Essential Mathematics for the Political and Social Research

JEFF GILL
Cambridge University Press

Lecture Slides, Chapter 5: Elementary Scalar Calculus

Chapter 5 Objectives

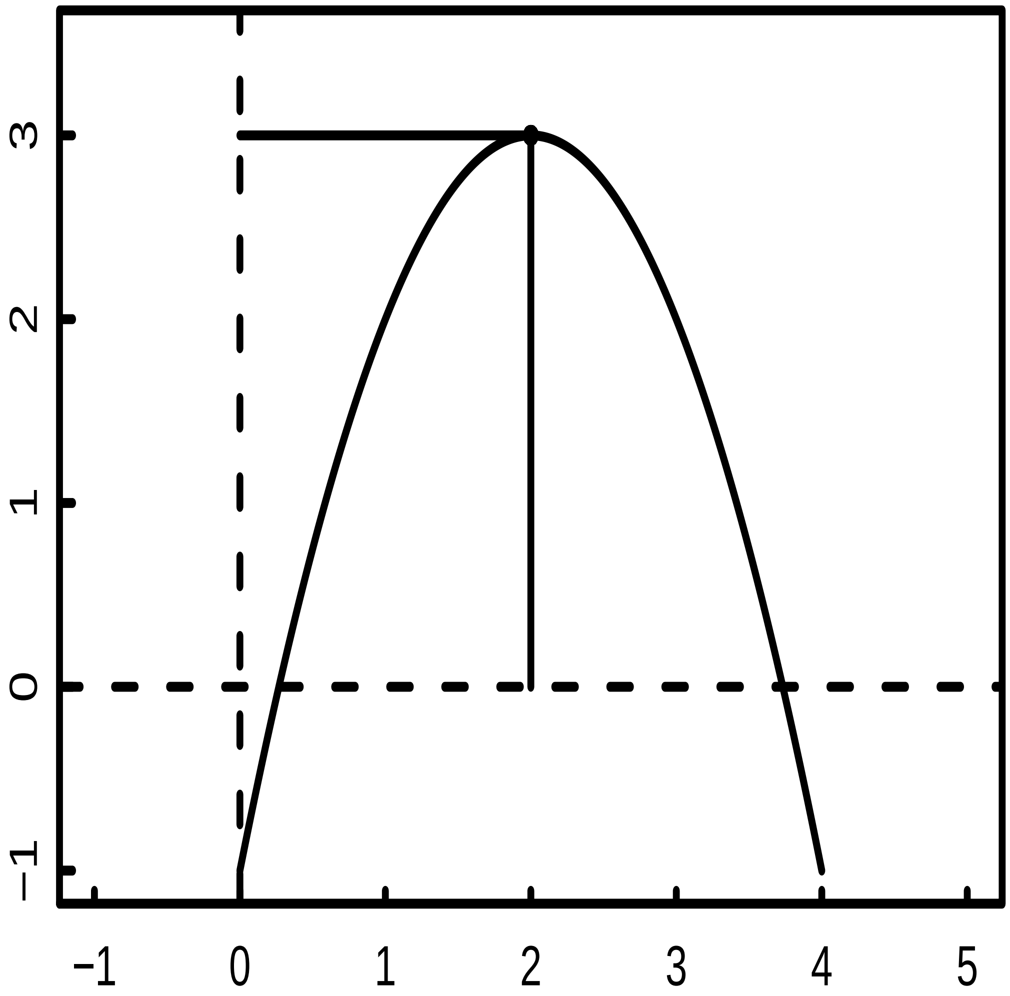
- This module introduces the basics of calculus operating on scalar quantities.
- Many find that the language and imagery of calculus are a lot more intimidating than the actual material (once they commit themselves to studying it).
- There are two primary tools of calculus to learn here: differentiation and integration.

Limits and Lines

- 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.
- We are specifically interested in how a function tends to or converges to some point or line in the limit.
- This shows "features" of the function that are not readily apparent otherwise.

Limits and Lines

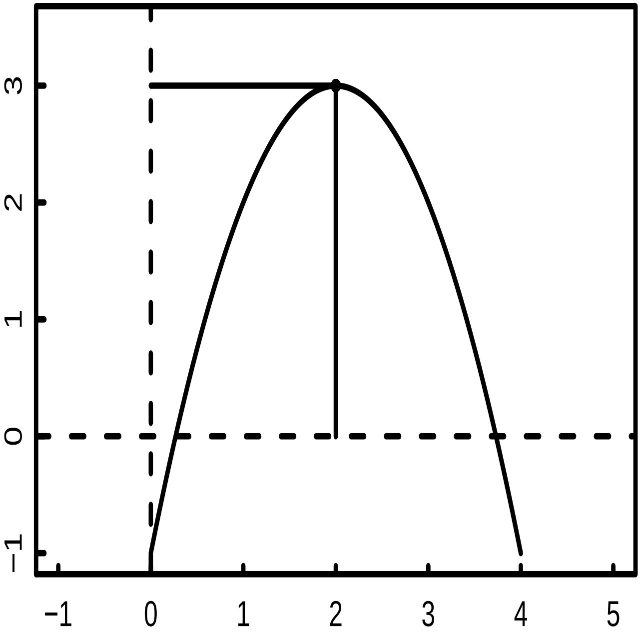
Consider the function $f(x) = 3 - (x - 2)^2$ over the domain (support) $[0 : 4]$. This function is *unimodal*, possessing a single maximum point, and it is *symmetric*, meaning that the shape is mirrored on either side of the line through the middle (which is the mode here). This function is graphed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



