BASIC CALCULUS REFRESHER

Ismor Fischer, Ph.D. Dept. of Statistics UW-Madison

1. Introduction.

This is a very condensed and simplified version of basic calculus, which is a prerequisite for many courses in Mathematics, Statistics, Engineering, Pharmacy, etc. It is *not* comprehensive, and absolutely *not* intended to be a substitute for a one-year freshman course in differential and integral calculus. You are strongly encouraged to do the included **Exercises** to reinforce the ideas. Important mathematical terms are in **boldface**; key formulas and concepts are boxed and highlighted. To view a color .pdf version of this document (recommended), see http://www.stat.wisc.edu/~ifischer.

2. Exponents – Basic Definitions and Properties

For *any* real number **base** x, we define **powers** of x: $x^0 = 1$, $x^1 = x$, $x^2 = x \cdot x$, $x^3 = x \cdot x \cdot x$, etc. (The exception is 0^0, which is considered **indeterminate**.) Powers are also called **exponents**.

Examples: $5^0 = 1$, $(-11.2)^1 = -11.2$, $(8.6)^2 = 8.6 \times 8.6 = 73.96$, $10^3 = 10 \times 10 \times 10 = 1000$, $(-3)^4 = (-3) \times (-3) \times (-3) \times (-3) = 81$.

Also, we can define *fractional* exponents in terms of **roots**, such as $x^{1/2} = \sqrt{x}$, the square root of x. Similarly, $x^{1/3} = \sqrt[3]{x}$, the cube root of x, $x^{2/3} = \left(\sqrt[3]{x}\right)^2$, etc. In general, we have $\boxed{x^{m/n} = \left(\sqrt[n]{x}\right)^m}$, i.e., the n^{th} root of x, raised to the m^{th} power.

Examples: $64^{1/2} = \sqrt{64} = 8$, $64^{3/2} = \left(\sqrt{64}\right)^3 = 8^3 = 512$, $64^{1/3} = \sqrt[3]{64} = 4$, $64^{2/3} = \left(\sqrt[3]{64}\right)^2 = 4^2 = 16$.

Finally, we can define *negative* exponents: $\boxed{x^{-r} = \dfrac{1}{x^r}}$. Thus, $x^{-1} = \dfrac{1}{x^1}$, $x^{-2} = \dfrac{1}{x^2}$, $x^{-1/2} = \dfrac{1}{x^{1/2}} = \dfrac{1}{\sqrt{x}}$, etc.

Examples: $10^{-1} = \dfrac{1}{10^1} = 0.1$, $7^{-2} = \dfrac{1}{7^2} = \dfrac{1}{49}$, $36^{-1/2} = \dfrac{1}{\sqrt{36}} = \dfrac{1}{6}$, $9^{-5/2} = \dfrac{1}{\left(\sqrt{9}\right)^5} = \dfrac{1}{3^5} = \dfrac{1}{243}$.

Properties of Exponents

1. $x^a \cdot x^b = x^{a+b}$ Examples: $x^3 \cdot x^2 = x^5$, $x^{1/2} \cdot x^{1/3} = x^{5/6}$, $x^3 \cdot x^{-1/2} = x^{5/2}$

2. $\dfrac{x^a}{x^b} = x^{a-b}$ Examples: $\dfrac{x^5}{x^3} = x^2$, $\dfrac{x^3}{x^5} = x^{-2}$, $\dfrac{x^3}{x^{1/2}} = x^{5/2}$

3. $(x^a)^b = x^{ab}$ Examples: $(x^3)^2 = x^6$, $(x^{-1/2})^7 = x^{-7/2}$, $(x^{2/3})^{5/7} = x^{10/21}$

3. *Functions and Their Graphs*

Input x → f → Output y

Descartes ~ 1640

If a quantity y always depends on another quantity x in such a way that every value of x corresponds to *one and only one* value of y, then we say that "y is a **function** of x," written $y = f(x)$; x is said to be the **independent variable**, y is the **dependent variable**. (Example: "Distance traveled per hour (y) *is a function of* velocity (x).") For a given function $y = f(x)$, the set of all ordered pairs of (x, y)-values that *algebraically* satisfy its equation is called the **graph** of the function, and can be represented *geometrically* by a collection of points in the XY-plane. (Recall that the ***XY*-plane** consists of two perpendicular copies of the **real number line** – a horizontal ***X*-axis**, and a vertical ***Y*-axis** – that intersect at a reference point (0, 0) called the **origin**, and which partition the plane into four disjoint regions called **quadrants**. Every point P in the plane can be represented by the **ordered pair** (x, y), where the first value is the ***x*-coordinate** – indicating its horizontal position relative to the origin – and the second value is the ***y*-coordinate** – indicating its vertical position relative to the origin. Thus, the point $P(4, 7)$ is 4 units to right of, and 7 units up from, the origin.)

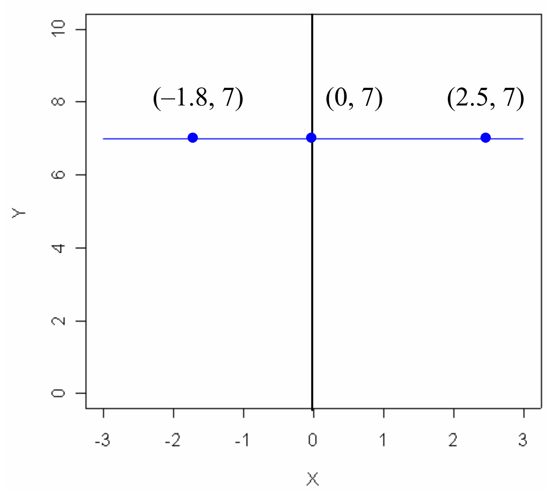
Examples: $y = f(x) = 7$; $y = f(x) = 2x + 3$; $y = f(x) = x^2$; $y = f(x) = x^{1/2}$; $y = f(x) = x^{-1}$; $y = f(x) = 2^x$.

The first three are examples of **polynomial** functions. (In particular, the first is **constant**, the second is **linear**, the third is **quadratic**.) The last is an **exponential** function; note that x is an exponent!

Let's consider these examples, one at a time.

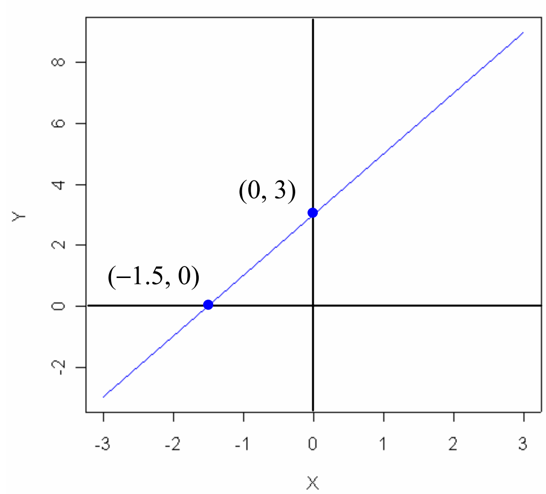
- $y = f(x) = 7$: If x = *any* value, then $y = 7$. That is, no matter what value of x is chosen, the value of the height y remains at a **constant** level of 7. Therefore, all points that satisfy this equation must have the form $(x, 7)$, and thus determine the graph of a horizontal line, 7 units up. A few typical points are plotted in the figure.

  Exercise: What would the graph of the equation $y = -4$ look like? $x = -4$? $y = 0$? $x = 0$?



- $y = f(x) = 2x + 3$: If $x = 0$, then $y = f(0) = 2(0) + 3 = 3$, so the point (0, 3) is on the graph of this function. Likewise, if $x = -1.5$, then $y = f(-1.5) = 2(-1.5) + 3 = 0$, so the point (–1.5, 0) is also on the graph of this function. (However, many points, such as (1, 1), do not satisfy the equation, and so do not lie on the graph.) The set of all points (x, y) that *do* satisfy this **linear** equation forms the graph of a line in the XY-plane, hence the name.

  Exercise: What would the graph of the line $y = x$ look like? $y = -x$? The **absolute value** function $y = |x|$?



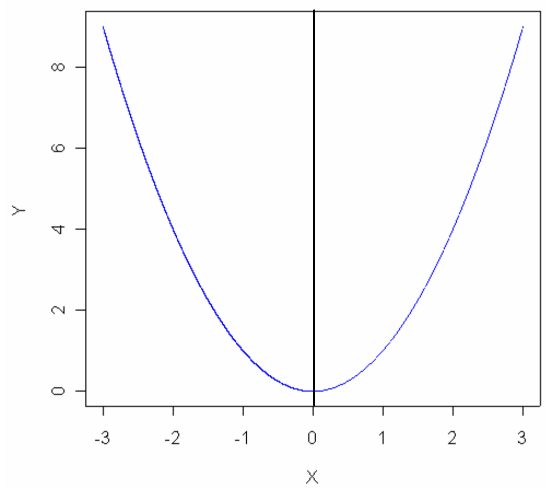
Notice that the line has the generic equation $\boxed{y = f(x) = mx + b}$, where b is the ***Y*-intercept** (in this example, $b = +3$), and m is the **slope** of the line (in this example, $m = +2$). In general, the slope of any line is defined as the ratio of "height change" Δy to "length change" Δx, that is,

$$\boxed{m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}}$$

for *any* two points (x_1, y_1) and (x_2, y_2) that lie on the line. For example, for the two points (0, 3) and (−1.5, 0) on our line, the slope is $m = \frac{\Delta y}{\Delta x} = \frac{0 - 3}{-1.5 - 0} = 2$, which confirms our observation.

