is especially handy where it is difficult to integrate the function in one piece (the example here is therefore somewhat artificial). Such cases occur when there are discontinuities or pieces of the area below the x-axis.

★ **Example 5.9: The Median Voter Theorem**. The simplest, most direct analysis of the aggregation of vote preferences in elections is the Median Voter Theorem. Duncan Black's (1958) early article identified the role of a specific voter whose position in a single issue dimension is at the median of other voters' preferences. His theorem roughly states that if all of the voters' preference distributions are unimodal, then the median voter will always be in the winning majority. This requires two primary restrictions. There must be a single issue dimension (unless the same person is the median voter in all relevant dimensions), and each voter must have a unimodal preference distribution. There are also two other assumptions generally of a less-important nature: All voters participate in the election, and all voters express their true preferences (sincere voting). There is a substantial literature that evaluates the median voter theorem after altering any of these assumptions [see Dion (1992), for example].

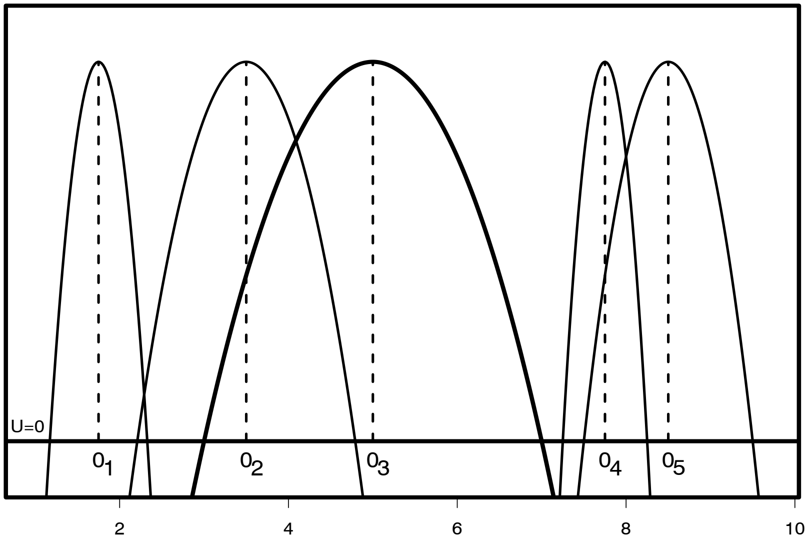
The Median Voter Theorem is displayed in Figure 5.8, which is a reproduction of Black's figure (1958, p.15). Shown are the preference curves for five hypothetical voters on an interval measured issue space (the x-axis), where the utility goes to zero at two points for each voter (one can also assume that the utility curves asymptotically approach zero as Black did). In the case given here it is clear that the voter with the mode at O_3 is the median voter for this system, and there is some overlap with the voter whose mode is at O_2. Since overlap represents some form of potential agreement, we might be interested in measuring this area.

These utility functions are often drawn or assumed to be parabolic shapes. The general form used here is

$$f(x) = 10 - (\mu_i - x)^2\omega_i,$$

where μ_i determines this voter's modal value and ω_i determines how fast their

Fig. 5.8. Placing the Median Voter



utility diminishes moving away from the mode (i.e., how "fat" the curve is for this voter). For the two voters under study, the utility equations are therefore

$$V_2\colon\ f(x) = 10 - (3.5 - x)^2(6) \qquad V_3\colon\ f(x) = 10 - (5 - x)^2(2.5),$$

so smaller values of ω produce more spread out utility curves. This is a case where we will need to integrate the area of overlap in two pieces, because the functions that define the area from above are different on either side of the intersection point.