Limits and Lines

- 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



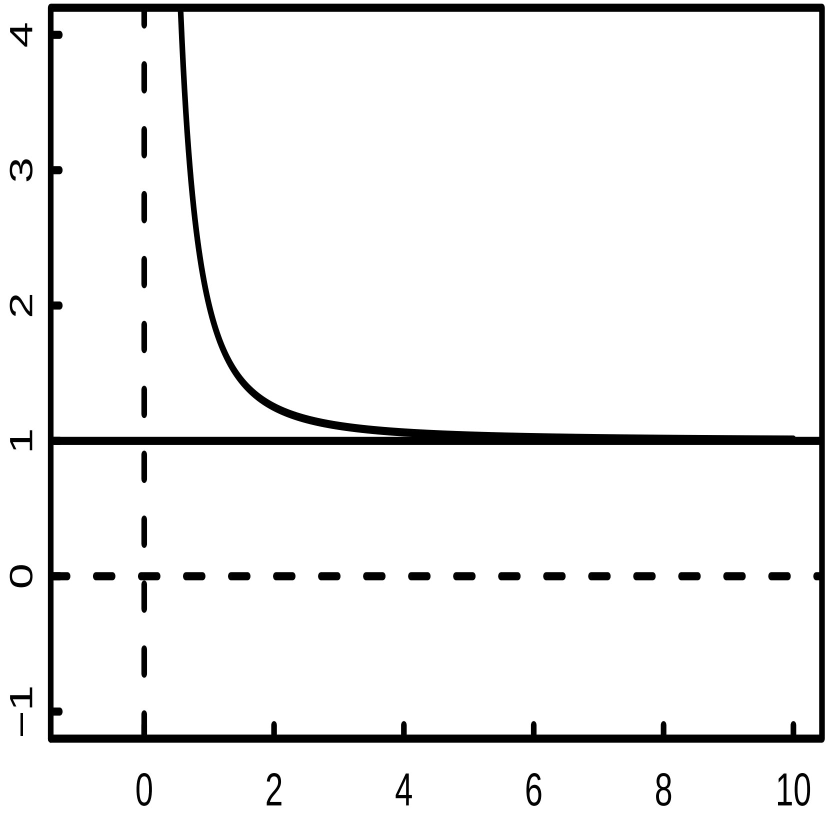
Limits and Lines

- It is shown in the graph as well as with the 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 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.

Limits and Lines

- 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.

Consider $f(x) = 1 + 1/x^2$ 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 the left?



Limits and Lines

- Let's look at tabulated values for $f(x) = 1 + 1/x^2$:

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

- As x gets arbitrarily large, $f(x)$ gets progressively closer to 1.
- What happens when we 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*: $\lim_{x\to\infty} f(x) = 1$.

Limits and Lines

- What happens at the limit $x = 0$ for $f(x) = 1 + 1/x^2$?
- 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: $\lim_{x \to 0^+} f(x) = \infty$.

Properties For Limits, 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

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

$\rightarrowtail$ Scalar Multiplication

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

$\rightarrowtail$ Division $\left(\lim_{x\to X} g(x) \neq 0\right)$

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

$\rightarrowtail$ Constants

$$\lim_{x\to X} k = k$$

$\rightarrowtail$ Natural Exponent

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

Limits of More Complex Functions

- **Quadratic Expression**.

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

- **Polynomial Ratio**.

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

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

Limits of More Complex Functions

- **Fractions and Exponents**.

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

- **Mixed Polynomials**.

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

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

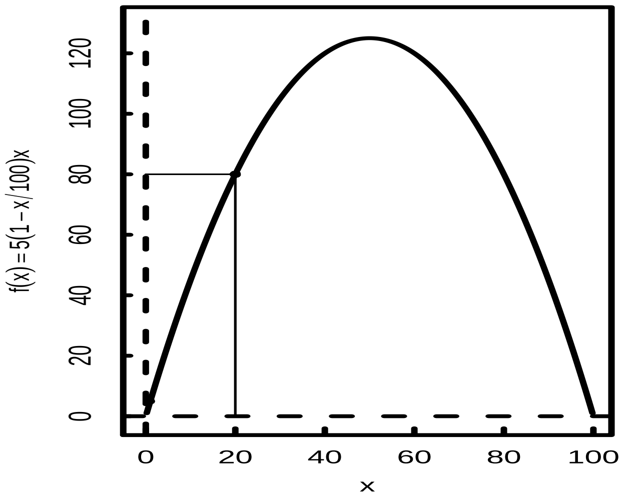
Understanding Rates, Changes, and Derivatives

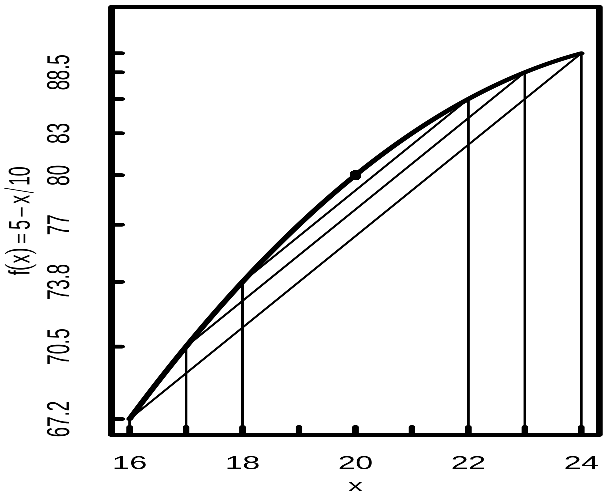
- We will look at a formal model from sociology that seeks to explain thresholds in voluntary racial segregation.
- Granovetter and Soong (1988) built on the foundational work of Thomas Schelling by mathematizing the idea that members of a racial group are progressively less likely to remain in a neighborhood 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 their 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 form of the function as:

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

Understanding Rates, Changes, and Derivatives

- Use the values number of whites $N_w = 100$ and tolerance $R = 5$, so $f(x) = 5(1 - \frac{x}{100})x = 5x - \frac{1}{20}x^2$.
- From the left panel 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.



