Chapter Five

Introduction to Calculus and the Derivative

In our experience, calculus and all things calculus-related prove the most stressful of the topics in this book for those students who have not had prior calculus coursework. We conjecture that this is due to the foreignness of the subject. While probability and linear algebra certainly have some complex concepts one must internalize, much of the routine manipulations students perform in applying these concepts use operations they are used to: addition, multiplication, etc. In contrast, calculus introduces two entirely new operators, the derivative and the integral, each with its own set of rules. Further, these operators are often taught as a lengthy set of rules, leading to stressful rote memorization and little true understanding of what are relatively straightforward concepts, at least as used in most of political science.[1] To try to avoid this, we're going to take a little more time with the topic. In this chapter we will cover the basics of Calculus and the derivative in what we hope is an intuitive manner, saving the rules of its use for the next chapter. If you are working through this chapter as part of a course and are not sure of something, this is the time to ask questions—before you end up trying to take derivatives without having a clear understanding what they *are*. The first section below provides a brief overview of calculus. The second section introduces the derivative informally, and the third provides a formal definition and shows how it works with a few functions.

5.1 A BRIEF INTRODUCTION TO CALCULUS

For our purposes, the primary use of calculus is that it allows us to deal with continuity in a consistent and productive manner. This is likely a useless claim at this point, so let us explain. As we discussed in Chapter 4, a continuous function is one that we can draw without lifting pencil from paper. Intuitively, such a function has no gaps or jumps in it.[2] Such functions are great for lots of applications, as we have noted, but they lead to some problems when we're trying to understand the concept of change. Change within a discrete function is pretty clear: if $f(1) = 1$ and $f(2) = 4$, and if f is not defined between 1 and 2, then we know moving from 1 to 2 results in a change of 3 in the function. Further, that's pretty much all we can say about it.

[1] As with the other topics in the book, calculus has plenty of complexity to it. We'll just be avoiding nearly all of it.

[2] We're going to assume that unless explicitly stated otherwise functions are defined over subsets of the real numbers and in one dimension in this chapter.

What if the function actually is defined between these two points, though? Let $f(x) = x^2$, now and consider $x \in [1, 2]$. For this function, $f(1) = 1$ and $f(2) = 4$, and we can still say that moving from 1 to 2 results in a change of 3, but we can say a lot more. For example, $f(1) = 1$ and $f(1.5) = 2.25$, and moving from 1 to 1.5 results in a change of 1.25, which is less than half the change resulting from moving from 1 to 2. Or, $f(1) = 1$ and $f(1.1) = 1.21$, and moving from 1 to 1.1 results in a change of 0.21. In fact, there's no limit to how many of these changes we can write down. But there is a *limit* to these changes, in the sense we discussed in Chapter 4: we can keep making the second point closer and closer to 1 until, at the limit, it is the same as 1. What is the change in f at this limit? As we'll see shortly, this is the *derivative*, and the study of these objects constitutes *differential calculus*.

Differential calculus thus deals with the study of infinitesimally small changes in a function. As we'll see in Section 3 below, the derivative of the function $f(x) = 3x$ is 3, which is the slope (rate of change) of the line $y = 3x$. This example illustrates the way in which a derivative breaks down functions, removing information about their value at *any* point and providing just the value of the change at that point. In this case, the rate of change is 3 at all x, and this is all the information the derivative gives us. In other words, the derivative provides for us a graph of the marginal rate of change in any variable that we can represent as a continuous function of another variable, with respect to the variable of which it is a function. This is powerful stuff, and we'll use it extensively in Chapter 8 to find maxima and minima of functions, which prove to be terribly important in both computational statistics and formal theory.

But what if we had a derivative, and wanted to build back up a function from it? Well, we'd need to start at the smallest value of x about which we cared and add up all the infinitesimal changes to $f(x)$ that occurred as we increased x from that point. But how do we add infinitesimal things? The answer is the integral, or antiderivative.[3] The integral is a tool for adding infinitesimals, the same way a sum ($\sum$) is a tool for adding discrete quantities. The symbol for an integral ($\int$) even looks like an "S" to help you remember. As we'll see in Chapter 7, integrating the function $f'(x) = 3$ returns the function $f(x) = 3x + C$, where the C is a constant that has to be added because the derivative doesn't contain information about the value of the function at $x = 0$, so we don't know where to start adding infinitesimals.

We go into much more detail about both of these objects in the coming chapters, but before doing so it's fair to ask why we'd want to do so. Calculus is foundational to higher math, so there are many answers, but two will prove particularly important in this book and political science more generally. Consider first a continuous function with a maximum. Intuitively, this means that there are some values of x for which the function increases in value, but for some other values of x the function has to stop doing so; otherwise it wouldn't have

[3]Technically speaking, only the indefinite integral is the antiderivative. We'll talk more about this in Chapter 7.

a maximum. At the point it stops doing so, and possibly *only* at that point, it is no longer increasing, but not yet decreasing either. This means at that point its *instantaneous* rate of change is zero, so the derivative at that point is zero. We'll discuss what instantaneous rate of change means shortly, and the procedure for finding maxima (and minima) in Chapter 8, but the important point is that *derivatives help us maximize (and minimize) functions.*

We put off until Chapter 11 of Part III of this book the most common use of integrals: in probability, specifically continuous probability distributions. There the integral will allow us to understand statistical inference with continuous variables and to compute expected values and expected utilities, which are vital when considering uncertainty of any sort in game theory. Needless to say, being able to make inferences and deal with uncertainty is necessary for quantitative and formal political research, and learning the ways of the integral will certainly pay off down the road.

5.2 WHAT IS THE DERIVATIVE?

Before we get too far ahead of ourselves, though, let's start with the fundamentals. What is a derivative? As we discussed briefly above, the **derivative** is the instantaneous rate of change of a function. That's it. The notation in the next section might be a bit intimidating, but the underlying concept is very straightforward. And it turns out that it can be quite useful to know what the rate of change is. This is true both for constructing theories of politics and for developing statistical techniques to test hypotheses.

You might be thinking that we snuck something in with that word *instantaneous.* Rate of change is certainly straightforward, but what is the *instantaneous* rate of change? Perhaps that is where things get tricky. If things do, in fact, get tricky, then we suppose the word *instantaneous* is the locus of the tricky bit. But we are going to forestall that issue for the moment and focus on *rate of change.*

5.2.1 Discrete Change

Perhaps one is interested in budgetary politics, the rise and decline of political parties, or arms races between countries. All of these topics can be considered with respect to growth rates over time. For example, we might calculate the percentage change from time $t - 1$ (e.g., 2000) to time t (e.g., 2001) in a budgetary expenditure, the seats a political party wins in the legislature, or the nuclear weapons a country builds. Alternatively, we might calculate and then plot the first difference of the series. The **first difference** of a variable is the value of that variable at time t minus the value of that variable at time $t - 1$, and it is a measure of discrete change. Table 5.1, for example, lists the total number of heavy weapons held by China for several years, as well as the annual discrete change (labeled first difference).

Table 5.1: Aggregate Heavy Weapons, China

Year	Total	First Difference
1995	37,095	——
1996	35,747	−1,348
1997	36,910	1,163
1998	37,032	122
1999	36,494	−538
2000	31,435	−5,059
2001	34,281	2,846

Source: SIPRI (`http://www.sipri.org/databases`).

Once we have calculated the first difference, we can calculate the **percentage change** using the formula $\frac{(x_{t+1}-x_t)}{x_t} \times 100\%$, where the subscript t indicates the first observation and the subscript $t + 1$ indicates the second observation.

Why discuss differences and percentage change in a chapter on the derivative? Our purpose is to get you comfortable thinking about rates of change using math with which you are familiar: arithmetic. The derivative is nothing more than a refinement of these ideas. And you guessed it: that refinement involves the word *instantaneous*.

Discrete change, then, is the first difference between two observations. It is a measure of change in a variable across two *discrete* moments in time. It follows that the size of a first difference is going to vary across different temporal scales. For example, we might have a measure of the number of times that the US Department of State lodges a complaint with a foreign government. We could calculate the first difference of that variable across two years, two quarters (i.e., two chunks of three months each), two months, or two weeks. The size of those differences will vary across the scale of discrete time we consider.

5.2.2 Instantaneous Change

The difference $(x_{t+1} - x_t)$ has a limitation: it can only represent the rate of change over two discrete moments in time. What if we want to know the rate of change at a specific moment in time, not over a discrete interval? The difference cannot help us if we want to know the rate of change at a specific point of a function. However, as we discussed in the previous section, if we could evaluate the rate of change at a point on a function by taking the *limit* of the difference as the interval gets smaller and smaller, then that would tell us what the **instantaneous rate of change** was at the point (or moment in time). And that is precisely what the derivative does.

The derivative is defined for a function with respect to a specific variable. We begin with univariate functions but will discuss multivariate functions as

well, very briefly at the end of this chapter and in more depth in Part V of the book. The derivative of $f(x)$ with respect to x tells us the instantaneous rate of change of the function at each point. Just as we can calculate the value of a function for a specific value of x we can calculate the value of the derivative of $f(x)$ for a specific value of x, but it is typically of more interest to have a general representation of the derivative of $f(x)$ over a range of values of x. And it turns out that there are rules for taking derivatives that make it possible for us to determine what that general representation is. We will cover these in the next chapter. First, we flesh out what we mean by instantaneous a bit more, and then, in the next section, formalize these notions in an intuitive fashion, making use of our study of limits in Chapter 4.