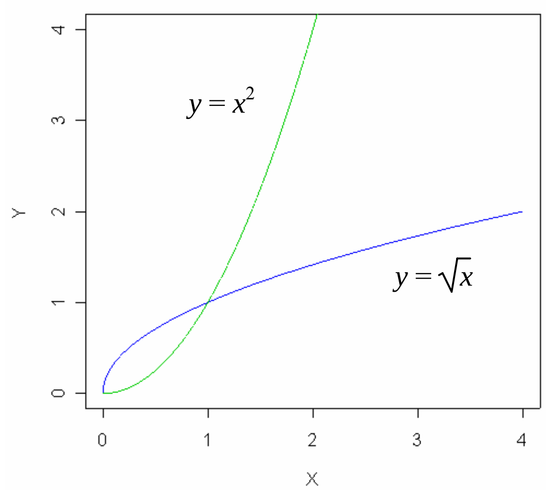
- $y = f(x) = x^2$: This is not the equation of a straight line (because of the "squaring" operation). The set of all points that satisfies this **quadratic** equation – e.g., (–3, 9), (–2, 4), (–1, 1), (0, 0), (1, 1), (2, 4), (3, 9), etc. – forms a curved **parabola** in the *XY*-plane. (In this case, the curve is said to be **concave up**, i.e., it "holds water." Similarly, the graph of $-x^2$ is **concave down**; it "spills water.")

  Exercise: How does this graph differ from $y = f(x) = x^3$? x^4? Find a pattern for $y = x^n$, for $n = 1, 2, 3, 4, \ldots$



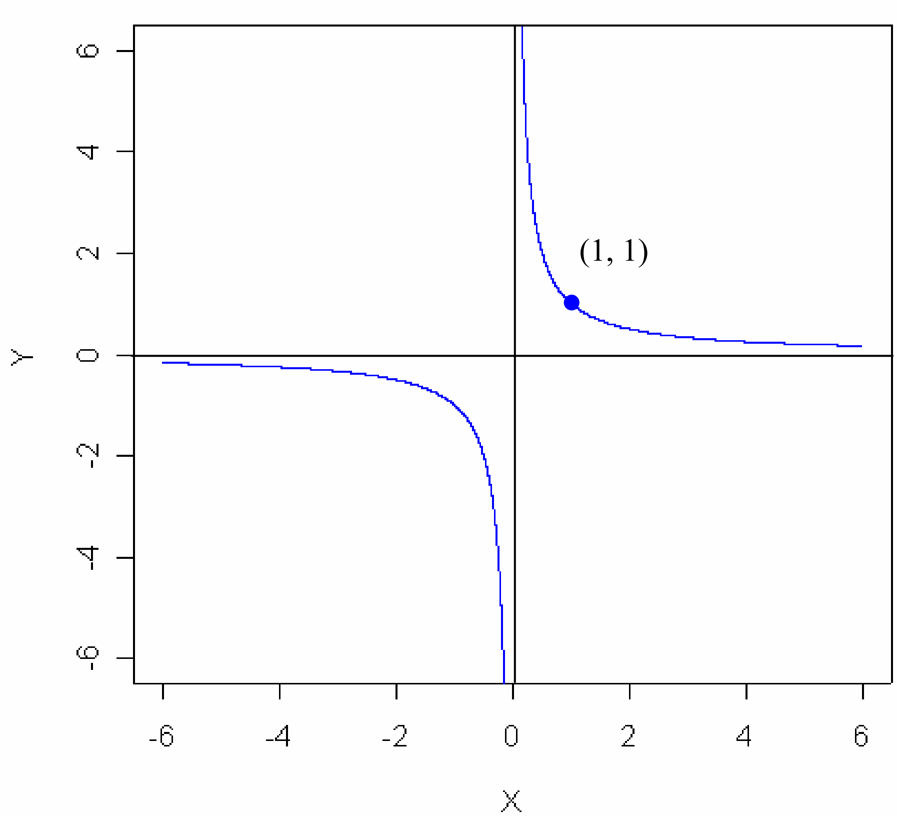
- $y = f(x) = x^{1/2} = \sqrt{x}$: The "square root" operation is not defined for negative values of x (e.g., $\sqrt{-64}$ does not exist as a real number, since both $(+8)^2 = +64$ and $(-8)^2 = +64$.) Hence the real-valued **domain** of this function is restricted to $x \geq 0$ (i.e., positive values and zero), where the "square root" operation *is* defined (e.g., $\sqrt{+64} = +8$). Pictured here is its graph, along with the first-quadrant portion of $y = x^2$ for comparison.

  Exercise: How does this graph differ from $y = f(x) = x^{1/3} = \sqrt[3]{x}$? *Hint*: What is its domain? (E.g., $\sqrt[3]{+64} = ??$, $\sqrt[3]{-64} = ???$) Graph this function together with $y = x^3$.



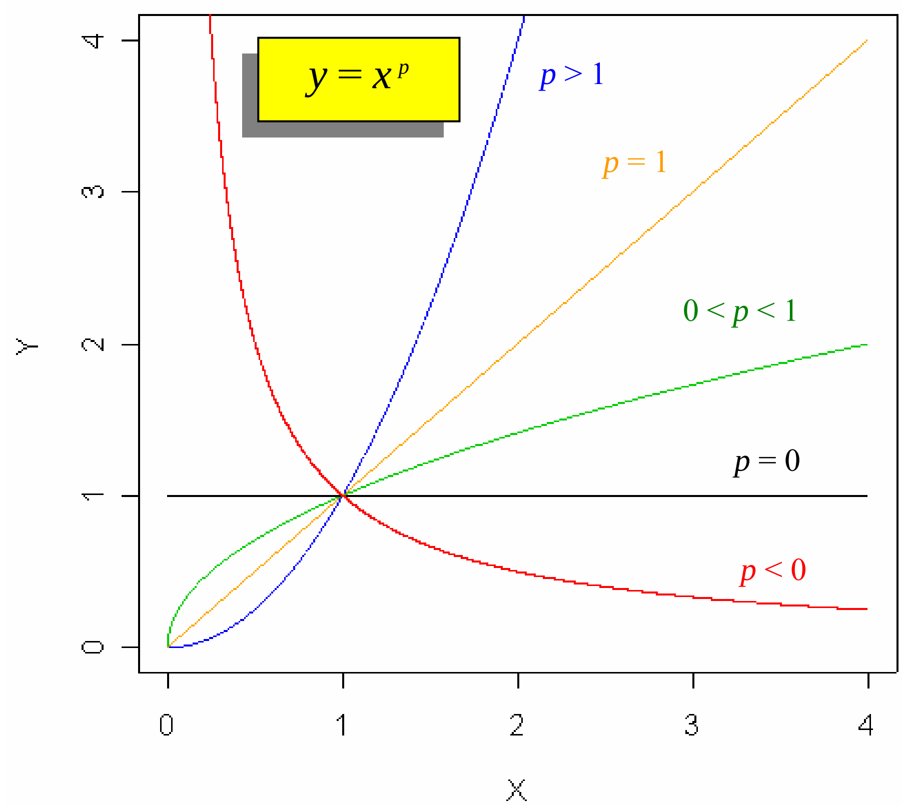
The "square" operation x^2 and "square root" operation $x^{1/2} = \sqrt{x}$ are examples of **inverse functions** of one another, for $x \geq 0$. That is, the effect of applying of either one, followed immediately by the other, lands you back to where you started from. More precisely, starting with a domain value x, the **composition** of the function $f(x)$ with its inverse – written $f^{-1}(x)$ – is equal to the initial value x itself. As in the figure above, when graphed together, the two functions exhibit symmetry across the "diagonal" line $y = x$ (not explicitly drawn).

- $y = f(x) = x^{-1} = \frac{1}{x}$: This is a bit more delicate. Let's first restrict our attention to *positive* domain x-values, i.e., $x > 0$. If $x = 1$, then $y = f(1) = \frac{1}{1} = 1$, so the point (1, 1) lies on the graph of this function. Now from here, as x grows *larger* (e.g., $x = 10, 100, 1000$, etc.), the values of the height $y = \frac{1}{x}$ $\left(\text{e.g., } \frac{1}{10} = 0.1,\ \frac{1}{100} = 0.01,\ \frac{1}{1000} = 0.001, \text{ etc.}\right)$ become *smaller*, *although they never actually reach* 0. Therefore, as we continue to move to the right, the graph approaches the X-axis as a **horizontal asymptote**, without ever actually touching it. Moreover, as x gets *smaller* from the point (1, 1) on the graph (e.g., $x = 0.1, 0.01, 0.001$, etc.), the values of the height $y = \frac{1}{x}$ $\left(\text{e.g., } \frac{1}{0.1} = 10,\ \frac{1}{0.01} = 100,\ \frac{1}{0.001} = 1000, \text{ etc.}\right)$ become *larger*. Therefore, as we continue to move to the left, the graph shoots upwards, approaching the Y-axis as a **vertical asymptote**, without ever actually touching it. (If $x = 0$, then y becomes infinite ($+\infty$), which is **undefined** as a real number, so $x = 0$ is not in the domain of this function.) A similar situation exists for *negative* domain x-values, i.e., $x < 0$. This is the graph of a **hyperbola**, which has two symmetric **branches**, one in the first quadrant and the other in the third.



Exercise: How does this graph differ from that of $y = f(x) = x^{-2} = \frac{1}{x^2}$? Why?

NOTE: The preceding examples are special cases of **power functions**, which have the general form $y = x^p$, for any real value of p, for $x > 0$. If $p > 0$, then the graph starts at the origin and continues to rise to infinity. (In particular, if $p > 1$, then the graph is **concave up**, such as the parabola $y = x^2$. If $p = 1$, the graph is the straight line $y = x$. And if $0 < p < 1$, then the graph is **concave down**, such as the parabola $y = x^{1/2} = \sqrt{x}$.) However, if $p < 0$, such as $y = x^{-1} = \frac{1}{x}$, or $y = x^{-2} = \frac{1}{x^2}$, then the Y-axis acts as a vertical asymptote for the graph, and the X-axis is a horizontal asymptote.



Exercise: Why is $y = x^x$ <u>not</u> a power function? Sketch its graph for $x > 0$.