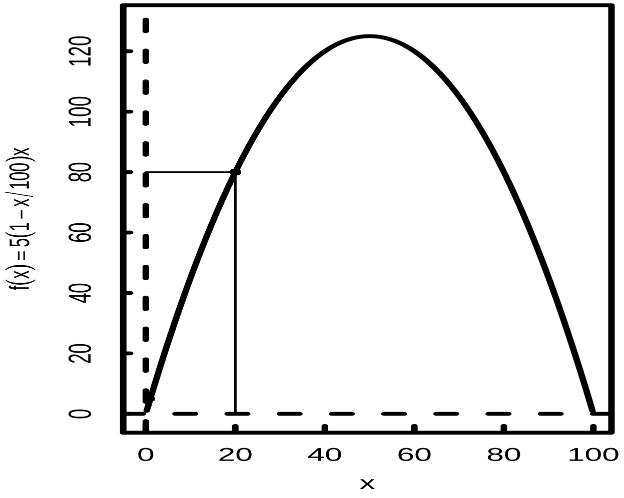


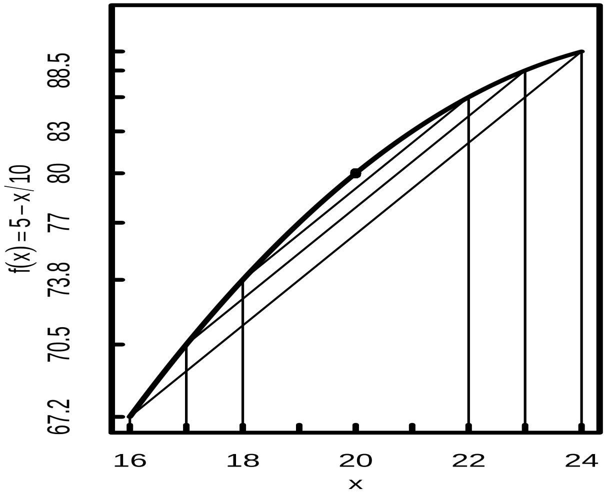
Understanding Rates, Changes, and Derivatives

- We are actually interested here in the *rate of change* of the tolerated number as opposed to the tolerated number itself:

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

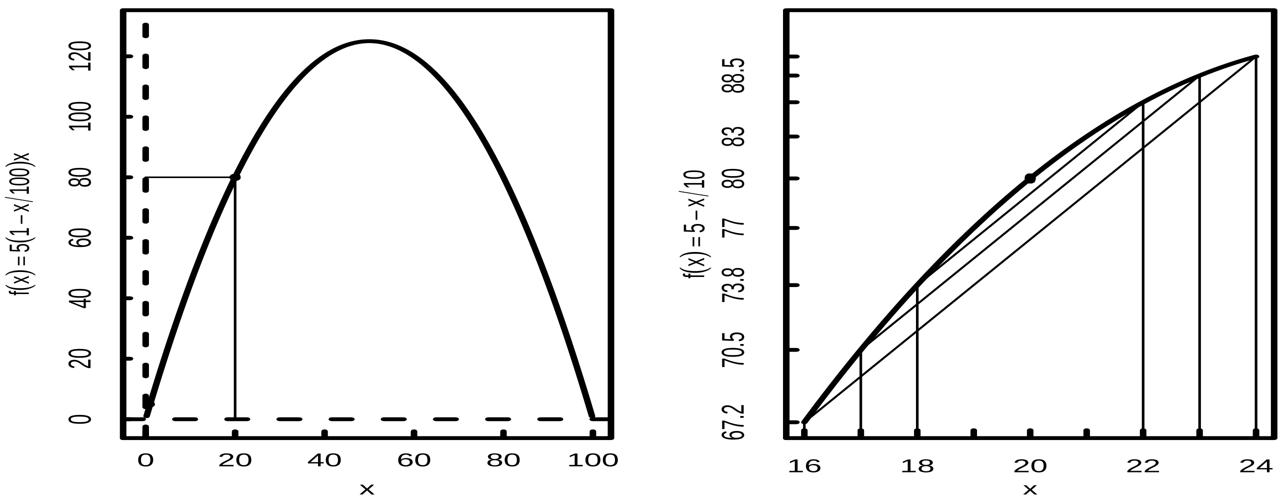
- Initially the rate of increase declines from a starting point until it reaches zero at the mode.
- Then the rate of decrease starts slowly and gradually picks up until the velocity is zero again.





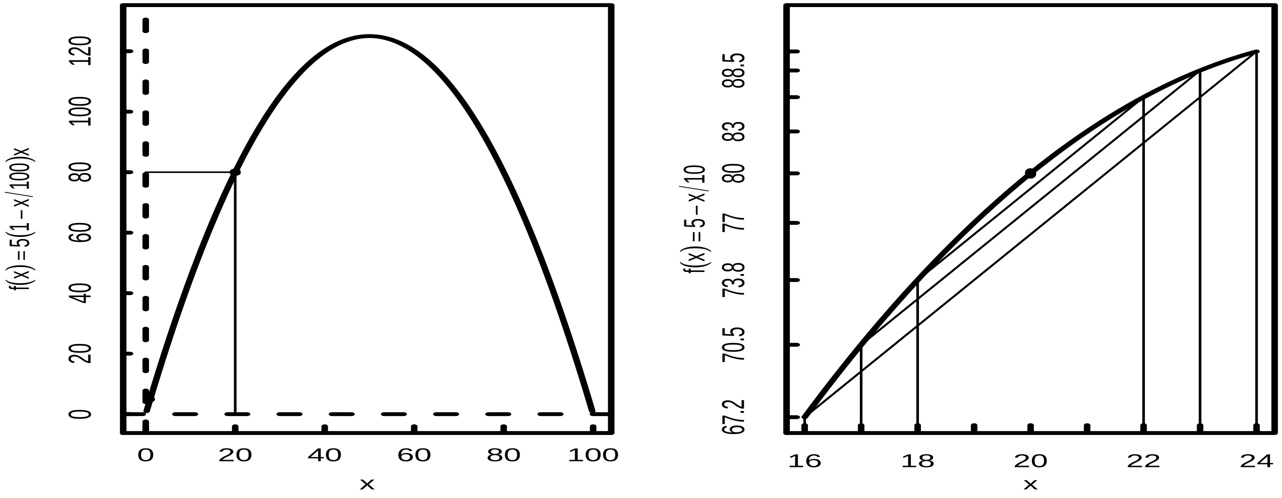
Understanding Rates, Changes, and Derivatives

- 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 the figure.
- A reasonable approximation can be made with line segments.
- The slope of the line segment is therefore an approximation to the instantaneous rate at $x = 20$, "rise-over-run," given by the segment
$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$
- 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.