5.2.3 Secants and Tangents

Let's return to the linear example $f(x) = 3x$. To figure out the discrete rate of change between any two points on this line we look at the amount of change on the y-axis relative to a particular amount of change on the x-axis. In other words, we compute for two points:

$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{y_2 - y_1}{x_2 - x_1}.$$

This should look familiar, as it is the equation for the slope of a line. For $f(x) = 3x$ and $x_2 = 2$ and $x_1 = 1$, we get $m = \frac{6-2}{2-1} = 3$, so the slope is 3. You can try other values of x to convince yourself that the slope is the same for all values of x. The rate of change between two discrete points is just the slope of the line connecting those two points, known as a **secant**.[4] We can't see it on a plot of a straight line because it overlaps the line (which is why we haven't created a figure to show you the line). However, in Figure 5.1, which plots the function $f(x) = x^2$, we can see the secant drawn between $x = 1$ and $x = 2$—it is above the curve.[5] The discrete change between these two points is 3. Of note, this change is not constant over values of x. The discrete change between $x = 2$ and $x = 3$ is $\frac{9-4}{3-2} = 5 > 3$, so the slope of the secant line between the points corresponding to $x = 2$ and $x = 3$ on the function is $m = 5$. Trying other points should convince you that the secant is increasing as x gets bigger.

This example illustrates that we can get an idea of change just by subtraction, but even in the case of a quadratic function that is familiar we can see that change itself is more complicated. If the rate of change is itself constantly changing, then how can we really speak about change? That's where the derivative and the instantaneous rate of change come in. Consider first the linear function $f(x) = 3x$. Since the secant overlaps the function itself, we can compare the function at closer and closer points ($x = 1$ and $x = 1.5$, $x = 1$ and $x = 1.1$, etc.) and still get a slope of 3 for the secant line. This is true for all differences, and if

[4] A secant is a line that intersects two points on a curve.

[5] Later we will see that this means $f(x) = x^2$ is a convex function.

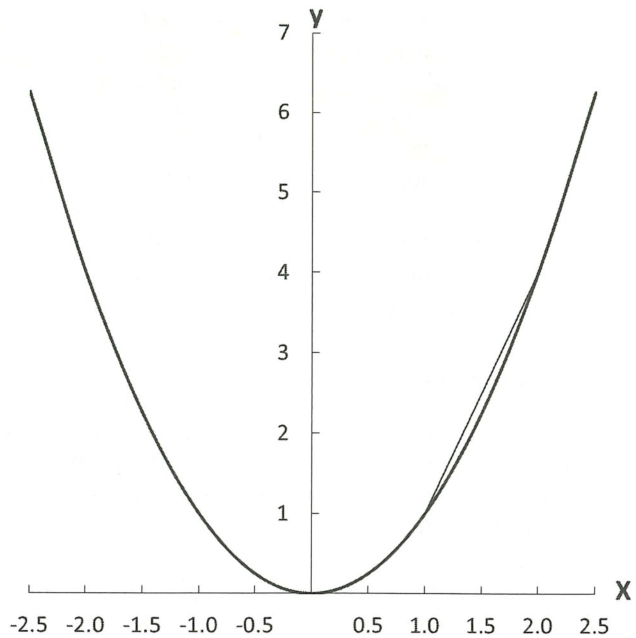

Figure 5.1: Graph of $y = x^2$ with Secant Line

we constructed a sequence of decreasing differences, such as $1, 0.1, 0.01, 0.001, \ldots$, represented by $x_n = \frac{1}{10^n}$, it would be true for every element in the sequence. It's not hard to see that it remains true in the limit as $n \to \infty$. Thus we can say that the instantaneous rate of change at $x = 1$ is also 3. A line that intersects the function at $x = 1$ and has a slope equal to the instantaneous rate of change of the function at $x = 1$ is known as the **tangent**[6] line of the function at the point $x = 1$. We could repeat this analysis to see that all tangent lines of the function have the same slope. In short, we can say the derivative of the function is 3, since this is the instantaneous rate of change at all points on the function.

Now turn one last time to $f(x) = x^2$. Since we already know that the slope of the secant lines change with different values of x, we should expect that changing the distance between points should also change the slopes of these lines. As we can see in Figure 5.2, it does. This plot adds secant lines between 1 and 1.5 and between 1 and 1.1 (which is barely visible) to the one between 1 and 2. As we can see, the slope of the secants decreases as the difference gets less. If this slope approaches some number as the difference approaches zero (i.e., in the limit), then we can talk about the instantaneous rate of change of the function at 1. Plugging numbers into the slope formula, we see that for the three secants we

[6]The tangent line is a line that just touches the curve at a given point.

drew, the slopes are $3, 2.5$, and 2.1, which seem to approach 2. Assuming this holds up under more formal analysis (as we'll see in the next section, it does), then the instantaneous rate of change of the function at $x = 1$ is 2. You can sort of see this graphically too in Figure 5.2: the slopes of the secants get flatter as they approach the tangent line, which has a slope of 2.

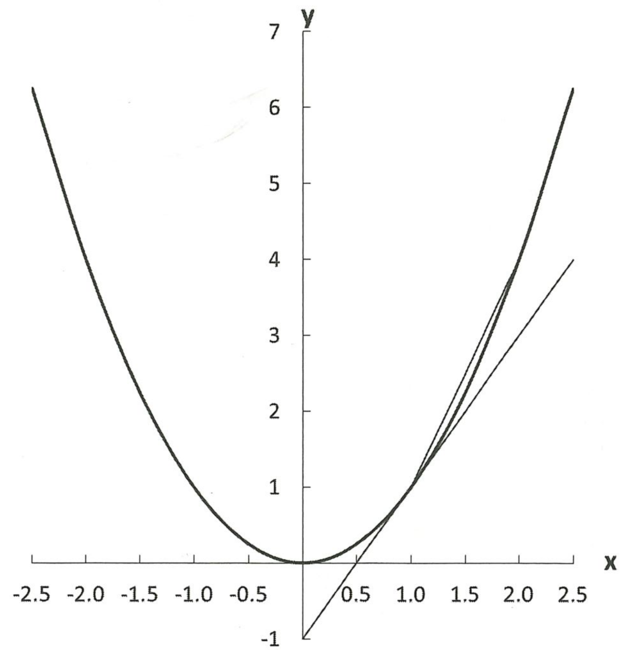

Figure 5.2: Graph of $y = x^2$ with Tangent Line

What about at other points? We could repeat this analysis over and over again, discovering that the slope of the tangent line is 2 at $x = 1$, 4 at $x = 2$, 6 at $x = 3$, and so on, but this is horribly inefficient, and will never give the full picture for the whole function because it's defined over the uncountable real numbers. We might guess that the slope of the tangent line at a point x is equal to $2x$ from the pattern above, and we'd be right in this case, but this would not work for harder problems. If we care about the instantaneous rate of change at all, and we hope we have convinced you that we do, we need a better technique.

5.3 THE DERIVATIVE, FORMALLY

At this point some readers might be readying themselves for the other shoe to drop. Though the graphical representation of secants and tangents might have been intuitive, surely one of the two fundamental pieces of calculus must be much more conceptually difficult. It turns out it's not. Before justifying this claim, however, we need a little notation.

5.3.1 Notation, Notation

Sir Isaac Newton, one of the fathers of calculus, represented a derivative this way: $\dot{y}$ is the derivative of $f(t)$ if $y = f(t)$. This is only used for derivatives with respect to time, however, and you will likely never see it in political science. Instead we turn to the notation of the other father of calculus, Baron Gottfried Wilhelm von Leibniz, who used the following notation to identify the derivative of a function: $\frac{d}{dx}f(x)$, which is read "the derivative of f of x with respect to x." One might say this more cumbersomely, but completely, as the instantaneous rate of change in f of x with respect to x.

If $y = f(x)$, then we can use Leibniz's notation to write the derivative of y with respect to x as $\frac{dy}{dx}$. Leibniz's notation sometimes throws students because it looks like a fraction. But it is not a fraction, so don't try to simplify it by eliminating the d's and concluding that $\frac{dy}{dx} = \frac{y}{x}$ (i.e., that the derivative is the ratio of y to x). That conclusion is *false*.

A benefit of Leibniz's notation is that it allows you to specify the variable with respect to which you are differentiating. This is the variable in the denominator. When this is clear, as it always is when there is only one variable, then we often use Lagrange's prime notation: $f'(x)$ (read "f prime x") is the derivative of $f(x)$. There is no difference between $\frac{d}{dx}f(x)$ and $f'(x)$; they mean the same thing. Finally, you may on occasion see Euler's differential operator D, indicating that the derivative is $Df(x)$. When there is potential confusion about the variable with respect to which the function is being differentiated, then we instead use $D_x f(x)$.