Exercise: Sketch the graph of the **piecewise-defined** functions

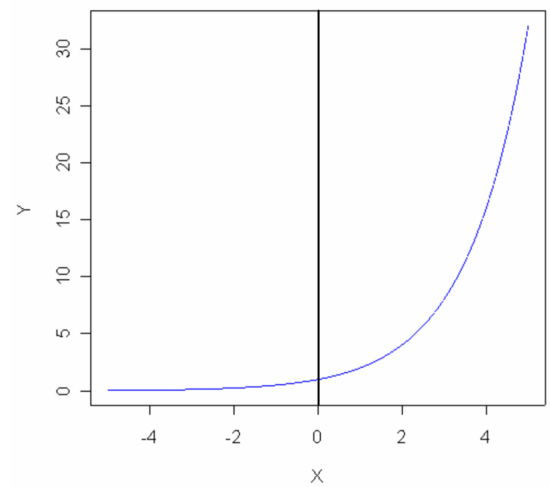
$$f(x) = \begin{cases} x^2, & \text{if } x \le 1 \\ x^3, & \text{if } x > 1 \end{cases}$$

This graph is the parabola $y = x^2$ up to and including the point (1, 1), then picks up with the curve $y = x^3$ after that. Note that this function is therefore **continuous** at $x = 1$, and hence for *all* real values of x.

$$g(x) = \begin{cases} x^2, & \text{if } x \le 1 \\ x^3 + 5, & \text{if } x > 1. \end{cases}$$

This graph is the parabola $y = x^2$ up to and including the point (1, 1), but then abruptly changes over to the curve $y = x^3 + 5$ after that, starting at (1, 6). Therefore, this graph has a break, or "jump discontinuity," at $x = 1$. (Think of switching a light from off = 0 to on = 1.) However, since it *is* continuous before and after that value, g is described as being **piecewise continuous**.

- $y = f(x) = 2^x$: This is *not* a power function! The graph of this increasing function is an **exponential growth** curve, which doubles in height y with every "unit increase" (i.e., + 1) in x. (Think of a population of bacteria that doubles its size ever hour.) If the exponent is changed from x to $-x$, then the resulting graph represents an **exponential decay** curve, *decreasing* by a constant factor of one-half from left to right. (Think of "half-life" of a radioactive isotope, e.g., carbon dating of prehistoric artifacts, fossils, etc.)



NOTE: For general **exponential functions** $y = f(x) = b^x$, the base b can be *any* positive constant. The resulting graph *increases* if $b > 1$ (e.g., $b = 2$ in the previous example), and *decreases* if $b < 1$ (e.g., $b = 1/2$). **Exercise:** How do the graphs of the functions $(1/2)^x$ and 2^{-x} compare, and why?

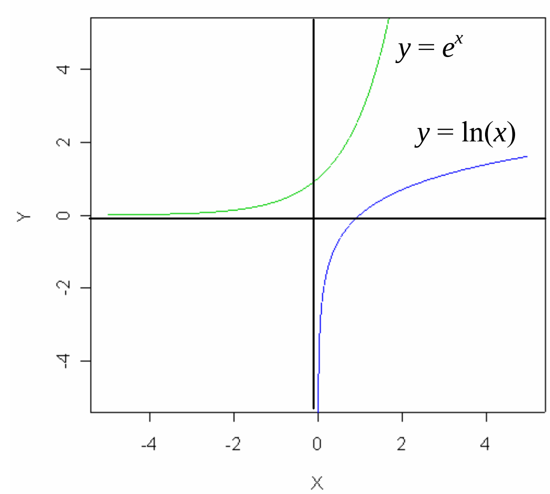
One very important special choice for calculus applications is called $e = 2.71828\ldots$, labeled for Swiss mathematician Leonhard Euler (pronounced "oiler"), 1707-1783. The resulting function, $y = f(x) = e^x$, is sometimes considered as "THE" exponential function, denoted **exp(x)**. Other exponential functions $y = f(x) = b^x$ can be equivalently expressed as $y = f(x) = e^{ax}$ (via $b = e^a$), with constant $a > 0$ (growth) or $a < 0$ (decay). For example, 2^x can be written $e^{0.6931\,x}$, so $a = 0.6931$. But given b, how do we get a?

The inverse of the exponential function $y = b^x$ is, *by definition*, the **logarithm** function $y = \log_b(x)$. That is, for a given base b, the "logarithm of x" is equal to the *exponent* to which the base b must be raised, in order to obtain the value x. For example, the statements in the following line are equivalent to those in the line below it, respectively:

$$10^2 = 100,\quad (37.4)^1 = 37.4,\quad (9)^{1/2} = \sqrt{9} = 3,\quad (98.6)^0 = 1,\quad 5^{-1} = \frac{1}{5^1} = 0.2,\quad 2^{-6} = \frac{1}{2^6} = 0.015625$$

$$\log_{10}(100) = 2,\quad \log_{37.4}(37.4) = 1,\quad \log_9(3) = 1/2,\quad \log_{98.6}(1) = 0,\quad \log_5(0.2) = -1,\quad \log_2(0.015625) = -6.$$

- The choice of base $b = 10$ results in the so-called **common logarithm** "$\log_{10}(x)$." But for calculus applications, the preferred base is "Euler's constant" $e = 2.71828\ldots$, resulting in the **natural logarithm** "$\log_e(x)$," usually denoted "**ln(x)**." And because it is the inverse of the exponential function e^x, it follows that $e^{\ln(x)} = x$ and $\ln(e^x) = x$. (For example, $\ln(2) = 0.6931$ because $e^{0.6931} = 2$.) The graphs of $y = e^x$ and $y = \ln(x)$ show the typical symmetry with respect to the line $y = x$ of inverse functions; see figure. Logarithms satisfy many properties that make them extremely useful for manipulating complex algebraic expressions, among other things…



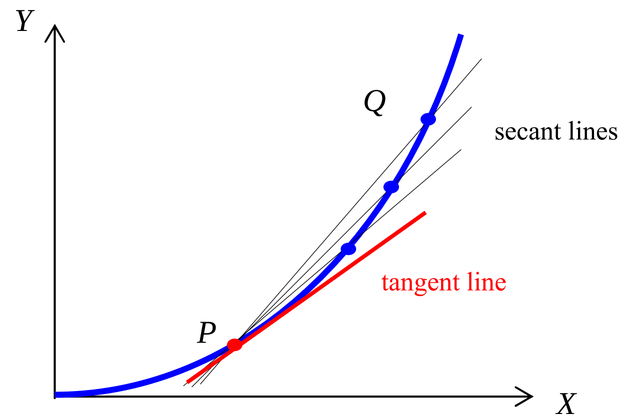
4. Limits and Derivatives

Newton, Leibniz ~ 1680

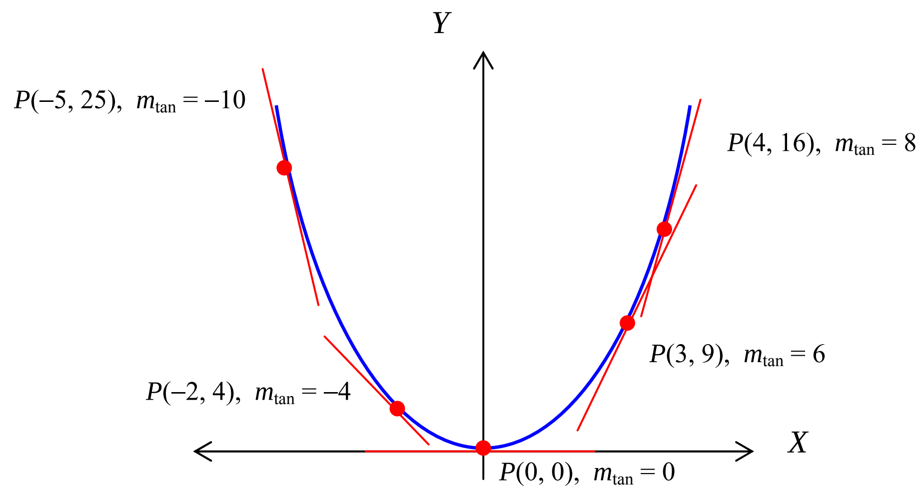
We saw above that as the values of x grow ever *larger*, the values of $\frac{1}{x}$ become ever *smaller*. We can't actually reach 0 exactly, but we can "sneak up" on it, forcing $\frac{1}{x}$ to become *as close to 0 as we like*, simply by making x large enough. (For instance, we can force $1/x < 10^{-500}$ by making $x > 10^{500}$.) In this context, we say that 0 is a **limiting value** of the $\frac{1}{x}$ values, as x gets arbitrarily large. A mathematically concise way to express this is a "limit statement":

$$\lim_{x \to +\infty} \frac{1}{x} = 0.$$

Many other limits are possible, but we now wish to consider a special kind. To motivate this, consider again the parabola example $y = f(x) = x^2$. The **average rate of change** between the two points $P(3, 9)$ and $Q(4, 16)$ on the graph can be calculated as the slope of the **secant line** connecting them, via the previous formula: $m_{\text{sec}} = \frac{\Delta y}{\Delta x} = \frac{16-9}{4-3} = 7$. Now suppose that we slide to a new point $Q(3.5, 12.25)$ *on the graph*, closer to $P(3, 9)$. The average rate of change is now $m_{\text{sec}} = \frac{\Delta y}{\Delta x} = \frac{12.25-9}{3.5-3} = 6.5$, the slope of the new secant line between P and Q. If we now slide to a new point $Q(3.1, 9.61)$ still closer to $P(3, 9)$, then the new slope is $m_{\text{sec}} = \frac{\Delta y}{\Delta x} = \frac{9.61-9}{3.1-3} = 6.1$, and so on. As Q approaches P, the slopes m_{sec} of the secant lines *appear* to get ever closer to 6 – the slope m_{tan} of the **tangent line** to the curve $y = x^2$ at the point $P(3, 9)$ – thus measuring the **instantaneous rate of change** of this function *at this point* $P(3, 9)$. (The same thing also happens if we approach P from the left side.) We can actually verify this by an explicit computation: From fixed point $P(3, 9)$ to any nearby point $Q(3 + \Delta x, (3 + \Delta x)^2)$ on the graph of $y = x^2$, we have $m_{\text{sec}} = \frac{\Delta y}{\Delta x} = \frac{(3+\Delta x)^2 - 9}{\Delta x} = \frac{9 + 6\,\Delta x + (\Delta x)^2 - 9}{\Delta x} = \frac{6\,\Delta x + (\Delta x)^2}{\Delta x} = \frac{\Delta x \cdot (6 + \Delta x)}{\Delta x} = 6 + \Delta x$. (We can check this formula against the m_{sec} values that we already computed: if $\Delta x = 1$, then $m_{\text{sec}} = 7$ ✓ ; if $\Delta x = 0.5$, then $m_{\text{sec}} = 6.5$ ✓; if $\Delta x = 0.1$, then $m_{\text{sec}} = 6.1$ ✓.) As *Q approaches P* – i.e., as Δx *approaches* 0 – this quantity $m_{\text{sec}} = 6 + \Delta x$ *approaches* the quantity $m_{\text{tan}} = 6$ as its limiting value, confirming what we initially suspected.