The first problem encountered is that we do not have any of the integral limits: the points where the parabolas intersect the x-axis (although we only need two of the four from looking at the figure), and the point where the two parabolas intersect. To obtain the latter we will equate the two forms and solve with the quadratic equation from page 33 (striking out the 10's and the multiplication by -1 here since they will cancel each other anyway). First expand the squares:

$$y = (3.5 - x)^2(6) \qquad y = (5 - x)^2(2.5)$$

$$y = 6x^2 - 42x + 73.5 \qquad y = 2.5x^2 - 25x + 62.5.$$

Now equate the two expressions and solve:

$$6x^2 - 42x + 73.5 = 2.5x^2 - 25x + 62.5$$

$$3.5x^2 - 17x + 11 = 0$$

$$x = \frac{-(-17) \pm \sqrt{(-17)^2 - 4(3.5)(11)}}{2(3.5)} = 4.088421 \text{ or } 0.768721.$$

This is a quadratic form, so we get two possible answers. To find the one we want, plug both potential values of x into one of the two parabolic forms and observe the y values:

$$y = 10 - 2.5(5 - (4.0884))^2 = 7.9226$$

$$y = 10 - 2.5(5 - (0.7687))^2 = -34.7593.$$

Because we want the point of intersection that exists above the x-axis, the choice between the two x values is now obvious to make. To get the roots of the two parabolas (the points where $y = 0$), we can again apply the quadratic equation to the two parabolic forms (using the original form):

$$x_2 = \frac{-(42) \pm \sqrt{(42)^2 - 4(-6)(-63.5)}}{2(-6)} = 2.209 \text{ or } 4.79$$

$$x_2 = \frac{-(25) \pm \sqrt{(25)^2 - 4(-2.5)(-52.5)}}{2(-2.5)} = 3 \text{ or } 7.$$

We know that we want the greater root of the first parabola and the lesser root of the second parabola (look at the picture), so we will use 3 and 4.79 as limits on the integrals.

The area to integrate now consists of the following two-part problem, solved by the antiderivative method:

$$A = \int_3^{4.0884} (-6x^2 + 42x - 63.5)dx + \int_{4.0884}^{4.79} (-2.5x^2 + 25x - 52.5)dx$$

$$= \left(-2x^3 + 21x^2 - 63.5x\right)\Big|_3^{4.0884} + \left(-\frac{5}{6}x^3 + \frac{25}{2}x^2 - 52.5x\right)\Big|_{4.0884}^{4.79}$$

$$= ((-45.27343) - (-55.5)) + ((-56.25895) - (-62.65137))$$

$$= 16.619.$$

So we now know the area of the overlapping region above the x-axis between voter 2 and voter 3. If we wanted to, we could calculate the other overlapping regions between voters and compare as a measure of utility similarity on the issue space.

5.6.2 Indefinite Integrals

Indefinite integrals are those that lack specific limits for the integration operation. The consequence of this is that there must be an arbitrary constant (labeled k here) added to the antiderivative to account for the constant component that would be removed by differentiating:

$$\int f(x)dx = F(x) + k.$$

That is, if $F(x) + k$ is the antiderivative of $f(x)$ and we were to calculate $\frac{d}{dx}(F(x) + k)$ with defined limits, then any value for k would disappear. The logic and utility of indefinite integrals is that we use them to relate functions rather than to measure specific areas in coordinate space, and the further study of this is called *differential equations*.

As an example of calculating indefinite integrals, we now solve one simple problem:

$$\int (1-3x)^{\frac{1}{2}}dx = -\frac{2}{9}(1-3x)^{\frac{3}{2}} + k.$$

Instead of a numeric value, we get a defined relationship between the functions $f(x) = (1-3x)^{\frac{1}{2}}$ and $F(x) = -\frac{2}{9}(1-3x)^{\frac{3}{2}}$.

5.6.3 Integrals Involving Logarithms and Exponents

We have already seen that the derivative of exponentials and logarithms are special cases: $\frac{d}{dx}e^x = e^x$ and $\frac{d}{dx}\log(x) = \frac{1}{x}$. For the most part, these are important rules to memorize, particularly in statistical work. The integration process with logarithms and exponents is only slightly more involved.

Recall that the chain rule applied to the exponential function takes the form

$$\frac{d}{dx}e^u = e^u \frac{du}{dx}.$$

This means that the form of the u function remains in the exponent but its derivative comes down. So, for example,

$$\frac{d}{dx}e^{3x^2-x} = e^{3x^2-x}(6x-1),$$

which is simple if one can remember the rule. For integration it is essential to keep track of the "reverse chain rule" that comes from this principle:

$$\int e^u du = e^u + k.$$

This means that the u function must be incorporated into the limit definition to reverse $\frac{du}{dx}$. In addition, we have to add a constant k that could have been there but was lost due to the derivative function operating on it ($\frac{d}{dx}(k) = 0$).