Formal notation isn't the only reference to derivatives that vary. You will sometimes see work that references **differentiation**. This is the process of taking a derivative, and it has the added bonus of a corresponding verb: to differentiate. That is, if one wants to calculate the derivative of $f(x)$, then one differentiates $f(x)$. Again, they mean the same thing.

5.3.2 Limits and Derivatives

Though it is reasonable to be confused at this point by the different ways of representing a derivative, believe us when we say this will pass. Notation is only a way of writing down a concept, and if the concept is clear, one will get used to the notation. We would assume that readers of this book do not have a great deal of trouble using "many" and "a lot," after all, notation that means more or less the same thing. Different notation exists simply for convenience in different applications.

With this notation in hand, let's return to the concept of the derivative, discussed in the previous section. The derivative of a function is the instantaneous rate of change at a point. As noted in the previous section, we can take the difference between two points and calculate the discrete rate of change (i.e., the rate of change between any two points). But the derivative is the continuous rate of change (i.e., the instantaneous rate of change at any given point). How do we calculate it?

We noted in the previous section that taking the difference between discrete points that are closer and closer together (i.e., two values of x that are increasingly close in value) resulted in determining this instantaneous rate of change, and that doing so is effectively equivalent to taking the limit as this difference goes to zero. It turns out that this verbal description is actually all we need to produce a formal definition of the derivative that relies on our prior definition of the limit of a function. As we calculate the discrete rate of change over smaller and smaller units, we get better and better approximations of the derivative. In other words, the discrete rate of change over very small units is an approximation of the continuous rate of change. Taking the limit as the small units approach zero makes the approximation exact.

To see how this works, start with the equation for the slope of the secant:

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

Now let $x_2 = x_1 + h$, where h is any real number. Rearranging, we see that $h = x_2 - x_1$ is the difference between the two discrete points. We can then write the slope as

$$m = \frac{f(x_1 + h) - f(x_1)}{h}.$$ [7]

Note that we have rearranged the equation for the slope of a line to more clearly express the rate of change in $f(x)$ over the interval h.

All that we have to do now is make the difference between points smaller and smaller until it approaches zero. That is, to get the derivative we take the limit as $h \to 0$:

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

This equation gives the slope of the tangent line to the function at any point. Note that the derivative is also a function in its own right: it varies with x and returns the value $f'(x)$.

The only thing missing from this formal presentation is the existence of derivatives. Derivatives can only be calculated at points at which limits exist because derivatives are in a sense limits themselves, but it is also necessary for a function to be continuous at a point to have a derivative at that point. In other words (recalling the logic in the section on proofs in Chapter 1), a function differentiable at a point is also necessarily continuous at the point. However, continuity is not sufficient for differentiability. For example, continuous func-

[7] You often see Δx used instead of h. There is no difference; these variables have no intrinsic meaning.

tions with "kinks," such as the absolute value function $f(x) = |x|$ at $x = 0$,[8] or with vertical tangent lines at a point, such as $f(x) = x^{\frac{1}{3}}$ at $x = 0$, don't have derivatives at those points. However, these functions are differentiable at all points *other than* $x = 0$.

5.3.3 Some Examples

Let's go through some examples to make this concrete, starting with the linear function $f(x) = 3x$. We calculate its derivative by plugging into the definition:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{3(x+h)-3x}{h} \\ &= \lim_{h\to 0} \frac{3x+3h-3x}{h} = \lim_{h\to 0} \frac{3h}{h} = 3. \end{aligned}$$

So the derivative is 3. We didn't even need to use the limit, because a linear function has a constant slope, equal to the slope of its tangent line (which overlaps the function, as noted in the previous section). In other words, only one line is tangent to any point on a linear equation, and it is the linear equation itself.

Now let's do a slightly more complicated one, $f(x) = x^2$. Again we plug into the definition:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^2-x^2}{h} \\ &= \lim_{h\to 0} \frac{x^2+2xh+h^2-x^2}{h} = \lim_{h\to 0} \frac{2xh+h^2}{h} = \lim_{h\to 0} 2x+h = 2x. \end{aligned}$$

We see that our earlier guess was correct: the derivative of $f(x) = x^2$ is $2x$. This means that the slope of the tangent line increases with x, and there is a different tangent line at each point.

Next, consider an even more complex derivative $f(x) = x^3 + x - 5$:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^3+(x+h)-5-(x^3+x-5)}{h} \\ &= \lim_{h\to 0} \frac{x^3+3x^2h+3xh^2+h^3+x+h-5-x^3-x+5}{h} \\ &= \lim_{h\to 0} \frac{3x^2h+3xh^2+h^3+h}{h} \\ &= \lim_{h\to 0} 3x^2+3xh+h^2+1 = 3x^2+1. \end{aligned}$$

[8] Draw the function to convince yourself that it has a negative constant slope for negative x and a positive constant slope for positive x, so that the point at which they meet can't have a well-defined instantaneous rate of change.

There are several things to note here. First, the constant term, 5, played absolutely no part—it cancels in the definition. Since this should be true for all constants, we should expect their derivatives to be zero. This makes sense: constants don't change by definition, so their rate of change should be zero.[9] Second, we know the slope of the linear term, x, is 1, by our first example. It appears as 1 by itself in the answer, suggesting that we'll be able to treat added terms separately. Third, only the second leading term in the expansion of $(x+h)^3$ (the term in x^2) survives the limit. This will be true in general, as the second term has h only to the first power, allowing it to cancel with the denominator and leave no h to go to zero in the limit. This will make longer expansions easier, as we can ignore all the other terms.

Finally, a derivative that will prove particularly useful in the next chapter $f(x) = \frac{1}{x}$:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{\frac{1}{x+h}-\frac{1}{x}}{h} \\ &= \lim_{h\to 0} \frac{\frac{x-(x+h)}{x(x+h)}}{h} = \lim_{h\to 0} \frac{-1}{x(x+h)} = -\frac{1}{x^2}. \end{aligned}$$

5.3.4 Multivariate Functions and the Partial Derivative

In general, we are keeping to functions of one variable in this part of the book, saving multivariate calculus for Part V. However, we recognize that some students will not get to that part on a first read-through despite our exhortations in the preface to read the first chapter of Part V. Given the importance of partial derivatives to understanding statistical work and formal theory, we offer an exceedingly brief discussion of the partial derivative here and again entreat our readers to visit the first chapter of Part V for more.

Without providing any formal definitions, let's extend our functions to consider two variables.[10] Write these as $f(x,z)$. Examples would be $f(x,z) = 3x^2z+2z$ or $y = \sqrt{x^2+z^2}$. Say we want to know how y changes with x, holding z constant. This is a fundamental question in both statistics and formal theory. We can find this via what is known as the **partial derivative**. It is written $\frac{\partial}{\partial x}f(x,y)$, or sometimes simply ∂_x, and means "treat every variable other than x as a constant, and just take the derivative with respect to x." Consequently, partial derivatives are no more complex to take than derivatives of one-dimensional functions. For instance, if $f(x,z) = 3z^3 - 3z^2 + \sqrt{z} + x$, then $\frac{\partial}{\partial x}f(x,z) = 1$, since only the last term has an x in it and it is linear in x. In essence, relative

[9] This is why we had to add the constant C in the integral we considered in Section 1 of this chapter. The derivative eliminates all information about constants in a function, which removes information about translations.

[10] We show in Part V of the book that all of this extends to more than two variables.

to the partial derivative with respect to x, $f(x,z) = 3z^3 - 3z^2 + \sqrt{z} + x$ is no different from $f(x,z) = x$.

Partial derivatives appear commonly in early statistics courses, which is why we bring them up here. The context is usually of an empirical model such as $y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz$. Here we have both a direct dependence on x by itself, and an interaction term with z. If we want to know how x affects y, holding z constant, we need to take the partial derivative with respect to x. Since this treats z as a constant, this partial derivative is $\frac{\partial}{\partial x} f(x,z) = \beta_1 + \beta_3 z$. When interpreting the effect of x on y, this is the expression of interest, which varies with z.

For example, Mondak et al. (2010) hypothesize that the size of a person's political discussion network will influence the impact that her personality characteristics will have on her political attitudes. They show that the extent to which a person is extroverted or conscientious influences her political attitudes, but that the size of the impact of extroversion or conscientiousness varies with the number of people with whom she has conversations about politics. Similarly, Hopkins (2010) argues that the local and national context interact to produce Americans' political attitudes toward immigrants. He finds that Americans' interpretation of demographic change varies depending on the local increase in immigrant population. Both of these studies present results based on discrete changes in the value of the intervening variable (network size for Mondak et al., 2010 and local increase in immigrant population for Hopkins, 2010), and use a graph to display those results. Alternatively they could have calculated, and reported, the first derivative. In statistical work economists frequently calculate and report derivatives (e.g., Eeckhout, Persico, and Todd, 2010) and have even adopted the convention of referring to these derivatives as marginal effects. In their statistical work, political scientists often make reference to marginal effects, though we use the term loosely, and more often than not it is used to describe discrete change, rather than continuous change arising from the derivative. Familiarity with the derivative permits greater precision in presentation as well as greater flexibility in choice about what results are of most interest given one's research question and hypothesis.