Suppose now we wish to find the instantaneous rate of change of $y = f(x) = x^2$ at some *other* point P on the graph, say at $P(4, 16)$ or $P(-5, 25)$ or even $P(0, 0)$. We can use the same calculation as we did above: the <u>average</u> rate of change of $y = x^2$ <u>between</u> *any* two generic points $P(x, x^2)$ and $Q(x + \Delta x, (x + \Delta x)^2)$ on its graph, is given by $m_{\text{sec}} = \dfrac{\Delta y}{\Delta x} = \dfrac{(x+\Delta x)^2 - x^2}{\Delta x} = \dfrac{x^2 + 2x\,\Delta x + (\Delta x)^2 - x^2}{\Delta x}$ $= \dfrac{2x\,\Delta x + (\Delta x)^2}{\Delta x} = \dfrac{\Delta x \cdot (2x + \Delta x)}{\Delta x} = 2x + \Delta x$. As Q *approaches* P – i.e., as Δx *approaches* 0 – this quantity *approaches* $m_{\tan} = 2x$ "in the limit," thereby *defining* the <u>instantaneous</u> rate of change of the function <u>at</u> the point P. (Note that if $x = 3$, these calculations agree with those previously done for m_{sec} and $m_{\tan}$.) Thus, for example, the instantaneous rate of change of the function $y = f(x) = x^2$ at the point $P(4, 16)$ is equal to $m_{\tan} = 8$, at $P(-5, 25)$ is $m_{\tan} = -10$, and at the origin $P(0, 0)$ is $m_{\tan} = 0$.



In principle, there is nothing that prevents us from applying these same ideas to other functions $y = f(x)$. To find the instantaneous rate of change at an arbitrary point P on its graph, we first calculate the **average rate of change** between $P(x, f(x))$ and a nearby point $Q(x + \Delta x, f(x + \Delta x))$ on its graph, as measured by $m_{\text{sec}} = \dfrac{\Delta y}{\Delta x} = \dfrac{f(x+\Delta x) - f(x)}{\Delta x}$. As Q *approaches* P – i.e., as Δx *approaches* 0 from both sides – this quantity becomes the **instantaneous rate of change** *at* P, defined by:

$$m_{\tan} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

This object – denoted compactly by $\dfrac{dy}{dx}$ – is called the **derivative** of the function $y = f(x)$, and is also symbolized by several other interchangeable notations: $\dfrac{d\,f(x)}{dx}$, $\dfrac{d}{dx}[f(x)]$, $f'(x)$, etc. (The last is sometimes referred to as "prime notation.") The process of calculating the derivative of a function is called **differentiation**. Thus, the derivative of the function $y = f(x) = x^2$ is equal to the function $\dfrac{dy}{dx} = f'(x) = 2x$. This can also be written more succinctly as $\dfrac{d\,(x^2)}{dx} = 2x$, $\dfrac{d}{dx}(x^2) = 2x$, or $dy = 2x\,dx$. The last form expresses the so-called **differential** dy in terms of the differential dx, which can be used to <u>estimate</u> a small output difference Δy in terms of a small input difference Δx, i.e., $\Delta y \approx 2x\,\Delta x$.

- Using methods very similar to those above, it is possible to prove that such a general rule exists for *any* power function x^p, not just $p = 2$. Namely, if $y = f(x) = x^p$, then $\frac{dy}{dx} = f'(x) = p\,x^{p-1}$, i.e.,

Power Rule $$\frac{d}{dx}(x^p) = p\,x^{p-1}.$$

Examples: If $y = x^3$, then $\frac{dy}{dx} = 3\,x^2$. If $y = x^{1/2}$, then $\frac{dy}{dx} = \frac{1}{2}x^{-1/2}$. If $y = x^{-1}$, then $\frac{dy}{dx} = -\,x^{-2}$. Also note that if $y = x = x^1$, then $\frac{dy}{dx} = 1\,x^0 = 1$, *as it should!* (The line $y = x$ has slope $m = 1$ *everywhere*.)

- Again using the preceding "limit definition" of a derivative, it can be proved that if $y = f(x) = b^x$, then $\frac{dy}{dx} = f'(x) = b^x \ln(b)$, i.e., $\frac{d}{dx}(b^x) = b^x \ln(b)$. Equivalently, if $y = f(x) = e^{ax}$, then $\frac{dy}{dx} = f'(x) = a\,e^{ax}$, i.e.,

Exponential Rule $$\frac{d}{dx}(b^x) = b^x \ln(b) \quad \text{– or equivalently –} \quad \frac{d}{dx}(e^{ax}) = a\,e^{ax}.$$

Examples: If $y = 2^x$, then $\frac{dy}{dx} = 2^x \ln(2) = 2^x\,(0.6931)$. If $y = 10^x$, then $\frac{dy}{dx} = 10^x \ln(10) = 10^x\,(2.3026)$. Hence, for any positive base b, the derivative of the function b^x is equal to the product of b^x *times a constant factor of* $\ln(b)$ that "hangs along for the ride." However, if $b = e$ (or equivalently, $a = 1$), then the derivative of $y = e^x$ is simply $\frac{dy}{dx} = e^x\,(1) = e^x$. In other words, the derivative of e^x is equal to just e^x itself! *This explains why base e is a particularly desirable* ***special case****; the constant multiple is just* 1. More examples: If $y = e^{x/2}$, then $\frac{dy}{dx} = \frac{1}{2}e^{x/2}$. If $y = e^{-x}$, then $\frac{dy}{dx} = (-1)\,e^{-x} = -e^{-x}$.

- Finally, exploiting the fact that exponentials and logarithms are inverses functions, we have for $x > 0$,

Logarithm Rule $$\frac{d}{dx}[\log_b(x)] = \frac{1}{x}\,\frac{1}{\ln(b)} \quad \Rightarrow \quad \textbf{Special case } (b = e)\textbf{:}\ \frac{d}{dx}[\ln(x)] = \frac{1}{x}.$$

- One last important example is worth making explicit: let $y = f(x) = 7$, whose graph is a horizontal line. The average rate of change (i.e., slope) between any two points on this graph is $\frac{\Delta y}{\Delta x} = \frac{7-7}{\Delta x} = 0$, and hence the instantaneous rate of change = 0 as well. In the same way, we have the following result.

If $y = f(x) = C$ (*any* constant) for all x, then the derivative $\frac{dy}{dx} = f'(x) = 0$ for all x.

(However, observe that a *vertical* line, having equation $x = C$, has an infinite – or **undefined** – slope.)

Not every function has a derivative everywhere! For example, the functions $y = f(x) = |x|$, $x^{1/3}$, and x^{-1} are *not* **differentiable** at $x = 0$, all for different reasons. Although the first two are **continuous** through the origin (0, 0), the first has a V-shaped graph; a *uniquely* defined tangent line does not exist at the "corner." The second graph has a *vertical* tangent line there, hence the slope is infinite. And as we've seen, the last function is *undefined* at the origin; $x = 0$ is not even in its domain, so any talk of a tangent line there is completely meaningless. But, many complex functions *are* indeed differentiable…

Properties of Derivatives

1. For any constant c, and any **differentiable** function $f(x)$,

$$\frac{d}{dx}[c\,f(x)] = c\,\frac{df}{dx}$$

$$[c\,f(x)]' = c\,f'(x)$$

Example: If $y = 5\,x^3$, then $\frac{dy}{dx} = 5\,(3\,x^2) = 15\,x^2$.

Example: If $y = \frac{1}{9}x$, then $\frac{dy}{dx} = \frac{1}{9}(1) = \frac{1}{9}$.

Example: If $y = -3\,e^{2x}$, then $\frac{dy}{dx} = -3\,(2\,e^{2x}) = -6\,e^{2x}$.

For any *two* differentiable functions $f(x)$ and $g(x)$,

2. Sum and Difference Rules

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

$$[\,f(x) \pm g(x)\,]' = f'(x) \pm g'(x)$$

Example: If $y = x^{3/2} - 7\,x^4 + 10\,e^{-3x} - 5$, then $\frac{dy}{dx} = \frac{3}{2}x^{1/2} - 28\,x^3 - 30\,e^{-3x}$.

3. Product Rule

$$[\,f(x)\,g(x)\,]' = f'(x)\,g(x) + f(x)\,g'(x)$$

Example: If $y = x^{11}\,e^{6x}$, then $\frac{dy}{dx} = (11\,x^{10})(e^{6x}) + (x^{11})(6\,e^{6x}) = (11 + 6x)\,x^{10}\,e^{6x}$.

4. Quotient Rule

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{[g(x)]^2}$$ provided $g(x) \neq 0$

Example: If $y = \frac{e^{4x}}{x^7 + 8}$, then $\frac{dy}{dx} = \frac{(x^7 + 8)\,4e^{4x} - (7x^6)\,e^{4x}}{(x^7 + 8)^2} = \frac{(4x^7 - 7x^6 + 32)\,e^{4x}}{(x^7 + 8)^2}$.