In the following example the function in the exponent is $f(x) = -x$, so we alter the limit multiplying by -1 so that the exponent value and limit are identical and the regular property of e applies:

$$\int_0^2 e^{-x}dx = -\int_0^2 e^{-x}d(-x) = -e^{-x}\Big|_0^2 = -e^{-2} - (-e^0) = 0.8646647.$$

Now consider a more complicated example:

$$\int \frac{e^x}{1+e^x} dx,$$

which seems very difficult until we make a substitution. First define $u = 1+e^x$, which changes the integral to

$$\int \frac{e^x}{u} dx.$$

This does not seem to help us much until we notice that $\frac{d}{dx}(1+e^x) = e^x = du$, meaning that we can make following second substitution:

$$\int \frac{e^x}{1+e^x} dx = \int \frac{du}{u}, \qquad \text{where } \; u = 1+e^x \; \text{ and } \; du = \frac{d}{dx}(1+e^x) = e^x.$$

So the seemingly difficult integral now has a very simple antiderivative (using the rule $\frac{d}{dy}\log(y) = 1/y$), which we can perform and then substitute back to the quantity of interest:

$$\int \frac{e^x}{1+e^x} dx = \int \frac{du}{u} = \log(u) + k = \log(1+e^x) + k.$$

What this demonstrates is that the rules governing exponents and logarithms for derivatives can be exploited when going in the opposite direction. When these sorts of substitutions are less obvious or the functions are more complicated, the next tool required is ***integration by parts***.

5.6.4 Integration by Parts

So far we have not developed a method for integrating functions that are products, although we did see that differentiating products is quite straightforward. Suppose we have an integral of the form

$$\int f(x)g(x)dx.$$

Often it is not always easy to see the structure of the antiderivative here. We will now derive a method, **integration by parts**, that gives a method for unwinding the product rule. The trick is to recharacterize part of the function into the $d()$ argument.

We will start with a basic derivation. Suppose first that we label $f(x) = u$ and $g(x) = v$, and note the shorthand versions of the derivatives $\frac{d}{dx}v = v'$ and $\frac{d}{dx}u = u'$. We can also rewrite these expressions as

$$dv = v'dx \qquad\qquad du = u'dx,$$

which will prove to be convenient. So the product rule is given by

$$\frac{d}{dx}(uv) = u\frac{d}{dx}v + v\frac{d}{dx}u = uv' + vu'.$$

Now integrate both sides of this equation with respect to x; simplify the left-hand side:

$$\int \frac{d}{dx}(uv)dx = \int uv'dx + \int vu'dx$$

$$uv = \int u(v'dx) + \int v(u'dx),$$

and plug in the definitions $dv = v'dx$, $du = u'dx$:

$$= \int udv + \int vdu.$$

By trivially rearranging this form we get the formula for integration by parts:

$$\int udv = uv - \int vdu.$$

This means that if we can rearrange the integral as a product of the function u and the derivative of another function dv, we can get uv, which is the product of u and the integral of dv minus a new integral, which will hopefully be easier to handle. If the latter integral requires it, we can repeat the process with new terms for u and v. We also need to readily obtain the integral of dv to get uv, so it is possible to choose terms that do not help.

As the last discussion probably foretells, there is some "art" associated with integration by parts. Specifically, how we split up the function to be integrated must be done strategically so that get more simple constituent parts on the right-hand side. Here is the classic textbook example:

$$\int x \log(x) dx,$$

which would be challenging without some procedure like the one described. The first objective is to see how we can split up $x\log(x)dx$ into udv. The two possibilities are

$$[u][dv] = [x][\log(x)]$$

$$[u][dv] = [\log(x)][x],$$

where the choice is clear since we cannot readily obtain $v = \int dvdx = \int \log(x)dx$. So picking the second arrangement gives the full mapping:

$$\boxed{\begin{array}{ll} u = \log(x) & dv = xdx \\ du = \frac{1}{x}dx & v = \frac{1}{2}x^2. \end{array}}$$