5.4 SUMMARY

This is the take-home point: derivatives help us study rates of change in continuous functions. The derivative is the instantaneous rate of change in a function at a point. If you are comfortable with discrete change (i.e., the difference between two values), then you will be able to master the derivative: it is the change in a function at a point rather than over two points. The key is to become familiar with the notation. The derivative is useful because it permits us to make precise statements about changes in relationships, and social scientists are often interested in the change that one variable causes in another, including the case where we want to hold all other variables constant. In a multivariate function the partial derivative provides us with precisely that information.

You will encounter the opinion that bringing mathematics (and especially calculus) to bear on the study of politics is a misguided effort motivated by "physics envy." This is a canard. One employs the tools of calculus to study politics if one wishes to sharpen one's deductions and/or one's ability to draw inferences by using statistical inference. Those are the primary motives that have merit. If one feels that one has sufficient specificity in one's theory, then there is no reason to employ calculus to develop theory. Similarly, if one feels that one can draw inferences sufficiently well without appeal to inferential statistics, then there is no reason to appeal to calculus as it is used in statistics. However, to the extent that one is dissatisfied with the rigor of deduction in one's theory, one may want to appeal to mathematics generally and, in some cases, calculus in particular to strengthen one's ability to deduce the implications of sets of assumptions and conjectures. And if one wishes to employ statistical inference, mathematics is essential and calculus is often helpful for understanding the appropriate use of that tool.

Finally, we remind our readers that many students find calculus quite challenging at first. However, as long as one can recall that the underlying concepts are fairly straightforward, we believe that it is merely a matter of time and practice before calculus becomes a useful tool in one's belt. To that end, we offer a few online resources. Those students who have never studied calculus before will likely find Daniel Kleitman's Calculus for Beginners and Artists site helpful: `http://www-math.mit.edu/~djk/calculus_beginners/`. Dan Sloughter's online text is also generally considered quite useful: `http://math.furman.edu/~dcs/book/`. Students who have studied calculus but would like an online refresher might try Harvey Mudd College's Calculus Tutorial, which includes review and quizzes: `http://www.math.hmc.edu/calculus/tutorials/`. Finally, the following introduction to the fundamental theorem of calculus is useful and contains some helpful Java figures: `http://ugrad.math.ubc.ca/coursedoc/math101/notes/integration/ftc.html`.

5.5 EXERCISES

1. Visit Daniel Kleitman's "Derivative and Tangent Line Applet" page here `http://www-math.mit.edu/~djk/calculus_beginners/tools/tools04.html`. Select the $y = x^2$ function, and change the x min and max values to -5 and 5. Change the y min and max values to -5 and 135. Click the Plot Function box to generate a new graph. Now click the Show Derivative box, and drag the slide bar to change the values of x. What is the relationship between the tangent line shown and the function being drawn? Now click the Show Second Derivative box and move the slide bar again. What function is now being drawn, and what is its relationship to the tangent

lines? (We discuss the second derivative in Chapter 8. For now we just want you to look at it.) Repeat this for the function $y = x^3$.

2. Use the definition of the derivative to find the derivative of y with respect to x for the following:
   a) $y = 6$.
   b) $y = 3x^2$.
   c) $y = x^3 - 2x^2 - 1$.
   d) $y = x^4 + 5x$.
   e) $y = x^8$.
   f) $y = 4x^3 - x + 1$.
   g) $y = 2x^4 + x^2 - 1$.
   h) $y = 5x^5 + 4x^4 + 3x^3 + 2x^2 + x + 1$.
   i) $y = 7x^4 - 9x^3 + 5x + 117$.
   j) $y = 27x^3 + 5x^2 - x + 13$.
3. For each of the following, first sketch $f(x)$. Then draw a secant between the points on f corresponding to $x = 1$ and $x = 3$ and to $x = 1$ and $x = 2$. Using these as a guide, draw the tangent line at $x = 1$. Guess its slope. Now use the definition of the derivative to find $f'(x)$. Finally, sketch the derivative and compare it to the tangent at $x = 1$.
   a) $f(x) = 2x^2 + 7$.
   b) $f(x) = x^3 - x + 1$.
4. For each of the following, find the partial derivative with respect to x.
   a) $f(x, z) = 3zx + 2z$.
   b) $f(x, z) = x^2 + 2z^2$.
   c) $f(x, z) = 3z^2 - z + 1$.
5. For each of the following, find the partial derivative with respect to z.
   a) $f(x, z) = 3x + 3z + 3$.
   b) $f(x, z) = 9x^2 + 3z^2$.
   c) $f(x, z) = 5xz + 7xz^2 + 9x^z$.
   d) $f(x, y, z) = 3x + 4y + 5z + 6$.
   e) $f(x, y, z) = 11z + 3x^2y + 5x^2z + 7z^2y$.
   f) $f(x, y, z) = 4x^2y^2z^2 + 8xyz + 12xy + 14x$.
   g) $f(x, y, z) = 8xyz^2 + 10x^2y^2 + 12x^2y + 14x^2z^2$.

Chapter Six

The Rules of Differentiation

In the previous chapter we introduced the derivative as an operator that takes a function and returns another function made up of the instantaneous rate of change of the first function at each point. We also presented the definition of the derivative in terms of the limit of the discrete rate of change in a function across two points as the difference between the points went to zero. From this definition we could, with some algebra, compute derivatives of some polynomial functions. We'll need to calculate derivatives for more complex functions in political science, however. Rather than go back to the definition each time, it will help to have some rules for the differentiation of specific functions and types of functions we can call on. This chapter presents those rules.

A common method of presentation of this material provides a list of many seemingly disconnected rules. While that type of presentation provides clear rules and an easy memorization device, we feel that it can lead to the perception that calculus is somehow more complicated than it is, fraught with specificities that one must memorize. In reality, most of the commonly used rules stem from a handful of properties of the derivative operator, and most of these properties may be deduced from the definition. To help make these connections, we present the material accordingly.

In Section 1 we develop the rules typically called the sum rule, product rule, quotient rule, and chain rule largely from the definition of the derivative itself. In Section 2 we use these rules and the definition of the derivative to offer derivatives of many common and special functions used in political science. In Section 3 we relent and provide a summary accounting of the rules for differentiation for ready reference, along with a discussion as to when to use each of the rules. If you already know or don't care about calculus but need to calculate some derivatives, you can skip to that section. You can also skip to that section if you are finding the derivations of the rules too difficult to follow the first time through. We expect that many students who have not had calculus before will be in this boat, and might benefit from some experience manipulating derivatives before going back to see whence came the rules they used to perform the manipulations.

6.1 RULES FOR DIFFERENTIATION

The definition of the derivative provided in the previous chapter implies several properties of the derivative that are useful in computation. We alluded to some of them in the previous chapter's examples; now we present them formally.

6.1.1 The Derivative Is a Linear Operator

In Chapter 3 we defined a linear function as one that satisfies the properties of additivity and scaling. A linear operator satisfies pretty much the same definition. That is, an operator like $\frac{d}{dx}$ is linear if $\frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}$ and $\frac{d(cf)}{dx} = c\frac{df}{dx}$ for any two (differentiable!) functions f and g and any constant c. Is the derivative linear? Let's check the definition

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

First we try additivity

$$\begin{aligned}
\frac{d(f+g)}{dx} &= \lim_{h\to 0} \frac{(f(x+h)+g(x+h)) - (f(x)+g(x))}{h} \\
&= \lim_{h\to 0} \frac{(f(x+h)-f(x)) + (g(x+h)-g(x))}{h} \\
&= \lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} + \frac{g(x+h)-g(x)}{h} \right) \\
&= \frac{df}{dx} + \frac{dg}{dx}.
\end{aligned}$$

That checks out, thanks to the fact that the limit of a sum is the sum of the limits. Scaling is easier, since the limit of a product is the product of the limits and the limit of a constant is a constant

$$\begin{aligned}
\frac{d(cf)}{dx} &= \lim_{h\to 0} \frac{c(f(x+h)) - c(f(x))}{h} \\
&= \lim_{h\to 0} c \left(\frac{f(x+h)-f(x)}{h} \right) \\
&= c\left(\frac{df}{dx} \right).
\end{aligned}$$

That was a little bit of algebra, but it got us something important: *the derivative is a linear operator*. This supersedes what are commonly referred to as the sum and difference rules: $(f+g)' = f' + g'$ and $(f-g)' = f' - g'$. Since the derivative is linear, we instead have the more general $(af+bg)' = af' + bg'$ for any two constants a and b and any two (differentiable) functions f and g. And, rather than having to memorize sum and difference rules, we just need to know that the derivative is a linear operator and know the characteristics of linearity, which come up in lots of contexts beyond just calculus.