5. Chain Rule **NOTE:** See below for a more detailed explanation.

$$[\,f(\,g(x)\,)\,]' = f'(\,g(x)\,) \times g'(x)$$

Example: If $y = (x^{2/3} + 2e^{-9x})^6$, then $\frac{dy}{dx} = 6\,(x^{2/3} + 2e^{-9x})^5\left(\frac{2}{3}x^{-1/3} - 18e^{-9x}\right)$.

Example: If $y = e^{-x^2/2}$, then $\frac{dy}{dx} = e^{-x^2/2}\left(-\frac{2\,x}{2}\right) = -x\,e^{-x^2/2}$.

The graph of this function is related to the "bell curve" of probability and statistics. Note that you *cannot* calculate its derivative by the "exponential rule" given above, because the exponent is a function of x, not just x itself (or a constant multiple ax)!

Example: If $y = \ln(7x^{10} + 8x^6 - 4x + 11)$, then $\frac{dy}{dx} = \frac{1}{7x^{10} + 8x^6 - 4x + 11}(70x^9 + 48x^5 - 4)$.

The **Chain Rule**, which can be written several different ways, bears some further explanation. It is a rule for differentiating a **composition** of two functions f and g, that is, a function *of* a function $y = f\,(\,g(x)\,)$. The function in the first example above can be viewed as composing the "outer" function $f(u) = u^6$, with the "inner" function $u = g(x) = x^{2/3} + 2e^{-9x}$. To find its derivative, first take the derivative of the outer function ($6u^5$, by the Power Rule given above), then multiply that by the derivative of the inner, and we get our answer. Similarly, the function in the second example can be viewed as composing the outer exponential function $f(u) = e^u$ (whose derivative, recall, is itself), with the inner power function $u = g(x) = -x^2/2$. And the last case can be seen as composing the outer logarithm function $f(u) = \ln(u)$ with the inner polynomial function $u = g(x) = 7x^{10} + 8x^6 - 4x + 11$.

Hence, if $y = f(u)$, and $u = g(x)$, then $\dfrac{dy}{dx} = f\,'(u)\,g\,'(x)$ is another way to express this procedure.

The logic behind the Chain Rule is actually quite simple and intuitive (though a formal proof involves certain technicalities that we do not pursue here). Imagine three cars traveling at different rates of speed over a given time interval: A travels at a rate of 60 mph, B travels at a rate of 40 mph, and C travels at a rate of 20 mph. In addition, suppose that A knows how fast B is traveling, and B knows how fast C is traveling, but A does *not* know how fast C is traveling. Over this time span, the average rate of change of A, *relative to* B, is equal to the ratio of their respective distances traveled, i.e., $\dfrac{\Delta A}{\Delta B} = \dfrac{60}{40} = \dfrac{3}{2}$, three-halves as much. Similarly, the average rate of change of B, *relative to* C, is equal to the ratio $\dfrac{\Delta B}{\Delta C} = \dfrac{40}{20} = \dfrac{2}{1}$, twice as much. Therefore, the average rate of change of A, *relative to* C, can be obtained by multiplying together these two quantities, via the elementary algebraic identity $\dfrac{\Delta A}{\Delta C} = \dfrac{\Delta A}{\Delta B} \times \dfrac{\Delta B}{\Delta C}$, or $\dfrac{3}{2} \times \dfrac{2}{1} = \dfrac{3}{1}$, three times as much. (Of course, if A *has* direct knowledge of C, this would not be necessary, for $\dfrac{\Delta A}{\Delta C} = \dfrac{60}{20} = \dfrac{3}{1}$. As it is, B acts an intermediate link in the chain, or an **auxiliary function**, which makes calculations easier in some contexts.) The idea behind the Chain Rule is that what is true for average rates of change also holds for instantaneous rates of change, as the time interval shrinks to 0 "in the limit," i.e., derivatives.

With this insight, an alternate – perhaps more illuminating – equivalent way to write the Chain Rule is $\dfrac{dy}{dx} = \dfrac{dy}{du} \cdot \dfrac{du}{dx}$ which, as the last three examples illustrate, specialize to a **General Power Rule**, **General Exponential Rule**, and **General Logarithm Rule** for differentiation, respectively:

If $y = u^p$, then $\dfrac{dy}{dx} = p\,u^{p-1}\,\dfrac{du}{dx}$. That is, $d(u^p) = p\,u^{p-1}\,du$.

If $y = e^u$, then $\dfrac{dy}{dx} = e^u\,\dfrac{du}{dx}$. That is, $d(e^u) = e^u\,du$.

If $y = \ln(u)$, then $\dfrac{dy}{dx} = \dfrac{1}{u}\,\dfrac{du}{dx}$. That is, $d[\ln(u)] = \dfrac{1}{u}\,du$.

Exercise: Recall from page 5 that $y = x^x$ is not a power function. Obtain its derivative via the technique of **implicit** (in particular, **logarithmic**) **differentiation**. (Not covered here.)

5. Applications: Estimation, Roots and Maxima & Minima, Related Rates

- As seen, for nearby points P and Q on the graph of a function $f(x)$, it follows that $m_{sec} \approx m_{tan}$, at least informally. That is, for a small change Δx, we have $\frac{\Delta y}{\Delta x} \approx \frac{dy}{dx} = f'(x)$, or $\boxed{\Delta y \approx f'(x)\,\Delta x}$. Hence the **first derivative** $f'(x)$ can be used in a crude local estimate of the amount of change Δy of the function $y = f(x)$.

  Example: Let $y = f(x) = (x^2 - 2x + 2)\,e^x$. The change in function value from say, $f(1)$ to $f(1.03)$, can be estimated by $\Delta y \approx f'(x)\,\Delta x = (x^2 e^x)\,\Delta x$, when $x = 1$ and $\Delta x = 0.03$, i.e., $\Delta y \approx 0.03\,e = 0.0815$, to four decimal places.

  Exercise: Compare this with the exact value of $\Delta y = f(1.03) - f(1)$.

  NOTE: A better estimate of the difference $\Delta y = f(x + \Delta x) - f(x)$ can be obtained by adding information from the **second derivative** $f''(x)$ (= derivative of $f'(x)$), the **third derivative** $f'''(x)$, etc., to the previous formula:

  $$\Delta y \approx \overbrace{\frac{1}{1!} f'(x)\,\Delta x}+ \frac{1}{2!} f''(x)\,(\Delta x)^2 + \frac{1}{3!} f'''(x)\,(\Delta x)^3 + \ldots + \frac{1}{n!} f^{(n)}(x)\,(\Delta x)^n$$

  (where $1! = 1$, $2! = 2 \times 1 = 2$, $3! = 3 \times 2 \times 1 = 6$, and in general "$n$ factorial" is defined as $n! = n \times (n-1) \times (n-2) \times \ldots \times 1$). The formal mathematical statement of equality between Δy and the sum of higher derivatives of f is an extremely useful result known as **Taylor's Formula**.

- Suppose we wish to solve for the **roots** of the equation $f(x) = 0$, i.e., the values where the graph of $f(x)$ intersects the X-axis (also called the **zeros** of the function $f(x)$). Algebraically, this can be extremely tedious or even impossible, so we often turn to **numerical techniques** which yield computer-generated approximations. In the popular **Newton-Raphson Method**, we start with an initial guess x_0, then produce a sequence of values $x_0, x_1, x_2, x_3, \ldots$ that converges to a numerical solution, by iterating the expression $\boxed{x - \frac{f(x)}{f'(x)}}$. (**Iteration** is the simple process of repeatedly applying the same formula to itself, as in a continual "feedback loop.")

  Example: Solve for x: $f(x) = x^3 - 21x^2 + 135x - 220 = 0$.

  We note that $f(2) = -26 < 0$, and $f(3) = +23 > 0$. Hence, *by continuity*, this **cubic** (degree 3) **polynomial** must have a zero somewhere between 2 and 3. Applying Newton's Method, we iterate the formula $x - \frac{f(x)}{f'(x)} = x - \frac{x^3 - 21x^2 + 135x - 220}{3x^2 - 42x + 135} = \frac{2x^3 - 21x^2 + 220}{3x^2 - 42x + 135}$. Starting with $x_0 = 2$, we generate the sequence $x_1 = 2.412698$, $x_2 = 2.461290$, $x_3 = 2.461941$, $x_4 = 2.461941$; thereafter, the iterates remain fixed at this value 2.461941. (Check: $f(2.461941) = 1.4 \times 10^{-5}$; the small difference from 0 is due to roundoff error.)

  Exercise: Try different initial values, e.g., $x_0 = 0, 1, 3, 4, 5, 6, 7$. Explain the different behaviors.

Notice the extremely rapid convergence to the root. In fact, it can be shown that the small error (between each value and the true solution) in each iteration is approximately *squared* in the next iteration, resulting in a much smaller error. This feature of **quadratic convergence** is a main reason why this is a favorite method. Why does it work at all? Suppose P_0 is a point on the graph of $f(x)$, whose x-coordinate x_0 is reasonably close to a root. Generally speaking, the tangent line at P_0 will then intersect the X-axis at a value much closer to the root. This value x_1 can then be used as the x-coordinate of a new point P_1 on the graph, and the cycle repeated until some predetermined error tolerance is reached. Algebraically formalizing this process results in the general formula given above.

- If a function $f(x)$ has either a **relative maximum** (i.e., **local maximum**) or a **relative minimum** (i.e., **local minimum**) at some value of x, and if $f(x)$ is *differentiable* there, then its tangent line must be horizontal, i.e., slope $m_{\tan} = f'(x) = 0$. This suggests that, in order to find such **relative extrema** (i.e., **local extrema**), we set the derivative $f'(x)$ equal to zero, and solve the resulting algebraic equation for the **critical values** of f, perhaps using a numerical approximation technique like Newton's Method described above. (*But beware*: *Not all critical values necessarily correspond to relative extrema!* More on this later…)

Example (cont'd): Find and classify the critical points of $y = f(x) = x^3 - 21x^2 + 135x - 220$.