This physical arrangement in the box is not accidental; it helps to organize the constituent pieces and their relationships. The top row multiplied together should give the integrand. The second row is the derivative and the antiderivative of each of the corresponding components above. We now have all of the pieces mapped out for the integration by parts procedure:

$$\int udv = uv - \int vdu.$$

$$= (\log(x))\left(\frac{1}{2}x^2\right) - \int \left(\frac{1}{2}x^2\right)\left(\frac{1}{x}dx\right)$$

$$= \frac{1}{2}x^2\log(x) - \int \frac{1}{2}xdx$$

$$= \frac{1}{2}x^2\log(x) - \frac{1}{2}(\frac{1}{2}x^2) + k$$

$$= \frac{1}{2}x^2\log(x) - \frac{1}{4}x^2 + k.$$

We benefited from a very simple integral in the second stage, because the antiderivative of $\frac{1}{2}x$ is straightforward. It can be the case that this integral is *more* difficult than the original one, which means that the choice of function assignment needs to be rethought.

5.6.4.1 *Application: The Gamma Function*

The **gamma function** (also called ***Euler's integral***) is given by

$$\Gamma(\omega) = \int_0^\infty t^{\omega-1}e^{-t}dt, \quad \omega > 0.$$

Here t is a "dummy" variable because it integrates away (it is a placeholder for the limits). The gamma function is a generalization of the factorial function that can be applied to any positive real number, not just integers. For integer values, though, there is the simple relation: $\Gamma(n) = (n-1)!$. Since the result of the gamma function for any given value of ω is finite, the gamma function shows that finite results can come from integrals with limit values that include infinity.

Suppose we wanted to integrate the gamma function for a known value of ω, say 3. The resulting integral to calculate is

$$\int_0^\infty t^2e^{-t}dt.$$

There are two obvious ways to split the integrand into u and dv. Consider this one first:

$$\boxed{\begin{array}{ll} u = e^{-t} & dv = t^2 \\ du = -e^{-t} & v = \frac{1}{3}t^3. \end{array}}$$

The problem here is that we are moving up ladders of the exponent of t, thus with each successive iteration of integration by parts we are actually making the subsequent $\int vdu$ integral ***more*** difficult. So this will not do. The other logical split is

$$\boxed{\begin{array}{ll} u = t^2 & dv = e^{-t} \\ du = 2t & v = -e^{-t}. \end{array}}$$

So we proceed with the integration by parts (omitting the limits on the integral for the moment):

$$\begin{aligned}\Gamma(3) &= uv - \int v du \\ &= (t^2)(-e^{-t}) - \int (-e^{-t})(2t)dt \\ &= -e^{-t}t^2 + 2\int e^{-t}tdt.\end{aligned}$$

Of course we now need to repeat the process to calculate the new integral on the right-hand side, so we will split this new integrand ($e^{-t}t$) up in a similar fashion:

$$\boxed{\begin{array}{ll} u = t & dv = e^{-t} \\ du = 1 & v = -e^{-t}. \end{array}}$$

The antiderivative property of e makes this easy (and repetitive). Now finish the integration:

$$\begin{aligned}\Gamma(3) &= -e^{-t}t^2 + 2\int_0^\infty e^{-t}tdt \\ &= -e^{-t}t^2 + 2\left[(t)(-e^{-t}) - \int_0^\infty (-e^{-t})(1)dt\right] \\ &= \left[-e^{-t}t^2 - 2e^{-t}t - 2e^{-t}\right]\Big|_0^\infty \\ &= [0-0-0] - [0-0-2] = 2,\end{aligned}$$

which makes sense because we know that $\Gamma(3) = 2! = 2$. Observe also that the limits on the function components $-e^{-t}t^2$ and $2e^{-t}t$ needed to be calculated with L'Hospital's rule.

★ **Example 5.10: Utility Functions from Voting**. Back in Chapter 1, in Example 1.3, we saw an important mathematical model of voting developed by Riker and Ordeshook (1968). In that same article they looked at the "utility" that a particular voter receives from having her preferred candidate receive a specific number of votes. Call x the number of votes for her preferred candidate, v the total number of voters (participating), and $u(x)$ the utility for x. Obviously this last term ranges from a maximum of $u(v)$, where all voters select this candidate, to $u(0)$, where no voters select this candidate. For this example, assume that utility increases linearly with proportion, x/v, and so these minimum and maximum utility values are really zero and one.

