**FINC 748: Investments** **Fall 2003**

**Lecture Note 0: Math Review**

---

The purpose of this note is to collect, in one place, most of the math needed to understand finance at the MBA level. This note is not intended to be overly technical, and is not chock-full of mathematical proofs. Rather, it is a chatty roadmap of facts and examples that will aid our understanding, without digressing too much within the body of the course. Ideally, you should read this note and understand it in full by the second week of class. You may later on find it helpful to use this note as a reference, as we encounter a particular mathematical concept, formula or calculation in our course.

---

- **THE BASICS**

  Sets, Functions and Graphs

  A **set** is a collection of **elements**. These elements may be numbers, names or anything at all.

  Example 1: We might define $X$ as the set of all first year MBA students at the Freeman School today. In mathematical notation, this set is represented as $X$={name1, name2,…., name 130}, where the 130 names are those of you and your classmates. Frequently, sets are denoted in text books as follows:

  $$X=\{x : x \text{ is a Freeman first year MBA student}\}$$

  This is read as: "$X$ is the set of all elements $x$ such that $x$ has a certain property" (here, the property is that $x$ is a Freeman first year MBA student). Note that usually, *sets are represented by upper case letters* while *elements of the set are represented as lower case letters*.

  Example 2: Y={……,-3,-2,-1,0,1,2,3,…….} is the set of all integers

  While in Example 1, we had a **finite** set of elements, in this example, we have a set of an **infinite** number of elements.

A **function** *f* is a relation (or a rule) defined between two sets *X* and *Y*. The function associates to each value in the set *X*, a corresponding value (*and only one corresponding value*) in the set *Y*. *X* is called the **domain** of the function, and *Y* is called the **co-domain** of the function.

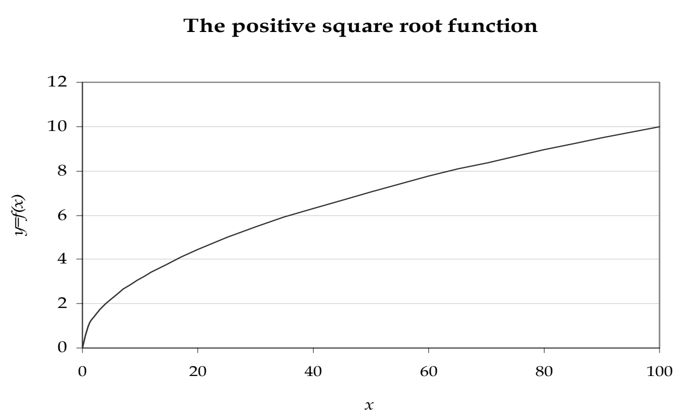
The set of all elements in *Y* associated with elements of *X* by the function is called the **range** of the function. Note that the range of the function may or may not be the entire set *Y*. Functions are sometimes denoted as $f : X \mapsto Y$. This notation is, however, rare. More frequently, one sees functions denoted as:

$$\boxed{y = f(x);\ x \in \mathrm{X}} \qquad \ldots \qquad (1)$$

The $\in$ in the above equation is the math symbol for "belongs to". That is, Equation (1) says that for every element *x* that belongs to the set X, the function *f* associates a value *y*. Let's see some examples:

<u>Example 3</u>: $y = +\sqrt{x}$ , the positive square root function

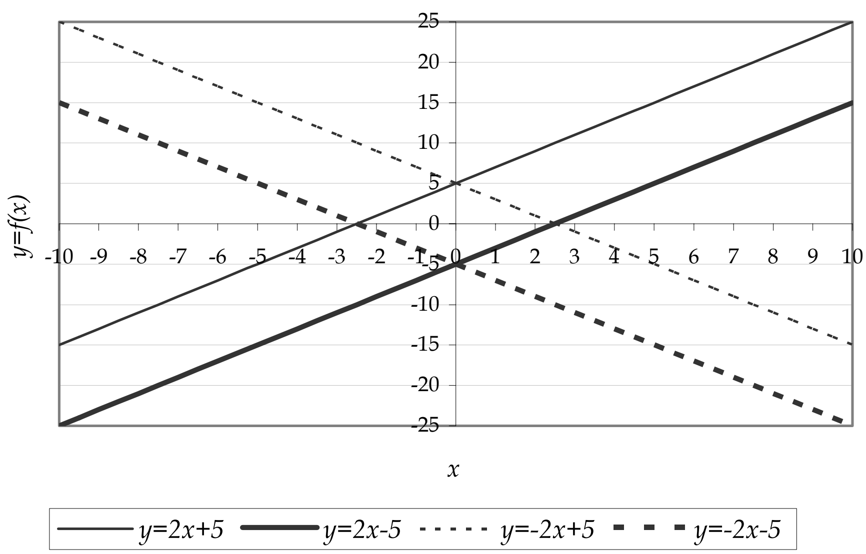
What are the domain and range of this function? We know that only non-negative numbers have *real* square roots[1], so this function is defined only for numbers 0 and higher. Further, we can take square roots not only of numbers such as 4, 9, and 16, but of 225.25, 174.5689 and so on. So we can conclude that the domain of this function is the set of *all non-negative real numbers*. It can be seen that the range of this function is also the set of *all non-negative real numbers*. Let's look at a graph of this function.