Let's try some examples before moving on. Let $h(x) = x + x^2$. Call $f(x) = x$ and $g(x) = x^2$. Linearity then yields $h'(x) = f'(x) + g'(x)$. We can look up these derivatives in the previous chapter to get $\frac{dh(x)}{dx} = 1 + 2x$. We can add a constant a to this too without changing anything; if $h(x) = (x+a) + x^2$, then still $h'(x) = 1 + 2x$, because the derivative of a constant is zero, as we saw in the previous chapter. So don't let yourself get distracted by constants. Either they vanish in the derivative if they are added, or they get pulled out to multiply the relevant terms in the derivative. For example, if $f(x) = ax$, where a is a constant, then $f'(x) = a$ since the derivative of x is one. We can expand these ideas to any number of terms. For example, if $h(x) = x^3 + 6x^2 - 3x + 1$, then $h'(x) = 3x^2 + 12x - 3$. All we've done here is take the derivative of each term separately. Let's work out this last one to see the application of the rule a bit more carefully

$$\begin{aligned}
\frac{dh(x)}{dx} &= \frac{d(x^3 + 6x^2 - 3x + 1)}{dx} \\
&= \frac{d(x^3)}{dx} + \frac{d(6x^2)}{dx} + \frac{d(-3x)}{dx} + \frac{d(1)}{dx} \\
&= \frac{d(x^3)}{dx} + 6\frac{d(x^2)}{dx} - 3\frac{d(x)}{dx} + \frac{d(1)}{dx} \\
&= 3x^2 + 6(2x) - 3(1) + 0 \\
&= 3x^2 + 12x - 3.
\end{aligned}$$

6.1.2 Chain Rule

Believe it or not, the fact that the derivative is a linear operator takes us much of the way toward providing rules for differentiation. To go further, though, we need to know how to deal with composite functions. Recall from Chapter 3 that a composite function looks like $g(f(x))$. To see why we're bringing this up now, note that most complex functions can be written as composite functions. Take, for example, $h(x) = e^{2x^2}$. If we let $f(x) = 2x^2$ and $g(x) = e^x$, then $h(x) = g(f(x))$. Thus we can break complicated functions down into the composition of simpler functions. If we can devise a rule for differentiating composite functions, a rule known as the **chain rule**, then we can simplify the differentiation of complex functions immensely. Given this incentive, we begin this task.[1]

We let f and g be differentiable, and use the definition of the derivative to get $\frac{dg(f(x))}{dx}$. We start with

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

[1]We should note that the following "proof" for the chain rule is "hand-wavy" even by the standards of this book. But the idea is more or less correct.

Then we multiply and divide by $f(x+h)-f(x)$:[2]

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

Since f is assumed differentiable, it is also continuous (see the previous chapter), so as $h \to 0$, $f(x+h)-f(x) \to 0$. Let's call $z = f(x+h)-f(x)$ and $u = f(x)$; rewriting the derivative with these substitutions and using the fact that the limit of a product is the product of a limit yields

$$\frac{dg(f(x))}{dx} = \lim_{z\to 0} \frac{g(u+z)-g(u)}{z} \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}.$$

We're almost done. The second term on the right-hand side (RHS) is $f'(x)$. The first term is $g'(u) = g'(f(x))$, where we've plugged back in our substitution for x.[3] Finally, this gives us the chain rule:

$$\frac{dg(f(x))}{dx} = \frac{dg(u)}{du}\frac{du}{dx}, \text{where } u = f(x).$$

In words, take the derivative of the outer function (g of x) evaluated at the inner function (f of x), then take the derivative of the inner function (f of x), and multiply the two terms. We can also write it as $(g(f(x)))' = g'(f(x))f'(x)$. The key concept here is that one can separate the derivative of a composite function into two derivatives: one of the outer function, evaluated at the inner function, and one of the inner function, evaluated at x.

This may have seemed complicated, but the chain rule is perhaps the most useful rule in differential calculus, and one of only a handful you should actually try to memorize.[4] Not only is it useful in its own regard, but it gives us many other rules as well.[5] But before getting into that, we start with some examples of its use.

Let's start with $h(x) = x^9$. We show below this is a surprisingly easy derivative to take, but right now you know only the definition of the derivative, and you might not feel like expanding $(x+h)^9$. So let's instead call $f(x) = x^3$, $g(x) = x^3$, and $h(x) = g(f(x))$; we check that $h(x) = (x^3)^3 = x^9$, recalling our rules on exponentiating exponents from Chapter 3. We've now seen several times that the derivative of x^3 is $3x^2$. So, plugging this into the chain rule, and letting $u = x^3$, we have $\frac{dg(f(x))}{dx} = \frac{dg(u)}{du}\frac{df(x)}{dx} = (3u^2)(3x^2) = 3(x^3)^2(3x^2) = 9x^8$. Not only is this much easier than expanding something to the ninth power, it

[2] For those keeping score at home, here's the "hand-wavy" part. The possibility of dividing by zero means a formal proof needs more care at this step than we offer.

[3] You'll see this technique a great deal in the next chapter, as the method of substitution in integration is closely related to the chain rule in differentiation.

[4] You will encounter many examples in political science publications, but for a specific one that uses the chain rule to find the long-run effects of a change in one variable in the context of a discrete difference equation, see Brandt and Williams (2001).

[5] It also proves immensely important in multivariate analysis, but we leave this until Part V of the book.

also lets us produce the derivative for a polynomial of any positive power we want, just by working up from ones we know already.

Of course, the two functions composed need not be equal. Let's return to our first example in this section, $h(x) = e^{2x^2}$. Again, we let $f(x) = 2x^2$ and $g(x) = e^x$, so $h(x) = g(f(x))$. We show below that $\frac{de^x}{dx} = e^x$, and we can use the fact that the derivative is a linear operator to find that the derivative of $2x^2$ is $2\frac{dx^2}{dx} = 2(2x) = 4x$. Given this, we can apply the chain rule to get $h'(x) = g'(u)f'(x) = e^u(4x) = 4xe^{2x^2}$, where in the last step we plugged back in for u.

One more example: let $h(x) = (2x-a)^2$. We'll set $g(x) = x^2$ and $f(x) = 2x - a$, so that $h(x) = g(f(x))$. Then $h'(x) = (g(f(x)))' = g'(f(x))f'(x) = 2(2x-a)(2) = 8x - 4a$, where in the second-to-last expression we evaluated $g'(x)$ at $f(x)$.

The major step in using the chain rule is to figure out how to assign $g(x)$ and $f(x)$. We'll talk more about how to do this in Section 3 below, but in general, you want each function to be one that you can differentiate easily. The tricky part in using the chain rule is to make sure that you substitute the inner function, $f(x)$, for the x in $g'(x)$. One way to make sure you do this right is to set $u = f(x)$ and then differentiate $g(u)$ with respect to u. This is what we'll usually do, and what we did in our first two examples. All you have to remember in this case is to plug back in for u at the end.

However, you won't always see it done that way, so it's good to get used to using the chain rule without this substitution, instead directly using $(g(f(x)))' = g'(f(x))f'(x)$ as we did in the last example. That is, you first calculate $g'(x)$ as if the function were not nested at all, and then you evaluate it at $f(x)$, which means, practically, replacing each x in $g'(x)$ with $f(x)$.[6] This is different from multiplying $g'(x)$ and $f(x)$.

There are many more examples like these, and we provide some at the end of the chapter. For now, however, we turn to using the chain rule to derive other rules. We start by considering the inverse function of $f(x)$, $f^{-1}(x)$. Recall that for an inverse function, $f(f^{-1}(x)) = x$. We can take the derivative of both sides, which is allowed since neither derivative is undefined or infinite. The RHS derivative is 1. To find the LHS, we use the chain rule to get

$$\frac{df(f^{-1}(x))}{dx} = \frac{df(u)}{du}\frac{df^{-1}(x)}{dx}, \text{where } u = f^{-1}(x).$$

The LHS of this equation is equal to one, as it is $\frac{dx}{dx}$. Assuming the derivative of the function exists and is not zero, the first term on the RHS is $f'(f^{-1}(x))$,

[6] If this is weird to you, think of evaluating a derivative at a point a. If we just had $g'(x)$, we could evaluate it at a to get $g'(a)$. (Remember that $g'(x)$ is a function, just as $g(x)$ is.) So this means replacing each x in $g'(x)$ with a. But instead we have a nested function, and we have to evaluate $g'(x)$ at the value that $f(x)$ maps a into: $f(a)$. So instead of replacing each x in $g'(x)$ with a, we replace each with $f(a)$. Both a and $f(a)$ are just values, so there's nothing much different here either way. This is true for all x in the domain, though, so we have to replace each x with a corresponding $f(x)$.

and we can divide by it. Putting these two facts together allows us to rearrange the equation to get the inverse function rule

$$\frac{df^{-1}(x)}{dx} = \frac{1}{f'(f^{-1}(x))}.$$

So we can get the derivative of the inverse of the function from the derivative of the function itself. Using this rule is no different from using the chain rule, so we leave examples until the exercises at the end of the chapter.

6.1.3 Products and Quotients

We can now deal with sums and differences of functions, as well as composite functions and functions multiplied by constants. The only remaining rule of immediate use is that for products and quotients of functions. Luckily, only the first need be memorized, and this too can be derived from the definition of the derivative if you happen to forget it.

We should note before this derivation that in contrast to sums and differences of functions, one does not simply differentiate the terms of a product. There are two ways to go about the task. The easier route is not always available, but if one can, do the multiplication first and then take the derivative of the resulting product. For example, if $y = f(x) \times g(x) = (2x+3) \times (x^2 - 15)$, then we can take the product, yielding $2x^3 + 3x^2 - 30x - 45$. Differentiating yields $6x^2 + 6x - 30$.