They next specify a value x_0 that is the respondent's estimate of what might happen in the election when they do *not* vote. There is a function $g(x - x_0)$ that for a single specified value x gives the subjectively estimated probability that her vote changes the outcome. So when $g(x - x_0)$ is high our potential voters feels that there is a reasonable probability that she can affect the election, and vice versa. The expected value of some outcome x is then $u(x)g(x - x_0)$ in the same sense that the expected value of betting **$1** on a fair coin is 50**¢**. Probability and expectation will be covered in detail in Chapter 7, but here it is sufficient to simply talk about them colloquially.

All this means that the expected value for the election before it occurs, given that our potential voter does not vote, is the integral of the expected value where the integration occurs over all possible outcomes:

$$EV = \int_0^v u(x)g(x - x_0)dx = \int_0^v \frac{x}{v} g(x - x_0)dx,$$

which still includes her subjective estimate of x_0, and we have plugged in the linear function $u(x) = x/v$ in the second part. Actually, it might be more appropriate to calculate this with a sum since v is discrete, but with a large value it will not make a substantial difference and the sum would be much harder. This formulation means that the rate of change in utility for a change in x_0 (she votes) is

$$\frac{\partial EV}{\partial x_0} = -\frac{1}{v} \int_0^v x \frac{\partial}{\partial x_0} g(x - x_0)dx,$$

which requires some technical "regularity" conditions to let the derivative pass inside the integral. To solve this integral, integration by parts is necessary where $x = u$ and $dv = \frac{\partial}{\partial x_0} g(x - x_0)dx$. This produces

$$\frac{\partial EV}{\partial x_0} = -\frac{1}{v}(vg(x - x_0)dx - 1).$$

This means that as $g(x-x_0)$ goes to zero (the expectation of actually affecting the election) voting utility simplifies to $1/v$, which returns us to the paradox of participation that resulted from the model on page 5. After developing this argument, Riker and Ordeshook saw the result as a refutation of the linear utility assumption for elections because a utility of $1/v$ fails to account for the reasonable number of people that show up at the polls in large elections.

5.7 Additional Topics: Calculus of Trigonometric Functions

This section contains a set of trigonometry topics that are less frequently used in the social sciences but may be useful as references. As before, readers may elect to skip this section.

5.7.1 Derivatives of Trigonometric Functions

The trigonometric functions do not provide particularly intuitive derivative forms, but fortunately the two main results are incredibly easy to remember. The derivative forms for the sine and cosine function are

$$\tfrac{d}{dx}\sin(x) = \cos(x) \qquad \tfrac{d}{dx}\cos(x) = -\sin(x).$$

So the only difficult part to recall is that there is a change of sign on the derivative of the cosine. Usually, this operation needs to be combined with the chain rule, because it is more common to have a compound function, $u = g(x)$, rather than just x. The same rules incorporating the chain rule are given by

$$\frac{d}{dx}\sin(u) = \cos(u)\frac{d}{dx}u \qquad \frac{d}{dx}\cos(u) = -\sin(u)\frac{d}{dx}u.$$

The derivative forms for the other basic trigonometric functions are given in the same notation by

$$\frac{d}{dx}\csc(u) = -\csc(u)\cot(u)\frac{d}{dx}u \qquad \frac{d}{dx}\sec(u) = \sec(u)\tan(u)\frac{d}{dx}u$$

$$\frac{d}{dx}\tan(u) = \sec^2(u)\frac{d}{dx}u \qquad \frac{d}{dx}\cot(u) = -\csc^2(u)\frac{d}{dx}u.$$

It is better not to worry about memorizing these and most people (other than undergraduates taking math exams) tend to look them up when necessary. Some numerical examples are helpful:

$$\frac{d}{dx}\sin(\frac{1}{2}\pi x) = \frac{1}{2}\pi\cos(\frac{1}{2}\pi x) \qquad \frac{d}{dy}\cot(y^2-y) = -\csc^2(y^2-y)(2y-1).$$

5.7.2 Integrals of Trigonometric Functions

The integrals of the basic trigonometric functions are like the derivative forms: easy to understand, annoying to memorize, and simple to look up (don't sell this book!). They can be given as either definite or indefinite integrals.