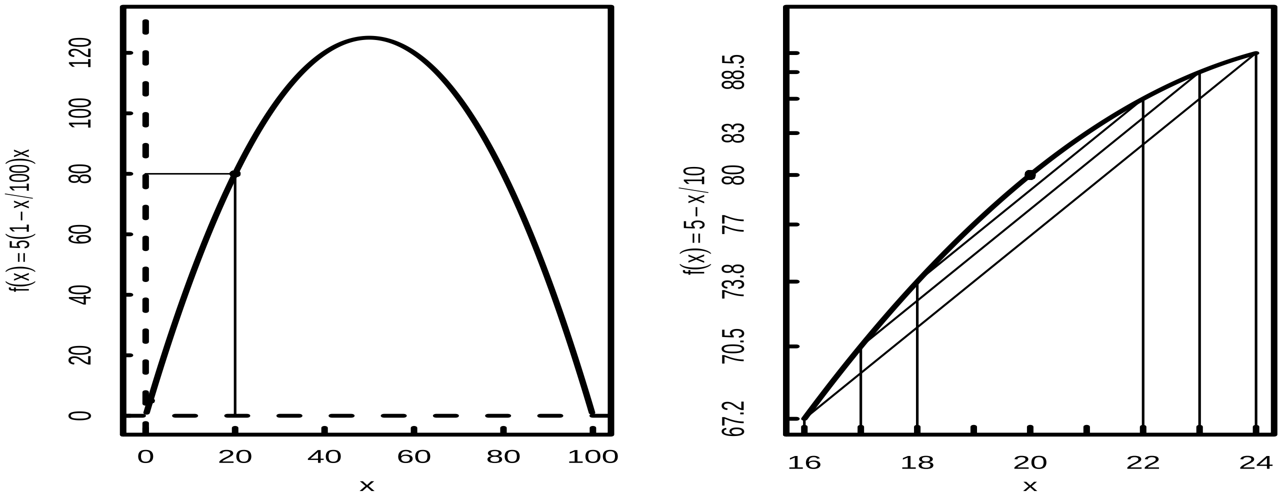
Understanding Rates, Changes, and Derivatives

- 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
$$m = \frac{f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)}{h} = \frac{f(x+h) - f(x)}{h}$$
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 the figure and has slope 2.6625.



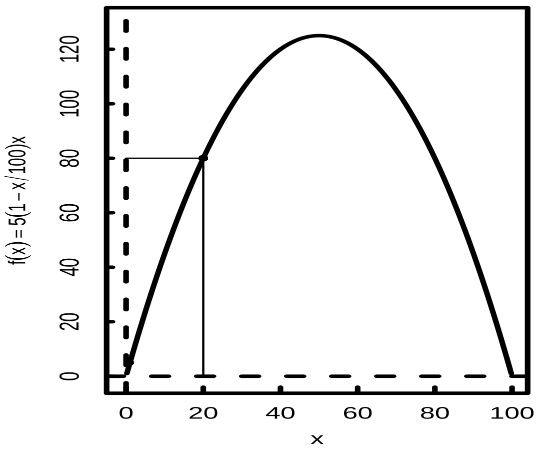
Understanding Rates, Changes, and Derivatives

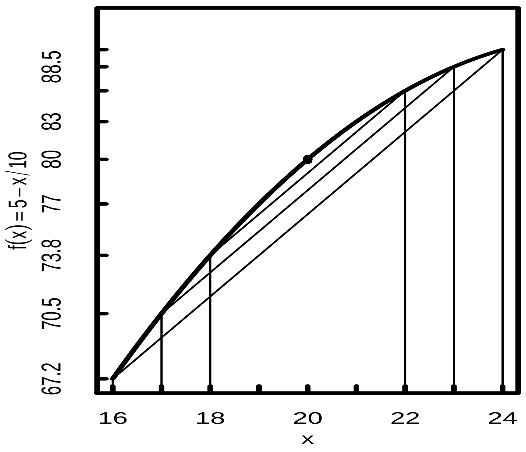
- This estimate is not quite right, but it is an average of a slightly faster rate of change (below) and a slightly slower rate of change (above).
- One way to improve the estimate is to decrease the width of the interval around the point.
- First go to 17–23 and then 18–22, and construct new line segments and therefore new estimates as shown in the second panel of the figure.
- At each reduction in interval width we are improving the estimate of the instantaneous rate of change at $x = 20$, so let's just keep going, making the estimate better and better.



The Derivative

- When should we stop? What if we never had to stop?
- 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 $\lim_{h \to 0}$ 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* to the curve at the point x, and is 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.





The Derivative

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

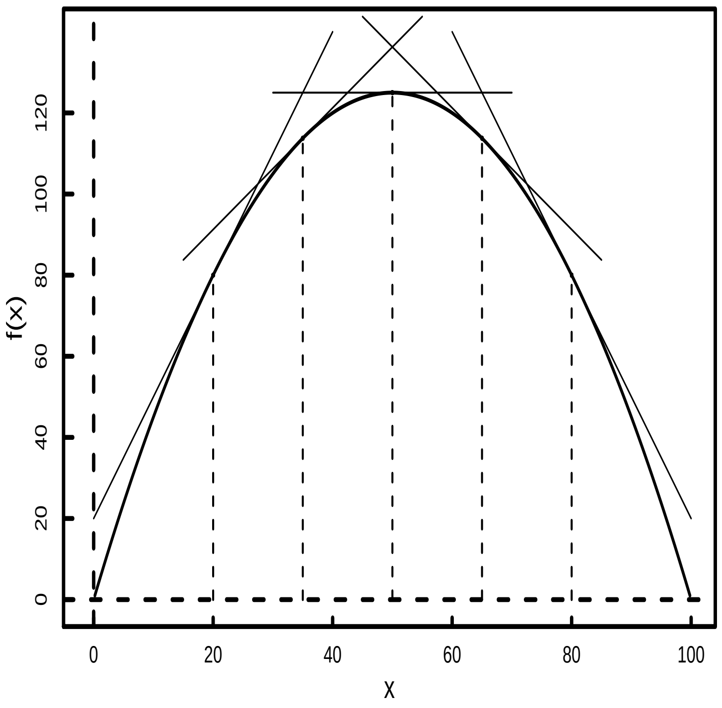
$$\begin{aligned}
\lim_{h\to 0} m(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{\left[5(x+h) - \frac{1}{20}(x+h)^2\right] - \left[5x - \frac{1}{20}x^2\right]}{h} \\
&= \lim_{h\to 0} \frac{5h - \frac{1}{20}(x^2 + 2xh + h^2) + \frac{1}{20}x^2}{h} \\
&= \lim_{h\to 0} \frac{5h - \frac{2}{20}xh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