We have $f'(x) = 3x^2 - 42x + 135 = 0$. As this is a **quadratic** (degree 2) **polynomial** equation, a numerical approximation technique is not necessary. We can use the **quadratic formula** to solve this explicitly, or simply observe that, via factoring, $3x^2 - 42x + 135 = 3(x-5)(x-9) = 0$. Hence there are *two* **critical values**, $x = 5$ and $x = 9$. Furthermore, since $f(5) = 55$ and $f(9) = 23$, it follows that the corresponding **critical points** on the graph of $f(x)$ are (5, 55) and (9, 23).

Once obtained, it is necessary to determine the exact nature of these critical points. Consider the first critical value, $x = 5$, where $f' = 0$. Let us now evaluate the derivative $f'(x)$ at two nearby values that bracket $x = 5$ on the left and right, say $x = 4$ and $x = 6$. We calculate that:

$m_{\tan}(4) = f'(4) = +15 > 0$, which indicates that the original function f is **increasing** at $x = 4$,

$m_{\tan}(5) = f'(5) = 0$, which indicates that f is **neither increasing nor decreasing** at $x = 5$,

$m_{\tan}(6) = f'(6) = -9 < 0$, which indicates that the original function f is **decreasing** at $x = 6$.

Hence, as we move from left to right in a **local neighborhood** of $x = 5$, the function $f(x)$ rises, levels off, then falls. This indicates that the point (5, 55) is a **relative maximum** for f, and demonstrates an application of the "**First Derivative Test**" for determining the nature of critical points. In an alternate method, the "**Second Derivative Test**," we evaluate $f''(x) = 6x - 42$ at the critical value $x = 5$, i.e.,

$f''(5) = -12 < 0$, which indicates that the original function f is **concave down** ("spills") at this value.

Hence, this also shows that the point (5, 55) is a **relative maximum** for f, consistent with the above.

Exercise: Show that: (1) $f'(8) < 0$, $f'(10) > 0$ **First Derivative Test**

(2) $f''(9) > 0$. **Second Derivative Test**

In either case, conclude that the point (9, 23) is a **relative minimum** for f.

Finally, notice that $f''(x) = 6x - 42 = 0$ when $x = 7$, and in a **local neighborhood** of that value,

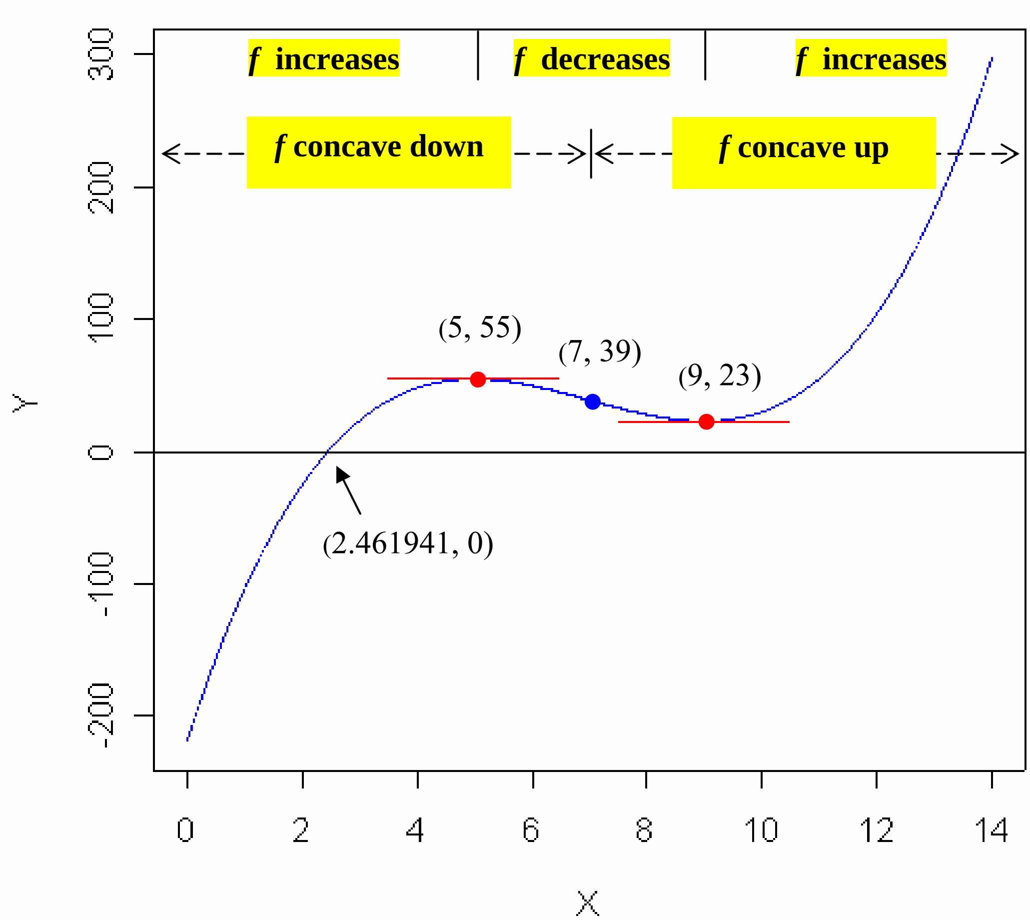
$f''(6) = -6 < 0$, which indicates that the original function f is **concave down** ("spills") at $x = 6$,

$f''(7) = 0$, which indicates that f is **neither concave down nor concave up** at $x = 7$,

$f''(8) = +6 > 0$, which indicates that the original function f is **concave up** ("holds") at $x = 8$.

Hence, across $x = 7$, there is change in **concavity** of the function $f(x)$. This indicates that (7, 39) is a **point of inflection** for f.

The full graph of $f(x) = x^3 - 21x^2 + 135x - 220$ is shown below, using all of this information.



As we have shown, this function has a **relative maximum** (i.e., **local maximum**) value = 55 at $x = 5$, and a **relative minimum** (i.e., **local minimum**) value = 23 at $x = 7$. But clearly, there are both higher and lower points on the graph! For example, if $x \geq 11$, then $f(x) \geq 55$; likewise, if $x \leq 3$, then $f(x) \leq 23$. (Why?) Therefore, this function has *no* **absolute maximum** (i.e., **global maximum**) value, and *no* **absolute minimum** (i.e., **global minimum**) value. However, if we restrict the domain to an interval that is "closed and bounded" (i.e., **compact**), then both absolute extrema are attained. For instance, in the interval [4, 10], the relative extreme points are also the absolute extreme points, i.e. the function attains its global maximum and minimum values of 55 (at $x = 5$) and 23 (at $x = 9$), respectively. However, in the interval [4, 12], the global maximum of the function is equal to 104, attained at the *right endpoint* $x = 12$. Similarly, in the interval [0, 12], the global minimum of the function is equal to –220, attained at the *left endpoint* $x = 0$.

Exercise: Graph each of the following.

$f(x) = x^3 - 21x^2 + 135x - 243$

$f(x) = x^3 - 21x^2 + 135x - 265$

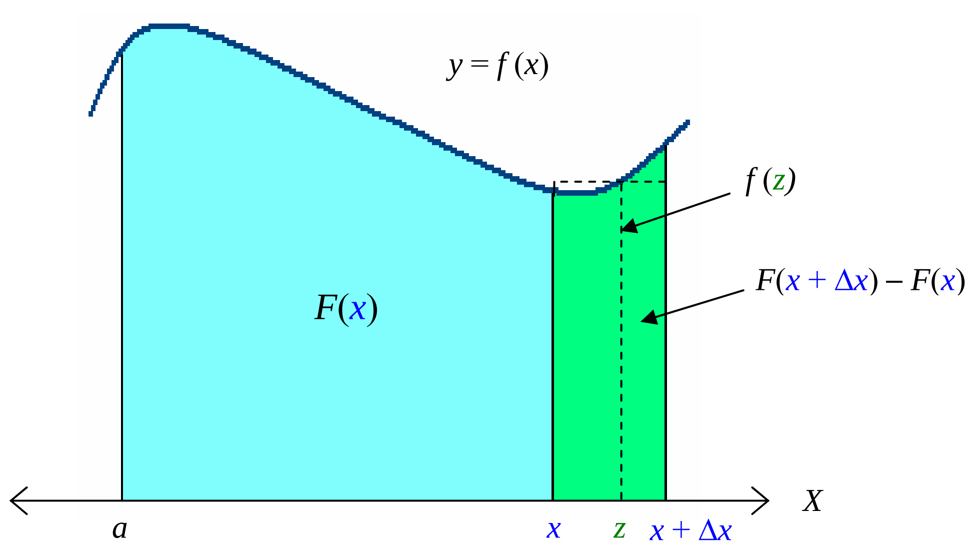
$f(x) = x^3 - 21x^2 + 147x - 265$

Exercise: The origin (0, 0) is a **critical point** for both $f(x) = x^3$ and $f(x) = x^4$. (Why?) Using the tests above, formally show that it is a **relative minimum** of the latter, but a **point of inflection** of the former. Graph both of these functions.

- The volume V of a spherical cell is functionally related to its radius r via the formula $V = \frac{4}{3}\pi r^3$. Hence, the derivative $\frac{dV}{dr} = 4\pi r^2$ naturally expresses the instantaneous rate of change of V with respect to r. But as V and r are functions of time, we can relate *both* their rates directly, by differentiating the original relation with respect to the time variable t, yielding the equation $\frac{dV}{dt} = 4\pi r^2 \frac{dr}{dt}$. Observant readers will note that the two equations are mathematically equivalent via the Chain Rule, since $\frac{dV}{dt} = \frac{dV}{dr}\frac{dr}{dt}$, but the second form illustrates a simple application of a general technique called **related rates**. That is, the "time derivative" of a functional relation between two or more variables results in a type of **differential equation** that relates their rates.