The two primary forms are

$$\int \sin(x)dx = -\cos(x) + k \qquad \int \cos(x)dx = \sin(x) + k.$$

It is also important to be able to manipulate these integrals for the reverse chain rule operation, for instance, $\int \sin(u)du = -\cos(u) + k$. The other four basic forms are more complicated. They are (in reverse chain rule notation)

$$\int \csc(u)du = \log|\csc(u) - \cot(u)| + k$$

$$\int \cot(u)du = \log|\sin(u)| + k.$$

$$\int \tan(u)du = -\log|\cos(u)| + k$$

$$\int \sec(u)du = \log|\sec(u) + \tan(u)| + k$$

Eight other fundamental trigonometric formulas are worth noting. These involve products and are given by

$$\int \sin^2(u)du = \frac{1}{2}u - \frac{1}{4}\sin(2u) + k$$

$$\int \cos(u)^2 du = \frac{1}{2}u + \frac{1}{4}\sin(2u) + k$$

$$\int \csc^2(u)du = -\cot(u) + k$$

$$\int \sec^2(u)du = \tan(u) + k$$

$$\int \tan^2(u)du = \tan(u) - u + k$$

$$\int \cot^2(u)du = -\cot(u) - u + k$$

$$\int \csc(u)\cot(u)du = -\csc(u) + k$$

$$\int \sec(u)\tan(u)du = \sec(u) + k.$$

There are actually many, many more of these and reference books abound with integral formulas that range from the basic (such as these) to the truly exotic.

5.8 New Terminology

Exercises

5.1 Find the following finite limits:

$$\lim_{x\to 4}\left[x^2 - 6x + 4\right] \qquad \lim_{x\to 0}\left[\frac{x-25}{x+5}\right]$$

$$\lim_{x\to 4}\left[\frac{x^2}{3x-2}\right] \qquad \lim_{y\to 0}\left[\frac{y^4-1}{y-1}\right].$$

5.2 Given

$$\lim_{y\to 0} f(y) = -2 \quad \text{and} \quad \lim_{y\to 0} g(y) = 2,$$

find

$$\lim_{y\to 0}\left[f(y) - 3f(y)g(y)\right].$$

5.3 Find the following infinite limits and graph:

$$\lim_{x\to\infty}\left[\frac{9x^2}{x^2+3}\right] \qquad \lim_{x\to\infty}\left[\frac{3x-4}{x+3}\right] \qquad \lim_{x\to\infty}\left[\frac{2^x-3}{2^x+1}\right].$$

5.4 Marshall and Stahura (1979) modeled racial mixes in U.S. suburbs over a 10-year period and found that changes between white and black proportions can be described with one of two functional forms: linear and quadratic:

Population Change, 1960 to 1970, White (y) as a Function of Black (x), by Suburb Size

Suburb	$y = 0.330 - 0.024x$
Under 25K	$y = 0.295 - 0.102x + 0.033x^2$
Suburb	$y = 0.199 - 0.112x$
25K–50K	$y = 0.193 - 0.144x + 0.103x^2$
Suburb	$y = 0.100 - 0.045x$
Over 50K	$y = 0.084 - 0.114x + 0.026x^2$

Comparing each suburb size separately, which of these two forms implies the greatest instantaneous change in y at $x = 0.5$? What is the interpretation on the minus sign for each coefficient on x and a positive coefficient for each coefficient on x^2?