Unfortunately, one cannot always multiply out before differentiating. In such cases the product rule is needed. If we have two functions, $f(x)$ and $g(x)$, then the derivative of their product is

$$\frac{d(f(x)g(x))}{dx} = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx}.$$

In words, the **product rule** states that the derivative of the product is a sum of product terms. In each such term, one and only one of the constituent functions is differentiated. Or, if you like, the derivative of the product of two terms is the derivative of the first term times the second term, plus the first term times the derivative of the second. Clear as mud, eh? We try to clarify by first deriving it, and then providing an example.

Let f and g be differentiable functions. The definition of the derivative gives

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

We next add and subtract the term $f(x)g(x+h)$ within the RHS to yield

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

Grouping terms and making use of the properties of the limit gives

$$\begin{aligned}(fg)' &= \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x)g(x+h)}{h} \\ &\quad + \lim_{h\to 0}\frac{f(x)g(x+h) - f(x)g(x)}{h}.\end{aligned}$$

Pulling out the common terms in each case and making use again of the properties of the limit gives

$$\begin{aligned}(fg)' &= \left(\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\right)\left(\lim_{h\to 0} g(x+h)\right) \\ &\quad + \left(\lim_{h\to 0} f(x)\right)\left(\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\right).\end{aligned}$$

Finally, taking the limits gives us the rule:

$$(fg)' = f'g + fg'.$$

Now for the example. Consider the functions above: $y = f(x)g(x) = (2x + 3)(x^2 - 15)$. Use the product rule

$$\begin{aligned}\frac{dy}{dx} &= \frac{d(2x+3)}{dx}(x^2-15) + (2x+3)\frac{d(x^2-15)}{dx} \\ &= (2)(x^2-15) + (2x+3)(2x) \\ &= (2x^2-30) + (4x^2+6x) \\ &= 6x^2 + 6x - 30.\end{aligned}$$

We get the same answer regardless of which way we go about it, which suggests that we have done it properly.

Instead of considering $y = f(x)g(x)$, we might also need to calculate $y = \frac{f(x)}{g(x)}$. There is a **quotient rule** for this that people often memorize (note that we must have $g \neq 0$)

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{\frac{df(x)}{dx}g(x) - f(x)\frac{dg(x)}{dx}}{g(x)^2}.$$

You are free to memorize this as well, and it's not so hard. The numerator is just like the product rule but with a negative sign instead of a positive one between the terms, and the denominator is just the square of the second term. But the quotient rule is just an application of the product rule combined with the chain rule. Let us explain. The only difference between the quotient and the product under consideration is that we are dividing by g instead of multiplying by it. So let's go back to multiplying by letting $h(x) = \frac{1}{g(x)}$. Now $y = f(x)h(x)$ as before. We already know that $y' = f'h + fh'$ from the product rule, so all we need is h'. We can think of h as a composite function: $h(x) = k(g(x))$, where $k(x) = \frac{1}{x}$. From the examples in the last chapter, we know that $k'(u) = -\frac{1}{u^2}$.

By the chain rule, then, $h'(x) = -\frac{1}{(g(x))^2}g'(x)$. We can plug this back into the result from the product rule to get $y' = \frac{f'}{g} - \frac{fg'}{g^2}$. Multiplying numerator and denominator of the first term by g and combining terms yields the quotient rule $y' = \frac{f'g - fg'}{g^2}$.

At this point you might be saying, "Of *course* I'm going to memorize the quotient rule! Who wants to re-derive all that every time?" That is a good point, but it misses the intent of this derivation. The fact that one can get the quotient rule from the product rule and the chain rule is indicative of the fact that one can compute derivatives using just the product rule and the chain rule; one does not need the quotient rule at all. Whether or not you want to bother memorizing it, then, depends on your facility with algebraic manipulation and fractions.

Let's try an example. Consider a situation where we are interested in the quotient $y = \frac{f(x)}{g(x)}$. The derivative of a ratio would be an example. Let $f(x) = (3x - 7)$ and $g(x) = (x^3 + 6)$. We need to find $\frac{d}{dx}\left(\frac{3x-7}{x^3+6}\right)$.

First we use the quotient rule to differentiate y

$$\begin{aligned}\frac{dy}{dx} &= \frac{\frac{d(3x-7)}{dx}(x^3+6) - (3x-7)\frac{d(x^3+6)}{dx}}{(x^3+6)^2} \\ &= \frac{(x^3+6)(3) - (3x-7)(3x^2)}{36 + 12x^3 + x^6} \\ &= \frac{(3x^3+18) - (9x^3 - 21x^2)}{36 + 12x^3 + x^6} \\ &= \frac{-6x^3 + 21x^2 + 18}{36 + 12x^3 + x^6}.\end{aligned}$$

Now we use the product and chain rules. Let $h(x) = \frac{1}{x^3+6}$. Then $h' = -\frac{1}{(x^3+6)^2}(3x^2)$ from the chain rule and $y' = f'h + fh'$ from the product rule. Since $f' = 3$, we can plug in to get

$$\begin{aligned}\frac{dy}{dx} &= \frac{3}{x^3+6} - \frac{(3x-7)(3x^2)}{(x^3+6)^2} \\ &= \frac{(x^3+6)(3) - (3x-7)(3x^2)}{36 + 12x^3 + x^6} \\ &= \frac{(3x^3+18) - (9x^3 - 21x^2)}{36 + 12x^3 + x^6} \\ &= \frac{-6x^3 + 21x^2 + 18}{36 + 12x^3 + x^6}.\end{aligned}$$

As you can see, the second method requires a bit more algebra and keeping track of a few more terms, but it doesn't require memorizing the quotient rule. They both give the same answer.

6.2 DERIVATIVES OF FUNCTIONS

To sum up the previous section, we have three useful rules of differentiation: the chain rule, the product rule, and the fact that the derivative is a linear operator. Other rules, such as the quotient, sum, difference, and inverse function rules, can all be derived from these three. Further, all three can be derived (more or less) from the definition of the derivative. We weren't pulling a fast one when we said there's not a lot to this: the derivative is the limit of discrete change in a function as the distance over which the change is measured goes to zero, and the rules of differentiation follow from this definition. If you get the concept of a derivative and can do a little algebra, you can get the rules.

This isn't quite fair, though, as we haven't told you how to differentiate specific functions yet, beyond those polynomials for which we could easily use the definition of the derivative. To remedy this, in this section we present derivatives of some common and special functions. Again, though, there will not be many of these to memorize.

6.2.1 Polynomials and Powers

By now you have most likely noticed that nearly all the functions you've dealt with can be written as a sum, difference, product, or quotient of terms like ax^n, where a and n are constants. For example, if $n = 0$, then we have $ax^0 = a$, which is just a constant. If $n = 1$, then we have $ax^1 = ax$, which is a linear function. If $n = 2$ we have ax^2, which is a quadratic function. If we add all three of these together we get a polynomial that describes a parabola. If $n = -1$, we get $ax^{-1} = \frac{a}{x}$. Finally, if $n = \frac{1}{2}$, we get $ax^{\frac{1}{2}} = a\sqrt{x}$, and if $n = \frac{1}{3}$, we get $ax^{\frac{1}{3}} = a\sqrt[3]{x}$.

We know we can treat each term in a sum or difference separately, thanks to the fact that the derivative is a linear operator. For this same reason we can pull the constant a out of the derivative. We also know how to deal with product, rational, and composite functions thanks to the product, quotient, and chain rules, respectively. Thus, if we can just differentiate anything of the form x^n, where n is a rational number, we can address a large proportion of derivatives in political science.

It turns out that it is easy to take this derivative. You might even have guessed the pattern from earlier examples such as $(x^3)' = 3x^2$ and $(x^9)' = 9x^8$. It's just

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

In words, move the value of the exponent to the front of the variable and subtract one from the exponent. This is true for any rational n and real x, other than times when x^n would be poorly defined.[7] This is generally referred to as the **power rule**.

[7]For example, if x were negative and n were one-half, or if both n and x were zero.

How did we get this? For n a positive integer there's a proof that is worth a moment of your time. Note that the first two terms in $(x+h)^n$ are x^n and nhx^{n-1}. All other terms have higher powers in h. As we noted in the previous chapter, all terms with powers in h in the numerator of the definition of the derivative will vanish in the limit, so all we need to care about are these first two terms. So that leaves

$$\frac{dx^n}{dx} = \lim_{h\to 0}\frac{(x+h)^n - x^n}{h} = \lim_{h\to 0}\frac{x^n + nhx^{n-1} - x^n}{h} = nx^{n-1},$$

where in the last step we canceled the x^n in the numerator and the h from top and bottom, leaving no h to worry about in the limit.

For more general n, the easiest way to see the rule is to use a slightly different definition of the derivative. Consider the derivative at a point c. As x approaches c, the difference $c - x$ goes to zero, just as h does. So let $h = c - x$. Then the derivative definition becomes

$$f'(c) = \lim_{x\to c}\frac{f(x) - f(c)}{x - c}.$$

If $f(x) = x^n$, then the numerator is $x^n - c^n$. We can factor out $(x - c)$ to get

$$(x-c)(x^{n-1} + cx^{n-2} + \ldots + c^{n-2}x + c^{n-1}).$$

After canceling the $(x - c)$ from numerator and denominator, taking the limit, and adding all the n identical terms, we get

$$(x^n)'|_c = \lim_{x\to c} x^{n-1} + cx^{n-2} + \ldots + c^{n-2}x + c^{n-1} = nc^{n-1}.$$

The $|_c$ here means evaluate the derivative at $x = c$. Since this is true for any point c, we have $\frac{dx^n}{dx} = nx^{n-1}$.