[1] Square roots of negative numbers are called *imaginary* numbers (as opposed to real numbers), and play a very important role in engineering and the physical sciences. But since we deal only with real numbers in finance, we will safely ignore this fact, and behave here as if imaginary numbers do not exist.

<u>Example 4</u>: I have graphed below four **linear** functions, i.e. functions whose graphs are straight lines. These are functions that are of the general form: $y = ax + b$. In the above equation, $a$ is called the **slope** of the function, while $b$ is called the **vertical intercept** of the function.

**Some linear functions**




Note two things:

1) The vertical intercept $b$ from the above equation is simply the point at which the graph intersects the vertical, or Y-axis.
2) Positive slopes correspond to upward sloping lines, while negative slopes correspond to downward sloping lines.

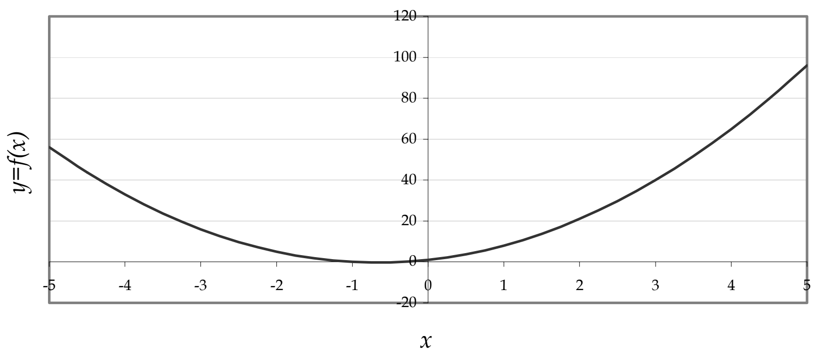
<u>Example 5</u>: Let's now see how a **quadratic** function looks like. The general form of a quadratic function is $y = ax^2 + bx + c$. Let us see two specific quadratic functions:

<u>5a</u>: $y = 3x^2 + 4x + 1$

<u>5b</u>: $y = -3x^2 - 4x - 1$

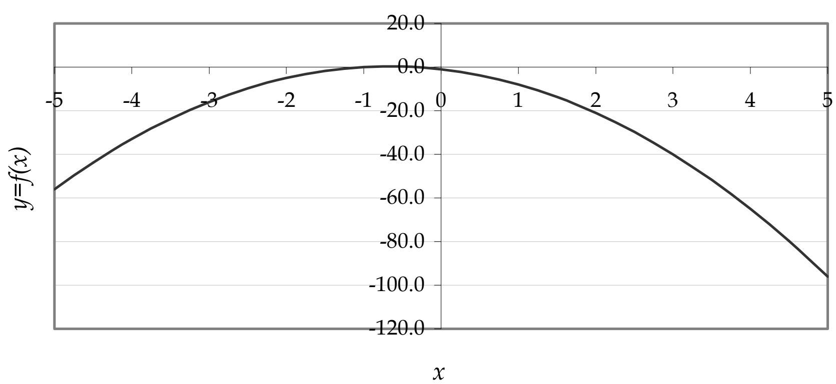
The first function is: $y = 3x^2 + 4x + 1$. This function is bowl-shaped, as you can see below. Such functions are called **convex** in shape.

**A convex quadratic function**




The next function, graphed below, is $y = -3x^2 - 4x - 1$. This function is the shape of an inverted bowl. Such functions are called **concave** in shape.