6. Integrals and Antiderivatives

Suppose we have a general function $y = f(x)$; for simplicity, let's assume that the function is non-negative (i.e., $f(x) \geq 0$) and **continuous** (i.e., informally, it has no breaks or jumps), as shown below. We wish to find the area under the graph of f, in an interval $[a, x]$, from some *fixed* lower value a, to any *variable* upper value x.



We formally define a new function

$$F(x) = \text{Area under the graph of } f \text{ in the interval } [a, x].$$

Clearly, because every value of x results in *one and only one* area (shown highlighted above in blue), *this is a function of x, by definition!* Moreover, F must also have a strong connection with f itself. To see what that connection must be, consider a nearby value $x + \Delta x$. Then,

$$F(x + \Delta x) = \text{Area under the graph of } f \text{ in the interval } [a, x + \Delta x],$$

and take the difference of these two areas (highlighted above in green):

$$F(x + \Delta x) - F(x) = \text{Area under the graph of } f \text{ in the interval } [x, x + \Delta x]$$

$$= \text{Area of the rectangle with height } f(z) \text{ and width } \Delta x$$
(where z is *some* value in the interval $[x, x + \Delta x]$)

$$= f(z)\,\Delta x.$$

Therefore, we have

$$\frac{F(x + \Delta x) - F(x)}{\Delta x} = f(z).$$

Now take the limit of both sides as $\Delta x \to 0$. We see that the left hand side becomes the *derivative of* $F(x)$ (recall the definition of the **derivative** of a function, previously given) and, noting that $z \to x$ as $\Delta x \to 0$, we see that the right hand side $f(z)$ becomes (via continuity) $f(x)$. Hence,

$$\boxed{F'(x) = f(x)},$$

i.e., F is an **antiderivative** of f.

Therefore, we formally express… $\boxed{F(x) = \int_a^x f(t)\,dt}$,

where the right-hand side $\int_a^x f(t)\,dt$ represents the **definite integral** of the function f from a to x. (In this context, f is called the **integrand**.) More generally, if F is *any* antiderivative of f, then the two functions are related via the **indefinite integral**: $\boxed{\int f(x)\,dx = F(x) + C}$, where C is an arbitrary constant.

<u>Example 1</u>: $F(x) = \frac{1}{10}x^{10} + C$ (where C is *any* constant) is the general antiderivative of $f(x) = x^9$, because $F'(x) = \frac{1}{10}(10x^9) + 0 = x^9 = f(x)$. ✓

We can write this relation succinctly as $\int x^9\,dx = \frac{1}{10}x^{10} + C$.

Example 2: $F(x) = 8\, e^{x/8} + C$ (where C is *any* constant) is the general antiderivative of $f(x) = e^{x/8}$, because $F'(x) = 8\left(\frac{1}{8} e^{x/8}\right) + 0 = e^{x/8} = f(x)$. ✓

We can write this relation succinctly as $\int e^{x/8}\, dx = 8\, e^{x/8} + C$.

NOTE: Integrals possess the analogues of Properties 1 and 2 for derivatives, found on page 10. *In particular, the integral of a constant multiple of a function, c f(x), is equal to that constant multiple c, times the integral of the function f(x).* Also, the integral of a sum (respectively, difference) of two functions is equal to the sum (respectively, difference) of the integrals. (The integral analogue for *products* corresponds to a technique known as **integration by parts**; not reviewed here.) These are extremely important properties for the applications that follow.

Properties of Integrals

1. For any constant c, and any **integrable** function $f(x)$,

$$\int [c\, f(x)]\, dx = c \int f(x)\, dx$$

For any *two* integrable functions $f(x)$ and $g(x)$,

2. **Sum and Difference Rules**

$$\int [f(x) \pm g(x)]\, dx = \int f(x)\, dx \pm \int g(x)\, dx$$

From the previous two examples, it is evident that the differentiation rules for power and exponential functions can be inverted (essentially by taking the integral of both sides) to the **General Power Rule**, **General Logarithm Rule**, and **General Exponential Rule** for integration:

$$\int u^p\, du = \begin{cases} \dfrac{u^{p+1}}{p+1} + C, & \text{if } p \neq -1 \\ \ln|u| + C, & \text{if } p = -1 \end{cases}$$

$$\int e^u\, du = e^u + C$$

NOTE: In order to use these formulas correctly, du must be present in the integrand (up to a *constant multiple*). To illustrate…

Example 3: $\int (x^5+2)^9\ 5x^4\,dx = \dfrac{(x^5+2)^{10}}{10} + C.$

$$\underbrace{\int (x^5+2)^9}_{\int u^9}\ \underbrace{5x^4\,dx}_{du} = \underbrace{\frac{(x^5+2)^{10}}{10} + C}_{\frac{u^{10}}{10} + C}$$

There are two ways to solve this problem. The first is to expand out the algebraic expression in the integrand, and integrate the resulting polynomial (of degree 49) term-by-term… *Yuk*. The second way, as illustrated, is to recognize that if we substitute $u = x^5 + 2$, then $du = 5x^4\,dx$, which is precisely the other factor in the integrand, *as is*! Therefore, in terms of the variable u, this is essentially just a "power rule" integration, carried out above. (To check the answer, take the derivative of the right-hand function, and verify that the original integrand is restored. *Don't forget to use the Chain Rule!*) Note that if the *constant multiple* 5 were absent from the original integrand, we could introduce and compensate for it. (This procedure is demonstrated in the next example.) However, if the x^4 were absent, *or were replaced by any other function*, then we would not be able to carry out the integration in the manner shown, since via Property 1 above, we can only balance *constant multiples*, not functions!

Example 4: $\displaystyle\int \frac{x^2}{\sqrt{1+x^3}}\,dx = \frac{1}{3}\int (1+x^3)^{-1/2}\ 3x^2\,dx = \frac{1}{3}\times 2\,(1+x^3)^{1/2} + C = \frac{2}{3}\sqrt{1+x^3} + C.$

$$\int \underbrace{(1+x^3)^{-1/2}}_{u^{-1/2}}\ \underbrace{3x^2\,dx}_{du} = \underbrace{2\,(1+x^3)^{1/2} + C}_{\frac{u^{+1/2}}{1/2} + C}$$

In this example, let $u = 1 + x^3$, so that $du = 3x^2\,dx$. This is present in the original integrand, *except* for the constant multiple of 3 (which we can introduce, provided we preserve the balance via multiplication by 1/3 on the outside of the integral sign), revealing that this is again a "power rule" integral. And again, if the x^2 were missing from the integrand, *or were replaced by any other function*, then we would not have been able to carry out the integration exactly in the manner shown. Verify that the answer is correct via differentiation.

Example 5: $\displaystyle\int \frac{x^2}{1+x^3}\,dx = \frac{1}{3}\int (1+x^3)^{-1}\ 3x^2\,dx = \frac{1}{3}\ \ln|1+x^3| + C.$

$$\int \underbrace{(1+x^3)^{-1}}_{u^{-1}}\ \underbrace{3x^2\,dx}_{du} = \underbrace{\ln|1+x^3| + C}_{\ln|u| + C}$$

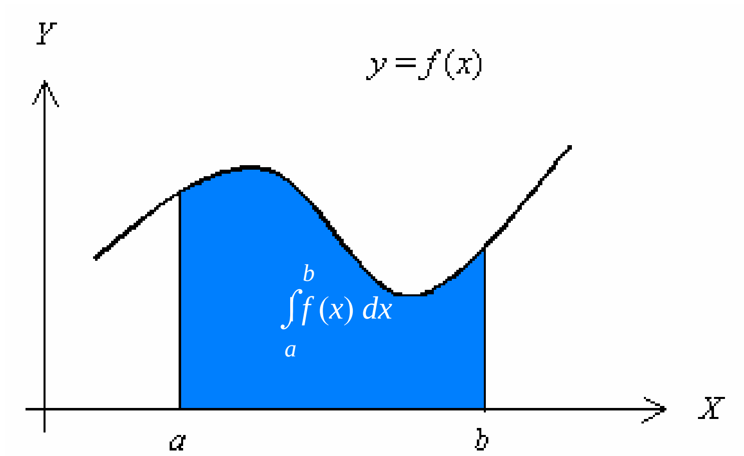
This is very similar to the previous example, except that it is a logarithmic, not a power, integral.

Example 6: $\int z\,e^{-z^2/2}\,dz = -\int e^{-z^2/2}\,(-z)\,dz = -\,e^{-z^2/2} + C.$

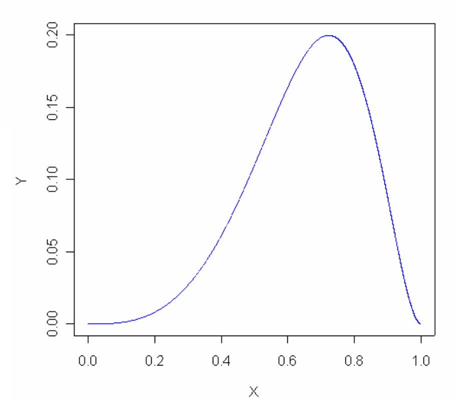
$$\int \underbrace{e^{-z^2/2}}_{e^u}\ \underbrace{(-z)\,dz}_{du} = \underbrace{e^{-z^2/2} + C}_{e^u + C}$$

In this example, if $u = -z^2/2$, then $du = -z\,dz$. This is present in the integrand, *except* for the constant multiple -1, which we can easily balance, and perform the subsequent exponential integration. Again, if the z were missing from the integrand, we would not be able to introduce and balance for it. In fact, it can be shown that without this factor of z, this integral is not expressible in terms of "elementary functions." Because it is related to the important "bell curve" of probability and statistics, the values of its corresponding *definite* integral are estimated and tabulated for general use.