The formula $\frac{dx^n}{dx} = nx^{n-1}$ is very flexible, and we advise you to memorize this relatively simple relation rather than try to reproduce it. Let's try some examples.

First, let's consider the simplest case, a constant. Let $y = f(x) = 5$. What is the derivative of this function? Our formula says that for $y = 5 = 5x^0$, $y' = 0(5x^{-1}) = 0$. Does this make sense? Recall that the derivative is the instantaneous rate of change of the function. Does y change? No, not in this case. Thus, the rate of change should be zero, as should the derivative of any constant. Some people call this the **constant rule**, but it is a simple application of our formula for x raised to a power.[8]

Now try a linear function, $y = f(x) = ax$. Since $x = x^1$ and $1x^{1-1} = 1x^0 = 1$, our formula implies that $f'(x) = \frac{d(ax)}{dx} = a$. The derivative of a linear function is a constant, which is just the slope of the line.

[8]As usual, we're ignoring some technical details, in this case the value at $x = 0$. But the constant rule we offer is true regardless, and this provides the intuition.

Higher exponents are more straightforward. Change the function to $y = g(x) = x^2$. Pulling down the exponent and replacing it with an exponent one less, as our formula requires, yields $2x^1 = 2x$. And we've already seen that if $y = x^3$, then $y' = 3x^2$, and if $y = x^9$, then $y' = 9x^8$.

We can combine this formula with the rules from the previous section to calculate the derivatives of rather difficult-looking functions. For example, what is the derivative of $y = 10x^3 + x^2 - 3x$? The answer is $30x^2 + 2x - 3$. Not too bad, eh?[9] And it only required linearity and the power rule. Or if $f(x) = (2x + 5)$ and $g = (x^3 + 2x + 1)$, you can multiply them out to get $y = f(x)g(x)$ and differentiate this, but you could also use the product rule: $y' = f'g + fg' = (2)(x^3 + 2x + 1) + (2x + 5)(3x^2 + 2)$. Finally, you can differentiate functions like $y = (x - c)^3$ by using the chain rule. Assign $g(x) = x^3$ and $f(x) = x - c$. Then $y' = g'(f(x))f'(x) = 3(x - c)^2$, since $f' = 1$. (This is in general true; because constants differentiate to zero, owing to the chain rule one can treat $(x \pm c)$ as x when differentiating. Note also that c can be left as a constant here; you need not substitute a real number to differentiate.) The point is, all polynomials are now differentiable by you via the application of only a handful of rules, all derived from the same definition.

Recall also that one can use exponents to represent fractions and radicals (or roots). For example, $x^{-n} = \frac{1}{x^n}$ and $x^{\frac{1}{n}} = \sqrt[n]{x}$. How does one apply the power rule to these functions? There are no tricks involved—just apply the rule. Let's try x^{-2}. Applying the rule yields $-2x^{-3}$. The case of a fractional exponent is the same. Consider the derivative of the function $y = x^{\frac{1}{3}}$. To apply the rule we move $\frac{1}{3}$ to the front of x and subtract one from $\frac{1}{3}$, yielding $\frac{1}{3}x^{-\frac{2}{3}}$. Again, we get a great deal of flexibility from a very simple rule.

6.2.2 Exponentials

As in Chapter 3, we've covered x^a, but now need to cover a^x. We start with the most commonly used exponential, e^x. The derivative of e^x is even simpler than that of x^n. In fact, it's just e^x, the same function with which we started. That is,

$$\frac{de^x}{dx} = e^x.$$

How can this be? To derive it we will need the definition of the derivative, as always, but we will also need to express the exponential function in a different way. It turns out that there are many ways to define the exponential function, all equivalent. You may see two common ones:

$$\begin{aligned} e^x &= \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n, \text{or} \\ &= \sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots. \end{aligned}$$

[9]Don't worry if you don't yet see this quickly. You will, though it'll come faster if you take each opportunity like this to practice and check your work against ours.

We use the second definition for our purposes.

Now to the proof:

$$\frac{de^x}{dx} = \lim_{h\to 0}\frac{e^{x+h}-e^x}{h} = \lim_{h\to 0}\frac{e^xe^h-e^x}{h} = \lim_{h\to 0}\frac{e^x(e^h-1)}{h} = e^x\lim_{h\to 0}\frac{e^h-1}{h}.$$

We've got the e^x now; we just need to show that the rest of the term on the RHS equals 1. This is where we use our definition of e^x:

$$\lim_{h\to 0}\frac{e^h-1}{h} = \lim_{h\to 0}\frac{-1+1+h+\frac{h^2}{2}+\frac{h^3}{6}+\dots}{h} = 1,$$

where we've used the fact that all terms of order h^2 and higher in the numerator vanish in the limit. This completes the proof.

To get the more general case of a^x for any a, we can use the properties of the exponential and logarithm to write it as $a^x = e^{x\ln(a)}$.[10] Next we use the chain rule. We let $g = e^x$ and $f = x\ln(a)$, so $\frac{da^x}{dx} = (e^{x\ln(a)})(\ln(a))$. Or, after rewriting the first term on the RHS one more time

$$\frac{da^x}{dx} = a^x(\ln(a)).$$

This is the **exponential rule**.

Believe it or not, between the power rule and this exponential rule, plus the three rules in the previous section, we've covered nearly all functions used regularly in political science. With these, even really complex functions are straightforward to differentiate (though they do require you to be careful with the algebra to avoid mistakes!).

For example, let $y = e^{x^2}$. We have one function (e^x) composed with another (x^2), so the chain rule is in order. Let $g(x) = e^x$ and $f(x) = x^2$. Then $y = g(f(x))$ and $y' = g'(f(x))f'(x) = (e^{x^2})(2x)$. This shows the nice thing about exponentials: they always return the original function, multiplied by the derivative of the term in the exponent. So the derivative of the really complicated function $y = e^{x^4-3x^2+1}$ is just $y' = (e^{x^4-3x^2+1})(4x^3-6x)$.

6.2.3 Logarithms

The only other function you'll encounter with any regularity is the logarithm. Recall from Chapter 3 that it is the inverse of the exponential function. So, for the natural log, $e^{\ln(x)} = x$. We can use the inverse function rule from the previous section directly, or we can just differentiate both sides and use the chain rule again. We'll do the latter. Let $g = e^x$ and $f = \ln(x)$. Then by the chain rule the LHS is $\frac{d\ln(x)}{dx}e^{\ln(x)} = \frac{d\ln(x)}{dx}x$, where we've again used the definition of an inverse function. The derivative of the RHS is just one. Thus, we get $\frac{d\ln(x)}{dx}x = 1$, or

$$\frac{d\ln(x)}{dx} = \frac{1}{x}.$$

[10] This is true because $e^{x\ln(a)} = e^{\ln(a^x)} = a^x$.

Usually people just memorize this, but as you can see, the derivation is pretty quick.

What about the more general logarithm $\log_a x$? Well, this is the inverse function of a^x, so $a^{\log_a(x)} = x$, and we can do the same proof over again. We leave it to you do this as an exercise, but we'll give you the answer

$$\frac{d\log_a(x)}{dx} = \frac{1}{x\ln(a)}.$$

While derivatives of the log don't have quite the nice properties of the exponential, they do have two things going for them. One, rather than just return the entire complicated exponential, they return one over the object of the log times the derivative of the object of the log, via the chain rule. So, $\ln(x)$ becomes $\frac{1}{x}$ under differentiation, since the derivative of the object of the log, x, is one. For a more complex example, let $y = \ln(x^5-2x^2+12)$. Then $y' = \frac{1}{x^5-2x^2+12}(5x^4-4x)$ via the chain rule. We can also write this as $y' = \frac{5x^4-4x}{x^5-2x^2+12}$, which illuminates the second thing derivatives of the log have going for them: they give you the *relative rate of change*, and are often used to this effect in the literature.

To see how this works, let $y = \ln(f(x))$ for some function $f(x)$. Then $y' = \frac{1}{f(x)}f'(x) = \frac{f'(x)}{f(x)}$. In words, the derivative of the log of a function is the derivative of the function divided by the function itself, giving the rate of change of the function relative to the value of the function.

6.2.4 Other Functions

This covers most of the functions that we see in political science. We mentioned a few other types of functions in Chapter 3, however, so we discuss their derivatives briefly here, though we'll not include proofs.

The derivatives of trigonometric functions turn them into each other. So $(\sin(x))' = \cos(x)$, $(\cos(x))' = -\sin(x)$, and $(\tan(x))' = 1+\tan^2(x)$.