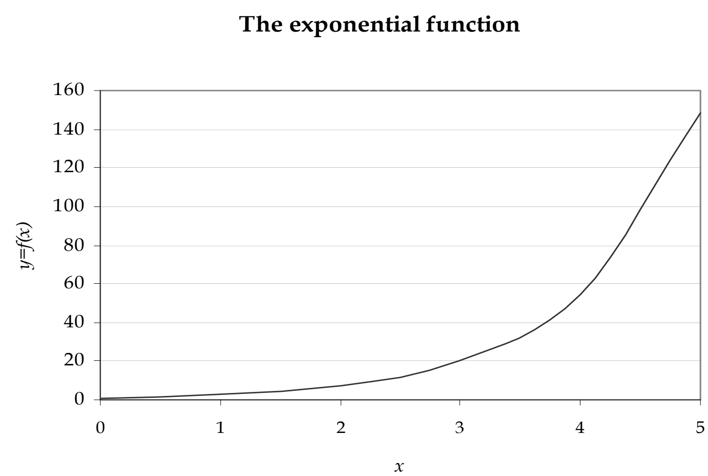
**A concave quadratic function**




Note that **convexity** and **concavity** are not properties of quadratic functions alone. Most (and I am deliberately being quite loose with "most" here) smooth, non-linear functions can be categorized as being either convex or concave. The shape of functions turns out to be very important, not only in this course, but in much of economics and finance. In particular, I will expect you to recognize the shape of a function when you see one.

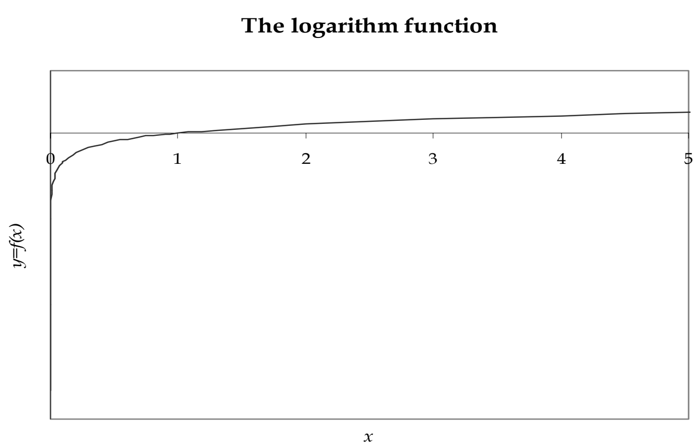
<u>Example 6</u>: Let's now look at a couple of functions that we encountered in option pricing.

<u>6a</u>: $y = e^x$




This function is called the **exponential** function. The number e is a special number in mathematics equal in value to 2.71828…., and is called the base of the natural logarithm.

<u>6b</u>: $y = \ln(x)$



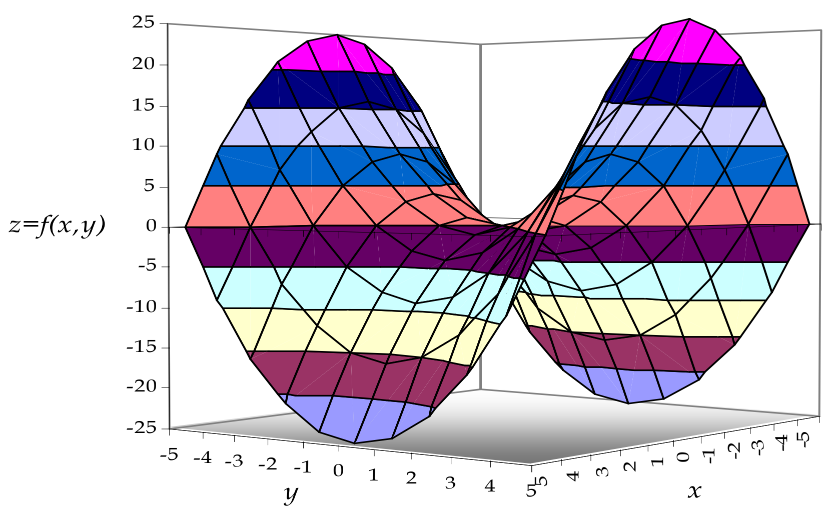

This function is called the (natural) **logarithm** function. As we shall see in later sections, the exponential and logarithm functions play a very important role in the mathematics of finance.

Before we go ahead, let us stop and think: why do we need functions in economics and finance? The answer is: because we need a way to quantify the relationship between two variables of interest. For example, we might want to understand how the market value of a company changes with the amount of debt it has. Or we might want to know the effect of a change in interest rates on the growth in Gross Domestic Product (GDP) of the economy. Or we might want to understand the impact of a stock's volatility on the price of a call option written on that stock.