Finally, all these results can be summarized into one elegant statement, the **Fundamental Theorem of Calculus** for **definite integrals**: $\boxed{\int_a^b f(x)\,dx = F(b) - F(a)}$. (Advanced techniques of integration – such as **integration by parts**, **trigonometric substitution**, **partial fractions**, etc. – will not be reviewed here.)



Example 7: $\int_0^1 x^3\,(1 - x^4)^2\,dx$



This definite integral represents the amount of area under the curve $f(x) = x^3\,(1 - x^4)^2$, from $x = 0$ to $x = 1$.

Method 1. Expand and integrate term-wise: $\int_0^1 x^3\,(1 - 2x^4 + x^8)\,dx \;=\; \int_0^1 (x^3 - 2x^7 + x^{11})\,dx$

$$= \left[\frac{x^4}{4} - \frac{2x^8}{8} + \frac{x^{12}}{12}\right]_0^1 = \left[\frac{1^4}{4} - \frac{2\,(1)^8}{8} + \frac{1^{12}}{12}\right] - \left[\frac{0^4}{4} - \frac{2\,(0)^8}{8} + \frac{0^{12}}{12}\right] = \frac{1}{12} - 0 = \frac{1}{12}.$$

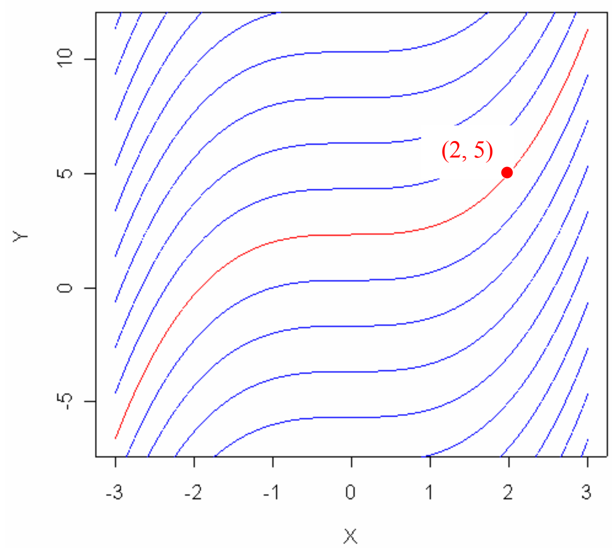
Method 2. Use the power function formula (if possible): If $u = 1 - x^4$, then $du = -4x^3\,dx$, and x^3 is indeed present in the integrand. Recall that the x-limits of integration should also be converted to u-limits: when $x = 0$, we get $u = 1 - 0^4 = 1$; when $x = 1$, we get $u = 1 - 1^4 = 0$.

$$-\frac{1}{4}\int_{x=0}^{x=1} \underbrace{(1 - x^4)^2}_{u^2}\,\underbrace{(-4x^3)\,dx}_{du} = -\frac{1}{4}\int_{u=1}^{u=0} u^2\,du = \frac{1}{4}\int_0^1 u^2\,du = \frac{1}{4}\left[\frac{u^3}{3}\right]_0^1 = \frac{1}{4}\left[\frac{1^3}{3} - \frac{0^3}{3}\right] = \frac{1}{12}.$$

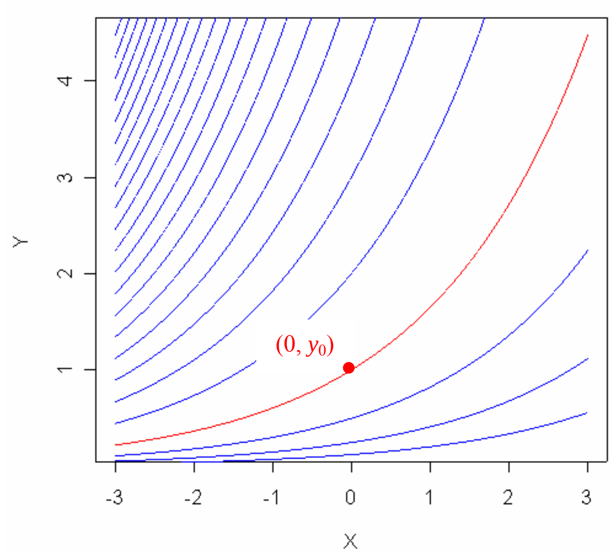
NOTE: **Numerical integration** techniques, such as the **Trapezoidal Rule**, are sometimes used also.

7. Differential Equations

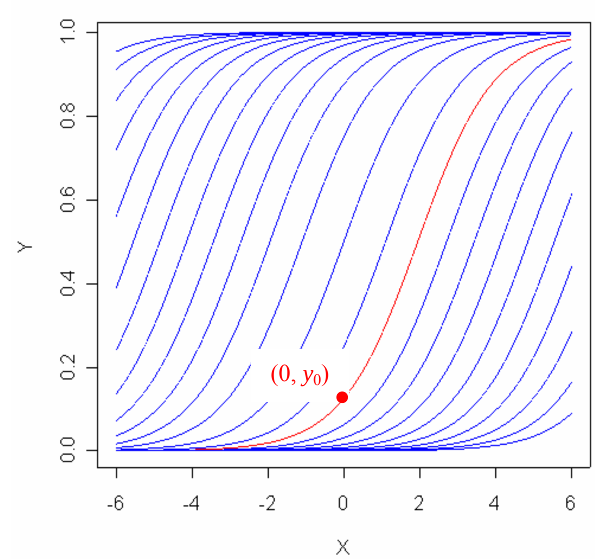
As a first example, suppose we wish to find a function $y = f(x)$ whose derivative $\frac{dy}{dx}$ is *given*, e;g., $\frac{dy}{dx} = x^2$. Formally, by **separation of variables** – y on the left, and x on the right – we can rewrite this **ordinary differential equation** (or **o.d.e.**) as $dy = x^2\, dx$. In this differential form, we can now integrate both sides explicitly to obtain $y = \frac{1}{3}x^3 + C$, where C is an arbitrary *additive* constant. Note that the "solution" therefore actually represents an entire *family* of functions; each function corresponds to a different value of C. Further specifying an **initial value** (or **initial condition**), such as $y(2) = 5$, singles out exactly one of them passing through the chosen point, in this case, $y = \frac{1}{3}x^3 + \frac{7}{3}$.



Now consider the case of finding a function $y = f(x)$ whose rate of change $\frac{dy}{dx}$ is proportional to y itself, i.e., $\frac{dy}{dx} = a\, y$, where a is a known **constant of proportionality** (either positive or negative). Separating variables produces $\frac{1}{y}\, dy = a\, dx$, integrating yields $\ln(y) = a\, x + C$, and solving gives $y = e^{ax + C} = e^{ax} e^{C}$, or $y = A\, e^{ax}$, where A is an arbitrary (positive) *multiplicative* constant. Hence this is a family of exponential curves, either increasing for $a > 0$ (as in unrestricted population growth, illustrated) or decreasing for $a < 0$ (as in radioactive isotope decay). Specifying an initial amount $y(0) = y_0$ "when the clock starts at time zero" yields the unique solution $y = y_0\, e^{ax}$.



Finally, suppose that population size $y = f(x)$ is restricted between 0 and 1, such that the rate of change $\frac{dy}{dx}$ is proportional to the product $y\,(1 - y)$, i.e., $\frac{dy}{dx} = a\, y\,(1 - y)$, where $a > 0$. With the initial condition $y(0) = y_0$, the solution is given by $y = \frac{y_0\, e^{ax}}{y_0\, e^{ax} + (1 - y_0)}$. This is known as the **logistic curve**, which initially resembles the exponential curve, but remains bounded as x gets large.



NOTE: Many types of differential equation exist, including those that cannot be explicitly solved using "elementary" techniques. Like integration above, such equations can be solved via **Euler's Method** and other, more sophisticated, numerical techniques.

8. Summary of Main Points

The instantaneous rate of change of a function $y = f(x)$ at a value of x in its domain, is given by its **derivative** $\frac{dy}{dx} = f'(x)$. This function is mathematically defined in terms of a particular "limiting value" of average rates of change over progressively smaller intervals (*when that limit exists*), and can be interpreted as the slope of the line tangent to the graph of $y = f(x)$. As complex functions are built up from simpler ones by taking sums, differences, products, quotients, and compositions, formulas exist for computing their derivatives as well. In particular, via the **Chain Rule** $\frac{d}{dx}[f(u)] = f'(u)\,\frac{du}{dx}$, we have:

$$\frac{d}{dx}(u^p) = p\,u^{p-1}\frac{du}{dx}$$ **General Power Rule**

$$\frac{d}{dx}(e^u) = e^u\frac{du}{dx}$$ **General Exponential Rule**

$$\frac{d}{dx}[\ln(u)] = \frac{1}{u}\frac{du}{dx}$$ **General Logarithm Rule**

Derivatives can be applied to **estimate** functions locally, find the **relative extrema** of functions, and **relate rates** of change of different but connected functions.

A function $f(x)$ has an **antiderivative** $F(x)$ if *its* derivative $\frac{dF}{dx} = f(x)$, or equivalently, $f(x)\,dx = dF$, and expressed in terms of an **indefinite integral**: $\int f(x)\,dx = F(x) + C$. In particular:

$$\int u^p\,du = \frac{u^{p+1}}{p+1} + C, \quad \text{if } p \neq -1$$ **General Power Rule**

$$\int u^{-1}\,du = \ln|u| + C$$ **General Logarithm Rule**

$$\int e^u\,du = e^u + C$$ **General Exponential Rule**

The corresponding **definite integral** $\int_a^b f(x)\,dx = F(b) - F(a)$ is commonly interpreted as the area under the graph of $y = f(x)$ in the interval $[a, b]$, though other interpretations do exist. Other quantities that can be interpreted as definite integrals include volume, surface area, arc length, amount of work done over a path, average value, probability, and many others.

Derivatives and integrals can be generally be used to analyze the dynamical behavior of complex systems, for example, via **differential equations** of various types. **Numerical methods** are often used when explicit solutions are intractable.

☞ Suggestions for improving this document? Send them to ifischer@wisc.edu.