Piecewise functions lack derivatives at points at which there are discontinuous jumps or kinks. The derivative of the rest of the function can itself be treated piecewise, omitting the troublesome points. For example, consider the function we presented in Chapter 3

$$f(x) = \begin{cases} -(x-2)^2 & : x \le 2, \\ \ln(x-2) & : x > 2. \end{cases}$$

We can write its derivative as

$$f'(x) = \begin{cases} -2(x-2) & : x < 2, \\ \frac{1}{x-2} & : x > 2. \end{cases}$$

Note that the point $x = 2$ has no derivative, though it is defined for $f(x)$.

Calc Rules

6.3 WHAT THE RULES ARE, AND WHEN TO USE THEM

As promised, in this section we list the **rules for differentiation** discussed in this chapter. In all these cases, f and g are assumed to be differentiable functions, and a is a constant. Table 6.1 lists the rules, and we hope that if you read the previous two sections as well, you have a pretty good idea of their origins. We then conclude with a brief discussion about how to use them.

Table 6.1: List of Rules of Differentiation

Sum rule	$(f(x)+g(x))' = f'(x)+g'(x)$
Difference rule	$(f(x)-g(x))' = f'(x)-g'(x)$
Multiply by constant rule	$f'(ax) = af'(x)$
Product rule	$(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$
Quotient rule	$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{(g(x))^2}$
Chain rule	$(g(f(x))' = g'(f(x))f'(x)$
Inverse function rule	$(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}$
Constant rule	$(a)' = 0$
Power rule	$(x^n)' = nx^{n-1}$
Exponential rule 1	$(e^x)' = e^x$
Exponential rule 2	$(a^x)' = a^x(\ln(a))$
Logarithm rule 1	$(\ln(x))' = \frac{1}{x}$
Logarithm rule 2	$(\log_a(x))' = \frac{1}{x(\ln(a))}$
Trigonometric rules	$(\sin(x))' = \cos(x)$
	$(\cos(x))' = -\sin(x)$
	$(\tan(x))' = 1+\tan^2(x)$
Piecewise rules	Treat each piece separately

There is one basic tactic of differentiation: use the rules on combining functions as sums, products, or composites to reduce the problem to the smallest pieces you know how to manage, differentiate these pieces, and then build up the answer again according to the rules. We work through one complex example here to show you what we mean; others can be found in the exercises.

Let $y = \frac{(5\ln(x+3))e^{3x^3-10x}}{5x^2+2}$. This is messy and there's not much we can do about that, but its derivative is solvable if we're methodical and use what we've learned. First we get rid of the quotient. So call $f(x) = (5\ln(x+3))e^{3x^3-10x}$ and $g(x) = 5x^2+2$. We've seen enough polynomials by now to know that $g'(x) = 10x$, so that's broken down enough.[11]

The integral f is still pretty complicated, though, so let's set $u(x) = 5\ln(x+3)$ and $v(x) = e^{3x^3-10x}$. We could break these down further if we liked, but we've

[11]If this is not true for you, then break it up further $g(x) = a(x)b(x)+c(x)$, where $a(x) = 5, b(x) = x^2$, and $c(x) = 2$.

seen examples of each type before, so we can take their derivatives: $u'(x) = \frac{5}{x+3}$ and $v'(x) = (9x^2-10)e^{3x^3-10x}$.[12]

Now we have all the components, so we can put it back together. We've used the quotient and product rules (and also the chain rule, if you've followed the footnotes), so

$$y' = \frac{f'g-fg'}{g^2} = \frac{(uv)'g-fg'}{g^2} = \frac{(u'v+uv')g-fg'}{g^2}.$$

All that's left is to plug the functions f, g, u, v into this to get our answer

$$\begin{aligned} y' &= (5x^2+2)^{-2}\Bigg[\left(\frac{5e^{3x^3-10x}}{x+3}+5\ln(x+3)(9x^2-10)e^{3x^3-10x}\right) \\ &\quad \times(5x^2+2)-(5\ln(x+3))e^{3x^3-10x}(10x)\Bigg]. \end{aligned}$$

This is long and a big mess, true (it's not really worth the effort to simplify), but the important point again is that, via a handful of rules, one can differentiate very complex functions by first breaking them down into manageable parts and then building the derivative back up again. At first these parts will be small, likely smaller than in our example, but with practice they'll get bigger and differentiation will get faster and less worrisome. And practice is key: there's nothing special about any of this beyond becoming accustomed to the techniques, and we hope we've convinced you that the techniques can be very useful.

6.4 EXERCISES

1. Find the derivative of y with respect to x for the following, using the rules in this chapter:
   - $y = 6$.
   - $y = 3x^2$.
   - $y = x^3-2x^2-1$.
   - $y = x^4+5x$.
   - $y = x^8$.
   - $y = x^{-3}$.
   - $y = ax^3+6$.
   - $y = 12x^{\frac{1}{2}}+c$.

[12]Again, if this is not clear, break them up further. In particular, let $d(x) = e^x$ and $k(x) = 3x^3-10x$ in $v(x)$ so that you can use the chain rule.

6.3 WHAT THE RULES ARE, AND WHEN TO USE THEM

As promised, in this section we list the **rules for differentiation** discussed in this chapter. In all these cases, f and g are assumed to be differentiable functions, and a is a constant. Table 6.1 lists the rules, and we hope that if you read the previous two sections as well, you have a pretty good idea of their origins. We then conclude with a brief discussion about how to use them.

Table 6.1: List of Rules of Differentiation

Sum rule	$(f(x)+g(x))' = f'(x)+g'(x)$
Difference rule	$(f(x)-g(x))' = f'(x)-g'(x)$
Multiply by constant rule	$f'(ax) = af'(x)$
Product rule	$(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$
Quotient rule	$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{(g(x))^2}$
Chain rule	$(g(f(x))' = g'(f(x))f'(x)$
Inverse function rule	$(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}$
Constant rule	$(a)' = 0$
Power rule	$(x^n)' = nx^{n-1}$
Exponential rule 1	$(e^x)' = e^x$
Exponential rule 2	$(a^x)' = a^x(\ln(a))$
Logarithm rule 1	$(\ln(x))' = \frac{1}{x}$
Logarithm rule 2	$(\log_a(x))' = \frac{1}{x(\ln(a))}$
Trigonometric rules	$(\sin(x))' = \cos(x)$
	$(\cos(x))' = -\sin(x)$
	$(\tan(x))' = 1+\tan^2(x)$
Piecewise rules	Treat each piece separately

There is one basic tactic of differentiation: use the rules on combining functions as sums, products, or composites to reduce the problem to the smallest pieces you know how to manage, differentiate these pieces, and then build up the answer again according to the rules. We work through one complex example here to show you what we mean; others can be found in the exercises.

Let $y = \frac{(5\ln(x+3))e^{3x^3-10x}}{5x^2+2}$. This is messy and there's not much we can do about that, but its derivative is solvable if we're methodical and use what we've learned. First we get rid of the quotient. So call $f(x) = (5\ln(x+3))e^{3x^3-10x}$ and $g(x) = 5x^2+2$. We've seen enough polynomials by now to know that $g'(x) = 10x$, so that's broken down enough.[11]

The integral f is still pretty complicated, though, so let's set $u(x) = 5\ln(x+3)$ and $v(x) = e^{3x^3-10x}$. We could break these down further if we liked, but we've

[11] If this is not true for you, then break it up further $g(x) = a(x)b(x)+c(x)$, where $a(x) = 5, b(x) = x^2$, and $c(x) = 2$.

seen examples of each type before, so we can take their derivatives: $u'(x) = \frac{5}{x+3}$ and $v'(x) = (9x^2-10)e^{3x^3-10x}$.[12]

Now we have all the components, so we can put it back together. We've used the quotient and product rules (and also the chain rule, if you've followed the footnotes), so

$$y' = \frac{f'g-fg'}{g^2} = \frac{(uv)'g-fg'}{g^2} = \frac{(u'v+uv')g-fg'}{g^2}.$$

All that's left is to plug the functions f, g, u, v into this to get our answer

$$\begin{aligned} y' &= (5x^2+2)^{-2}\Bigg[\left(\frac{5e^{3x^3-10x}}{x+3} + 5\ln(x+3)(9x^2-10)e^{3x^3-10x}\right) \\ &\quad \times(5x^2+2) - (5\ln(x+3))e^{3x^3-10x}(10x)\Bigg]. \end{aligned}$$

This is long and a big mess, true (it's not really worth the effort to simplify), but the important point again is that, via a handful of rules, one can differentiate very complex functions by first breaking them down into manageable parts and then building the derivative back up again. At first these parts will be small, likely smaller than in our example, but with practice they'll get bigger and differentiation will get faster and less worrisome. And practice is key: there's nothing special about any of this beyond becoming accustomed to the techniques, and we hope we've convinced you that the techniques can be very useful.

6.4 EXERCISES

1. Find the derivative of y with respect to x for the following, using the rules in this chapter:
   - $y = 6$.
   - $y = 3x^2$.
   - $y = x^3 - 2x^2 - 1$.
   - $y = x^4 + 5x$.
   - $y = x^8$.
   - $y = x^{-3}$.
   - $y = ax^3 + 6$.
   - $y = 12x^{\frac{1}{2}} + c$.

[12] Again, if this is not clear, break them up further. In particular, let $d(x) = e^x$ and $k(x) = 3x^3 - 10x$ in $v(x)$ so that you can use the chain rule.