- 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 $f'(20) = 3$.

The Derivative

- The figure shows tangent lines plotted at various points on $f(x)$.
- Note that the tangent line at the maxima is "flat," having slope zero.

Figure 1: Tangent Lines on $f(x) = 5x - \frac{1}{20}x^2$



Derivative Notation

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

Derivative Notation

- 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 $u = f(x)$.

- However, 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.

Derivative Summary, So Far

Summary of Derivative Theory

$\rightarrowtail$	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
$\rightarrowtail$	Definition	$f'(x) = \lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h}$
$\rightarrowtail$	Tangent Line	$f'(x)$ is the slope of the line tangent to $f()$ at x; this is the limit of the enclosed secant lines

No Corners

- The second existence condition in that list 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
$$\lim_{\Delta x \to 0^-} \frac{\Delta f(x)}{\Delta x} \neq \lim_{\Delta x \to 0^+} \frac{\Delta f(x)}{\Delta x}.$$
- So 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, $\Delta x \to 0^+$.
- 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.

Derivatives for Analyzing Legislative Committee Size

- Francis (1982) wanted to find criteria for determining "optimal" committee sizes in Congress or other legislatures, since committees are key organizational and procedural components of American (and some other) legislatures.

- 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, X, in the following partly-linear specifications:

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

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

- 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 before.

Derivatives for Analyzing Legislative Committee Size

- Since g is necessarily positive, increasing committee size increases decision costs.
- 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, adding
$$Y = Y_d + Y_e = a_d + b_d X^g + a_e + b_e X^k.$$
- This gives a single equation that expresses the full identified costs to members as a function of X.
- 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.

Derivatives On Polynomials

- The derivative operating on a polynomial multiplies the expression by the exponent, it decreases the exponent by one, and it gets rid of any isolated constant terms.
- The consequence in this case is that the first derivative of the Francis model is
$$\frac{d}{dx}Y = gb_d X^{g-1} + kb_e X^{k-1},$$
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.

Power Rule

- 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

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

- Examples include

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

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

Constant Rule

- 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).
- 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.

Sum Rule

- 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,

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

Product Rule

- 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 not a very difficult formula to remember due to its symmetry.
- For example,

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

(where we also used the sum property).

- Check this answer by multiplying the functions first and then taking the derivative

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

Quotient Rule

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

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

- For example,

$$\begin{aligned}\frac{d}{dx}\left(\frac{3x^2-3}{4x^3-2x}\right) &= \frac{(4x^3-2x)\frac{d}{dx}(3x^2-3) - (3x^2-3)\frac{d}{dx}(4x^3-2x)}{(4x^3-2x)(4x^3-2x)} \\ &= \frac{(4x^3-2x)(6x) - (3x^2-3)(12x^2-2)}{16x^6-16x^4+4x^2} \\ &= \frac{-6x^4+15x^2-3}{8x^6-8x^4+2x^2}.\end{aligned}$$

Quotient Rule

- 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*.

Chain Rule

- The *chain rule* provides a means of differentiating nested functions.
- Previously 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.

Chain Rule

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

Combining Derivative Operations

- 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 following calculation:

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

which requires the chain rule and other operations.

- Now define u such that

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

- This redefines the problem as

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

Combining Derivative Operations

- 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 $(y^2+1)^{-1}$.

Example: Productivity and Interdependence in Work Groups

- 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.
- 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.
- These are extreme cases, but the point is to show differences, so they are likely to be maximally revealing.

Example: Productivity and Interdependence in Work Groups

- 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 shows here is that if even one worker is totally unproductive, $p_j = 0$, then the entire organization is unproductive.

Example: Productivity and Interdependence in Work Groups

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

Example: Productivity and Interdependence in Work Groups

- 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 was 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.

Derivatives of Logarithms and Exponents

- 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.
- 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.

Derivatives of Logarithms and Exponents

- Derivatives of the logarithm are given by
$$\frac{d}{dx}\log(x) = \frac{1}{x},$$
for the natural log (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
$$\frac{d}{dx}\log_b(x) = \frac{1}{x\mathrm{ln(b)}}, \text{ because } \log_b(x) = \frac{\mathrm{ln(x)}}{\mathrm{ln(b)}}$$
with the associated chain rule version:
$$\frac{d}{dx}\log_b(u) = \frac{1}{u\mathrm{ln(b)}}\frac{d}{dx}u.$$

Logarithm Example

- Find the derivative of the logarithmic function $y = \log(2x^2 - 1)$:

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

Logarithmic Differentiation

- *Logarithmic differentiation* uses the log function to make the process of differentiating a difficult function easier.
- The idea is to log the function, take the derivative of this version, and then compensate back at the last step.
- For example,
$$y = \frac{(3x^2 - 3)^{\frac{1}{3}}(4x - 2)^{\frac{1}{4}}}{(x^2 + 1)^{\frac{1}{2}}},$$
which would be quite involved using the product rule, the quotient rule, and the chain rule.
- 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.