5.5 Calculate the following derivatives:

$$\frac{d}{dx}3x^{\frac{1}{3}} \qquad \frac{d}{dt}(14t-7)$$

$$\frac{d}{dy}(y^3+3y^2-12) \qquad \frac{d}{dx}\left(\frac{1}{100}x^{25}-\frac{1}{10}x^{0.25}\right)$$

$$\frac{d}{dx}(x^2+1)(x^3-1) \qquad \frac{d}{dy}(y^3-7)\left(1+\frac{1}{y^2}\right)$$

$$\frac{d}{dy}(y-y^{-1})(y-y^{-2}) \qquad \frac{d}{dx}\left(\frac{4x-12x^2}{x^3-4x^2}\right)$$

$$\frac{d}{dy}\exp[y^2-3y+2] \qquad \frac{d}{dx}\log(2\pi x^2).$$

5.6 Stephan and McMullin (1981) considered transportation issues and time minimization as a determinant of the distribution of county seats in the United States and elsewhere. The key trade-off is this: If territories are too small, then there may be insufficient economic resources to sustain necessary services, and if territories are too large, then travel distances swamp economic advantages from scale. Define s as the average distance to traverse, v as the average speed, h as the total maintenance time required (paid for by the population), and p as the population size. The model for time proposed is $T = s/v + h/p$. Distance is proportional to area, so substitute in $s = g\sqrt{a}$ and $p = ad$, where g is a proportionality constant and a is area. Now find the condition for a that minimizes time by taking the first derivative of T with respect to a, setting it equal to zero, and solving. Show that this is a minimum by taking an additional (second) derivative with respect to a and noting that it must be positive.

5.7 Calculate the derivative of the following function using logarithmic differentiation:

$$y = \frac{(2x^3-3)^{\frac{5}{2}}}{(x^2-1)^{\frac{2}{3}}(9x^2-1)^{\frac{1}{2}}}.$$

5.8 Use L'Hospital's rule to evaluate the following limits:

$$\lim_{x\to 1}\frac{\sin(1-x)}{x^2-1} \qquad \lim_{x\to 0}\frac{x}{\sin(\pi x)}$$

$$\lim_{x\to\infty}\frac{\log(x)}{x} \qquad \lim_{x\to\infty}\frac{3^x}{x^3}$$

$$\lim_{y\to 0}\frac{e^y\sin(y)}{\log(1-y)} \qquad \lim_{y\to 0}\frac{(\sin(y))^2-\sin(y^2)}{y}$$

$$\lim_{y\to\infty}\frac{3e^y}{y^3} \qquad \lim_{x\to\infty}\frac{x\log(x)}{x+\log(x)}.$$

5.9 Using the limit of Riemann integrals, calculate the following integrals. Show steps.

$$R=\int_2^3\sqrt{x}dx \qquad R=\int_1^9\frac{1}{1+y^2}dy$$

$$R=\int_0^{0.1}x^2dx \qquad R=\int_1^9 1dx.$$

5.10 Solve the following definite integrals using the antiderivative method.

$$\int_6^8 x^3dx \qquad \int_1^9 2y^5dy \qquad \int_{-1}^0(3x^2-1)dx$$

$$\int_{-1}^1(14+x^2)dx \qquad \int_1^2\frac{1}{t^2}dt \qquad \int_2^4 e^y dy$$

$$\int_{-1}^2 x\sqrt{1+x^2}dx \qquad \int_{-100}^{100}(x^2-2/x)dx \qquad \int_2^4\sqrt{t}dt.$$

5.11 Calculate the area of the following function that lies above the x-axis and over the domain $[-10, 10]$:

$$f(x) = 4x^2 + 12x - 18.$$

5.12 In Figure 5.8, the fourth and fifth voters have ideal points at 7.75 and 8.5, respectively, and $\omega_4 = 40$, $\omega_5 = 10$. Calculate the area of the overlapping region above the x-axis.

5.13 Calculate the following indefinite integrals:

$$\int (2+12x^2)^{3/2}dx \qquad \int (x^2 - x^{-\frac{1}{2}})dx \qquad \int \frac{y}{\sqrt{9+3y^3}}dy$$

$$\int 360t^6 dt \qquad \int \frac{x^2}{1-x^2}dx \qquad \int (11-21y^9)^2 dy.$$

5.14 Calculate the following integrals using integration by parts:

$$\int x\cos(x)dx \qquad \int x\log(x^2+1)dx \qquad \int e^x x^{-2}dx$$

$$\int 2x\log(x)^2 dx \qquad \int \sec(x)^3 dx \qquad \int xe^{(3x^2-3)}dx.$$

5.15 For the gamma function (Section 5.6.4.1):

(a) Prove that $\Gamma(1) = 1$.