In Examples 3 through 6, we looked at the basics of **univariate calculus,** i.e. *functions of one variable*. Each value of $y$ depended *only* on the value of $x$. As $x$ varies (which is why it is called a **variable**), the corresponding value of the function $y$ changes along with it. This is perfectly fine for representing some relationships. But frequently, an economic quantity might depend on more than one variable. For instance, GDP growth might depend on government spending, in addition to interest rates. What then? We can easily extend the logic of the previous examples to handle this new wrinkle. Let's now dig a bit deeper into the basics of **multivariate calculus** (more than one variable).

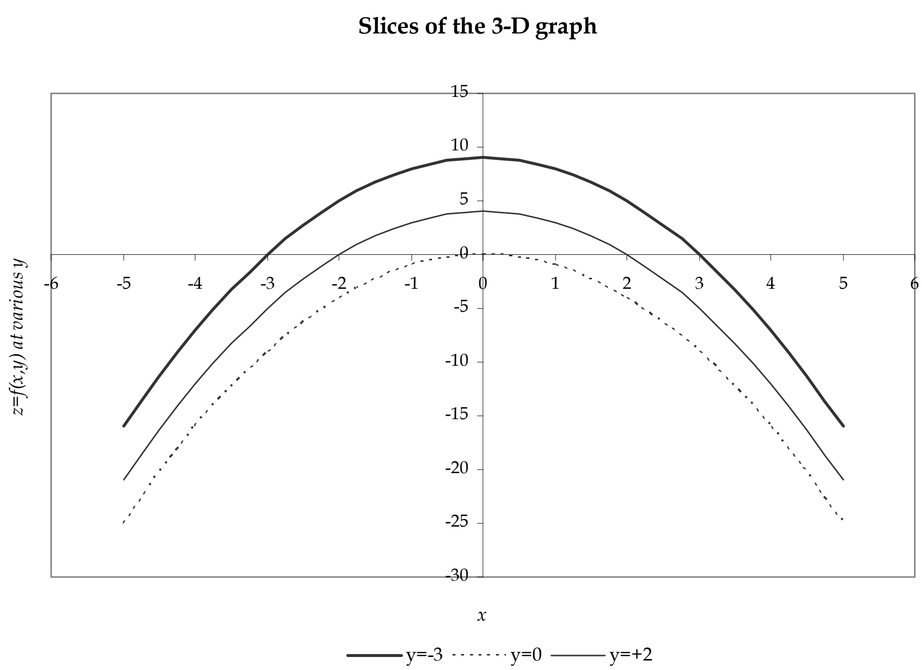
Example 7: Let's look at the function: $z = f(x, y) = x^2 - y^2$

A two-variable function



Here, since we have two variables, we need two axes to represent their variation. We have represented the function value $f(x,y)$ for each combination of $x$ and $y$ on the third axis.

If this graph appears too confusing, we have to understand that there are two things changing at the same time at every point. To get clearer pictures in our mind, we frequently fix the value of one variable, and see what happens to the function value when the other variable is allowed to change. For example, fixing the value of $y$ at -3, 0 and 2 gives us the following graph (rather, three pictures on the same graph).




The best way to think of this graph is as an assembly of vertical "slices" of the original 3-dimensional graph from last page at the points $y$=-3, $y$=0, and $y$=+2.

- **UNIVARIATE CALCULUS**

Continuity

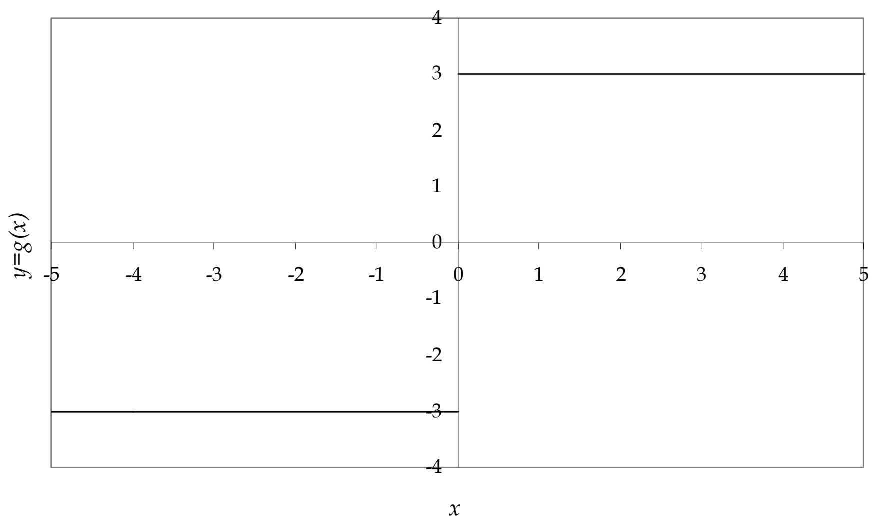
The concept of a **continuous** function is central to all calculus. Economists routinely write statements like "we assume that the so-and-so function is twice continuously differentiable" in journal articles and in text books. What does it mean when they say such high-falutin' things?