Logarithmic Differentiation

- 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:

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

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

Example: 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?

Example: Security Trade-Offs for Arms Versus Alliances

- Consider a state labeled i and its erstwhile ally labeled j.
- They each have military capability labeled M_i and M_j, correspondingly.
- This includes such factors as the numbers of soldiers, quantity and quality of military hardware, as well as geographic constraints.
- It is 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 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.

Example: Security Trade-Offs for Arms Versus Alliances

- 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.

Example: Security Trade-Offs for Arms Versus Alliances

- 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.

- The three marginal effects from the security equation are given by

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

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

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

Example: Security Trade-Offs for Arms Versus Alliances

- 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.
- The marginal effect of M_j includes the term T, meaing 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.

L'Hospital's Rule

- One problem, already mentioned, caused by zero is when it ends up in the denominator of a fraction.
- Differential 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* (Johannes Bernoulli) states that:

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

  provided that $g'(x) \neq 0$.
- 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.

L'Hospital's Rule

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

$$\begin{aligned}\lim_{x\to 0}\frac{x}{\log(1-x)} &= \lim_{x\to 0}\frac{\frac{d}{dx}x}{\frac{d}{dx}\log(1-x)}\\ &= \lim_{x\to 0}\frac{1}{\frac{1}{-(1-x)}} = -1.\end{aligned}$$

L'Hospital's Rule

- 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,
$$\lim_{x\to a} f(x)/g(x) = \lim_{x\to a} f'(x)/g'(x).$$

L'Hospital's Rule

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

$$\begin{aligned}
\lim_{x\to\infty} \frac{(\log(x))^2}{x^2 \log(x)} &= \lim_{x\to\infty} \frac{\frac{d}{dx}(\log(x))^2}{\frac{d}{dx}x^2 \log(x)} \\
&= \lim_{x\to\infty} \frac{2\log(x)\frac{1}{x}}{2x\log(x) + x^2\frac{1}{x}} \\
&= \lim_{x\to\infty} \frac{\log(x)}{x^2\log(x) + \frac{1}{2}x^2}.
\end{aligned}$$

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

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

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.

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 $f'(\hat{x}) = 0$.
- Interpretation: 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.

Rolle's Theorem, Intuition

- 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.

Rolle's Theorem

- 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.
- 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

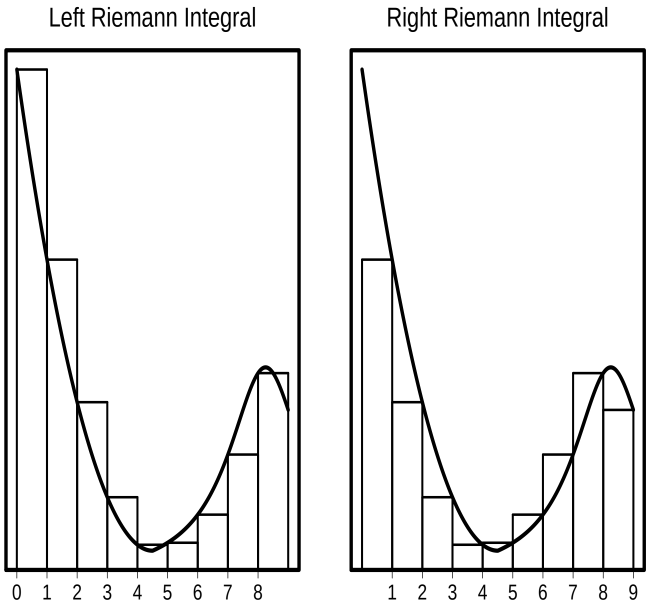
- Assume a function $f(x)$ that is continuous on the closed interval $[a\!:\!b]$ and differentiable on the open interval $(a\!:\!b)$.
- There is now guaranteed to be at least one point $\hat{x}$ in $(a\!:\!b)$ such that $f(b) - f(a) = f'(\hat{x})(b - a)$.
- This theorem just says that between the function 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 $f(a) = f(b) = 0$.

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., $f(x) \geq 0,\ \forall x \in [a\!:\!b]$):



- *Integration* is a calculus procedure for measuring areas and is as fundamental a process as differentiation.

Riemann Integrals

- So how would we go about measuring such an area?
- First "slice up" the area under the curve with a set of bars that are approximately as high as the curve at different places: like a histogram approximation of R where we simply sum up the sizes of the set of rectangles.
- 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 $f(a + ih)$.

Riemann Integrals

- 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.

Integrals As Limits

- 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.

- The limit of the Riemann process is the true area under the curve.

Integrals As Limits

- There is specific terminology for what we have done: the *definite integral*:
$$R = \int_a^b f(x)dx,$$
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 make this "definite" since they 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.

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 right Riemann integral:

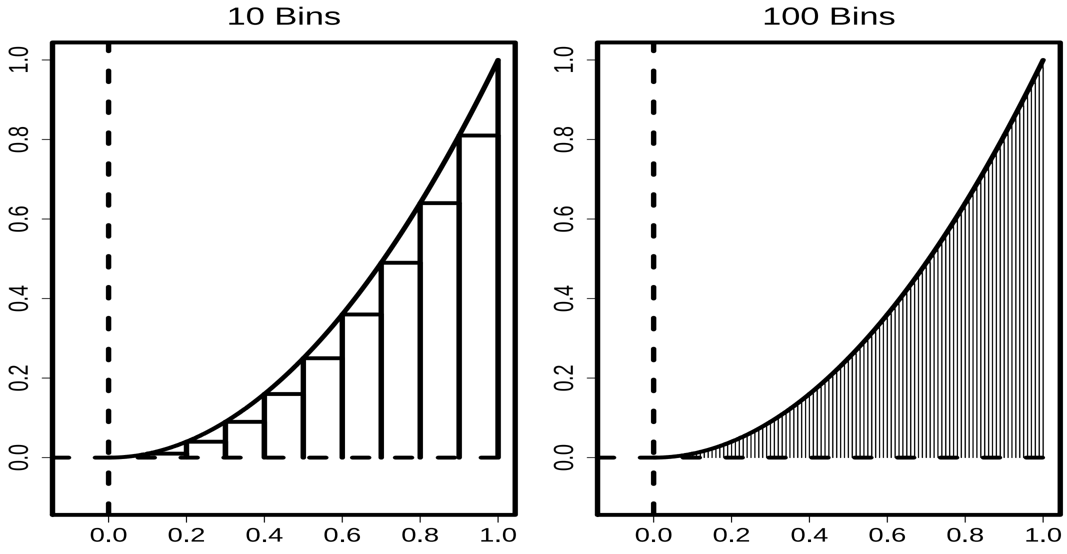
$$
\begin{aligned}
R = \lim_{h\to\infty} \sum_{i=1}^{h} \frac{1}{h} f(x) &= \lim_{h\to\infty} \sum_{i=1}^{h} \frac{1}{h}(i/h)^2 \\
&= \lim_{h\to\infty} \frac{1}{h^3} \sum_{i=1}^{h} i^2 = \lim_{h\to\infty} \frac{1}{h^3} \frac{h(h+1)(2h+1)}{6} \\
&= \lim_{h\to\infty} \frac{1}{6}(2 + \frac{3}{h} + \frac{1}{h^2})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