(b) Express the binomial function, $\binom{n}{y} = \frac{n!}{y!(n-y)!}$, in terms of the gamma function instead of factorial notation for integers.

(c) What is $\Gamma\left(\frac{1}{2}\right)$?

5.16 Another version of the factorial function is the **double factorial**, with odd and even versions:

$$(2n)!! = (2n)\cdot(2n-2)\cdots(2n-4)\cdots 6\cdot 4\cdot 2$$
$$(2n-1)!! = (2n-1)\cdot(2n-3)\cdots(2n-5)\cdots 5\cdot 3\cdot 1$$
$$= \pi^{-\frac{1}{2}}2^n\Gamma\left(n+\frac{1}{2}\right).$$

Show that $(2n)!! = 2^n n!$ and $(2n-1)!! = \frac{(2n-1)!}{2^n n!}$.

5.17 Blackwell and Girshick (1954) derived the result below in the context of mixed strategies in game theory. Game theory is a tool in the social sciences where the motivations, strategies, and rewards to hypothesized competing actors are analyzed mathematically to make predictions or explain observed behavior. This idea originated formally with von Neumann and Morgenstern's 1944 book. Simplified, an actor employs a mixed strategy when she has a set of alternative actions each with a known or assumed probability of success, and the choice of actions is made by randomly selecting one with that associated probability. Thus if there are three alternatives with success probabilities $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$, then there is a 50% chance of picking the first, and so on. Blackwell and Girshick (p.54) extended this idea to continuously measured alternatives between zero and one (i.e., a smooth function rather than a discrete listing). The first player accordingly chooses value $x \in [0\!:\!1]$, and the second player chooses value $y \in [0\!:\!1]$, and the function that defines the "game" is given by $M(x, y) = f(x - y)$, where

$$f(t) = \begin{cases} t(1-t), & \text{for } 0 \geq t \geq 1 \\ f(t+1), & \text{for } -1 \geq t \geq 0. \end{cases}$$

In other words, it matters which of x and y is larger in this game. Here is the surprising part. For any fixed value of y (call this y_0), the expected value of the game to the first player is $\frac{1}{6}$. To show this we integrate over a range of alternatives available to this player:

$$\begin{aligned} \int_0^1 M(x, y)dx &= \int_0^1 f(x - y_0)dx \\ &= \underbrace{\int_0^{y_0} f(x - y_0)dx}_{x<y_0} + \underbrace{\int_{y_0}^1 f(x - y_0)dx}_{x>y_0}, \end{aligned}$$

where breaking the integral into two pieces is necessary because the first one contains the case where $-1 \geq t \geq 0$ and the second one contains the case where $0 \geq t \geq 1$. Substitute in the two function values (t or $t + 1$) and integrate over x to obtain exactly $\frac{1}{6}$.

5.18 Calculate the following derivatives of trigonometric functions:

$$\frac{d}{dx}\tan(9x) \qquad \frac{d}{dy}(\cos(3y) - \sin(2y))$$

$$\frac{d}{dx}\csc(\sin(x/2)) \qquad \frac{d}{dx}\cot(1 - x^2)$$

$$\frac{d}{dy}\tan(y)^2 \qquad \frac{d}{dy}\left(\frac{\tan(y) - \sin(y)}{\tan(y) + \sin(y)}\right)^{\frac{1}{3}}$$

$$\frac{d}{dx}\left(9x^2 + \csc(x)^3 - \tan(2x)\right) \qquad \frac{d}{dx}(x + \sec(3x)).$$

5.19 Show that the Mean Value Theorem is a special case of Rolle's Theorem by generalizing the starting and stopping points.

5.20 Calculate $f(x) = x^3$ over the domain $[0:1]$ using limits only (no power rule), as was done for $f(x) = x^2$ on page 208.

5.21 From the appendix to Krehbiel (2000), take the partial derivative of

$$\frac{50((M^2 - M/2 + \delta)^2 + 100(1 - M/2 + \delta) - M)}{M((1 - M/2 + \delta) - (M/2 + \delta))}$$

with respect to M and show that it is decreasing in $M \in (0\!:\!1)$ (i.e., as M increases) for all values $\delta \in (0\!:\!1)$.