- The step out of the summation was accomplished by a well-known trick:

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

Application: Limits of a Riemann Integral

- This process is shown in the figure 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.
- Riemann Sums for $f(x) = x^2$ over [0:1]:



Riemann Sums

- 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.
- 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.

The Antiderivative

- 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 as done here.
- The derivative of the antiderivative returns the original form: $F'(x) = f(x)$.

The Antiderivative

- The antiderivative is a function in the regular sense, so we can treat it as such and apply the Mean Value Theorem discussed on 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:

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

The Antiderivative

- 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:

$$F(x_1) - F(x_1), F(x_2) - F(x_2), \ldots F(x_{H-1}) - F(x_{H-1}).$$

- 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

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

  which is just $\int_a^b f(x)dx$ from above.

The Fundamental Theorem of Calculus

- 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.

Integrating Polynomials with Antiderivatives

- The use of antiderivatives for solving definite integrals is especially helpful with polynomial functions.
- 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

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

Integrating Polynomials with Antiderivatives

- 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.

Properties of Definite Integrals

Properties of Definite Integrals

$\rightarrowtail$ Constants $\quad \int_a^b kf(x)dx = k\int_a^b f(x)dx$

$\rightarrowtail$ Additive Property $\quad \int_a^b (f(x)+g(x))dx$

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

$\rightarrowtail$ Linear Functions $\quad \int_a^b (k_1 f(x) + k_2 g(x))dx$

$$= k_1 \int_a^b f(x)dx + k_2 \int_a^b g(x)dx$$

$\rightarrowtail$ Intermediate Values $\quad$ 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 $\quad \int_a^b f(x)dx = -\int_b^a f(x)dx$

(The first two properties are obvious by now and the third is just a combination of the first two.)

Integration By Pieces

- The fourth property above 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.
- For example 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:

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

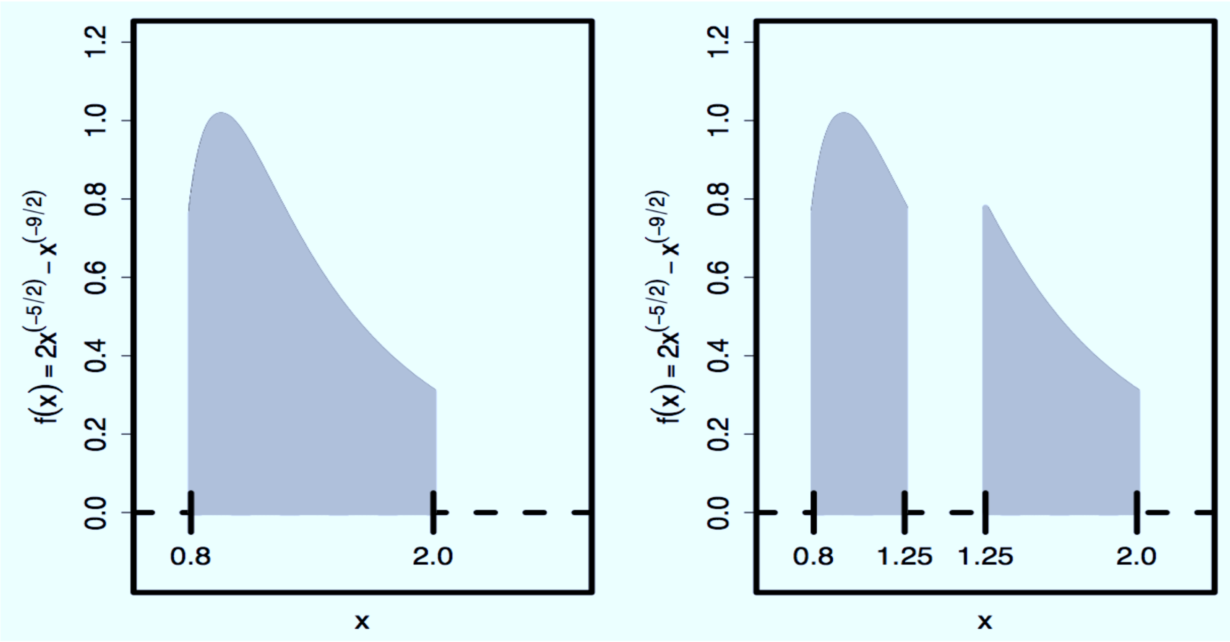
$$= \left[\left(-\frac{2}{3}\right)2x^{-\frac{3}{2}} - \left(-\frac{2}{7}\right)x^{-\frac{7}{2}}\right]\Bigg|_{0.8}^{1.25}$$

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

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

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

Integration By Pieces, Illustration



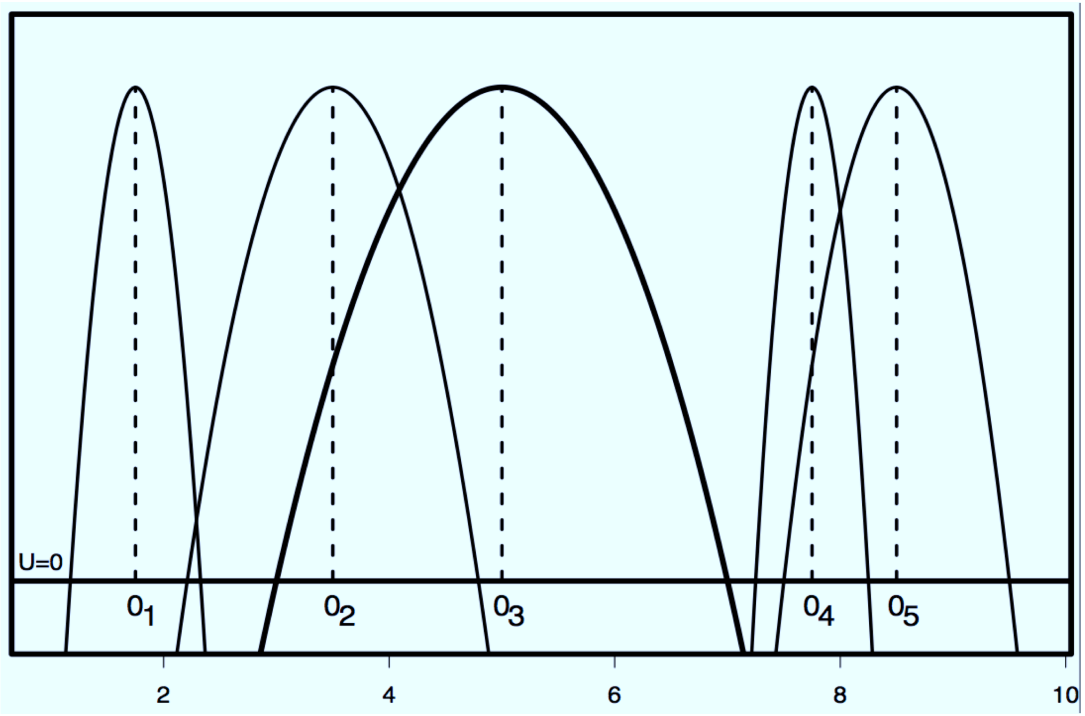
The Median Voter Theorem

- 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.
- This is shown in the following figure based on Black's figure (1958, p.15).
- Displayed 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.

The Median Voter Theorem

- 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.



The Median Voter Theorem

- These utility functions are often drawn or assumed to be parabolic shapes and the 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 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
$$V_2\colon\ f(x) = 10 - (3.5 - x)^2(6) \qquad V_3\colon\ f(x) = 10 - (5 - x)^2(2.5).$$

- 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 Median Voter Theorem

- 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 (striking out the 10's and the multiplication by -1 here since they will cancel each other anyway).
- First expand the squares:

$$\begin{aligned} y &= (3.5-x)^2(6) & \qquad y &= (5-x)^2(2.5) \\ y &= 6x^2 - 42x + 73.5 & \qquad y &= 2.5x^2 - 25x + 62.5. \end{aligned}$$

The Median Voter Theorem

- Now equate the two expressions and solve:

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

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

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

- 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:

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

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

The Median Voter Theorem

- 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):

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

$$x_3 = \frac{-(25) \pm \sqrt{(25)^2 - 4(-2.5)(-52.5)}}{2(-2.5)} = 3 \text{ or } 7.$$

- 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 Median Voter Theorem

- 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.

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:

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

- 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*.

Calculating Indefinite Integrals

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

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

- Where k is some arbitrary unknown constant.
- We have to add a constant k that could have been there but was lost due to the derivative function operating on it ($\frac{d}{dx}(k) = 0$).
- Instead of a numeric value, we get a defined relationship between the functions

$$f(x) = (1-3x)^{\frac{1}{2}},$$

and

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

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 $\frac{d}{dx}\log(x) = \frac{1}{x}$.
- Recall also 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.
- 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 $\frac{du}{dx}$.

Example Calculations

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

Example Calculations

- 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:

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

Integration by Parts

- 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.

Integration by Parts

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

Integration by Parts

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

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

Integration by Parts

- 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.
- There is a little bit of “art” here as it is possible to choose terms that make the problem harder.

Integration by Parts Example

- Calculate:
$$\int x \log(x) dx,$$
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 $v = \int dv dx = \int \log(x) dx$.

- Picking the second arrangement gives the full mapping:

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

Integration by Parts Example

- 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:

$$\begin{aligned}\int udv = uv - \int vdu. &= (\log(x))\left(\frac{1}{2}x^2\right) - \int \left(\frac{1}{2}x^2\right)\left(\frac{1}{x}dx\right)\\ &= \frac{1}{2}x^2\log(x) - \int \frac{1}{2}xdx\\ &= \frac{1}{2}x^2\log(x) - \frac{1}{2}(\frac{1}{2}x^2) + k\\ &= \frac{1}{2}x^2\log(x) - \frac{1}{4}x^2 + k.\end{aligned}$$

- We benefited from a very simple integral in the second stage, because the antiderivative of $\frac{1}{2}x$ is straightforward.

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.

Application: The Gamma Function

- Suppose we wanted to integrate the gamma function for a known value of ω, say 3.
- The resulting integral to calculate is
$$\int_0^\infty t^2 e^{-t} dt.$$
- 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
- The other logical split is
$$\boxed{\begin{array}{ll} u = t^2 & dv = e^{-t} \\ du = 2t & v = -e^{-t}. \end{array}}$$

Application: The Gamma Function

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

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

- 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).

Application: The Gamma Function

- Finish the integration:

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

- 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.

Utility Functions from Voting

- In the article by Riker and Ordeshook (1968) they also 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.
- 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.

Utility Functions from Voting

- 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 cents.
- 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.
- 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.

Utility Functions from Voting

- 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.

- 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.