The best way to understand this concept is, not surprisingly, pictures. Geometrically speaking, a function is continuous if its graph has no breaks. All the functions we have seen so far in Examples 3 through 7 are continuous. Let us look at a **discontinuous** function called the **step function** to see the difference.

Example 8: The step function

This function can be written as: $y = g(x) = \begin{cases} +3, & x \geq 0 \\ -3, & x < 0 \end{cases}$

The discontinuous step function




At the point $x$=0, the function abruptly *jumps* from a value of -3 to +3. The transition from -3 to +3 is not continuous or smooth. That's why this is a discontinuous function.

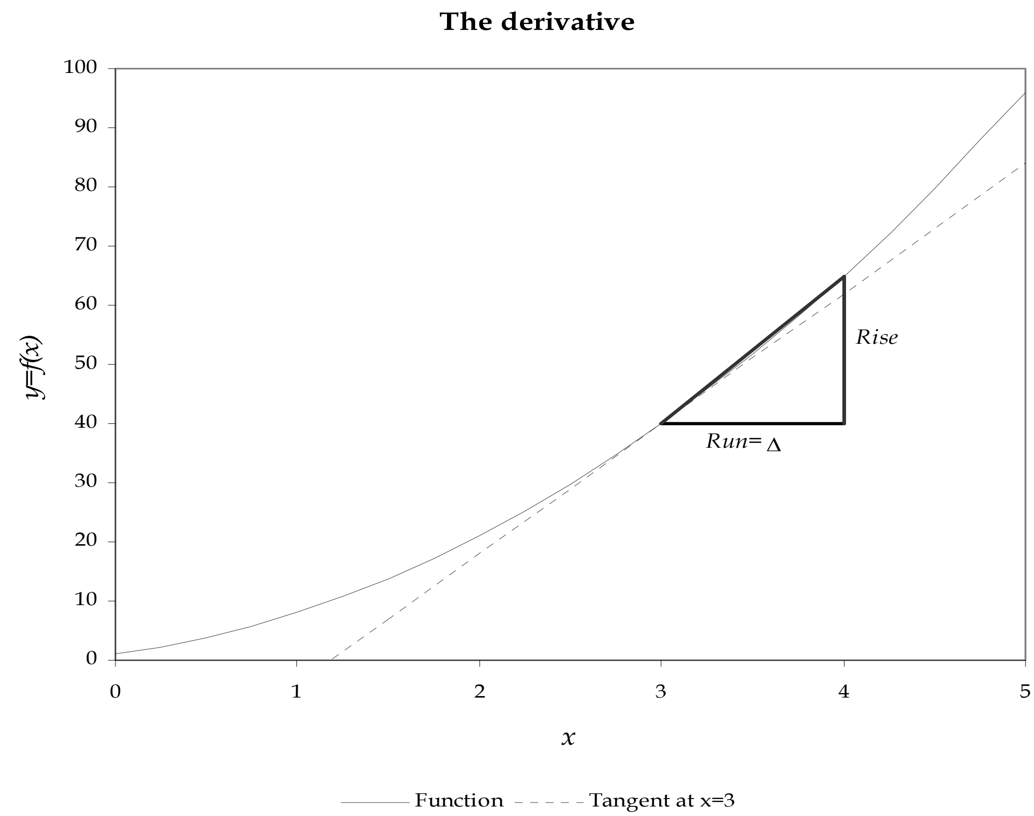
Differentiability and derivatives of a (univariate) function

The derivative[2] of a (univariate) function is defined as follows.

Let $y=f(x)$ be a function. Then the **derivative** of $f(x)$ with respect to (henceforth w.r.t.) $x$ is defined as: $f^{'}(x) = \frac{dy}{dx} = \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}$ … (2)

[2] Do not confuse this mathematical term "derivative" with the term "derivative" in finance, as in "Options, futures and other *derivatives*" by John C. Hull

Example 5a (extended): To see what this really means, let us take a function, $y = 3x^2 + 4x + 1$. You will recognize this as the quadratic function from Example 5a. Let us try and find the derivative of this function at the point *x*=3. I have reproduced below the graph from Example 5a, along with a few additions. Notice I have focused only on (zoomed into, if you will) the part of the graph for $x \geq 0$.

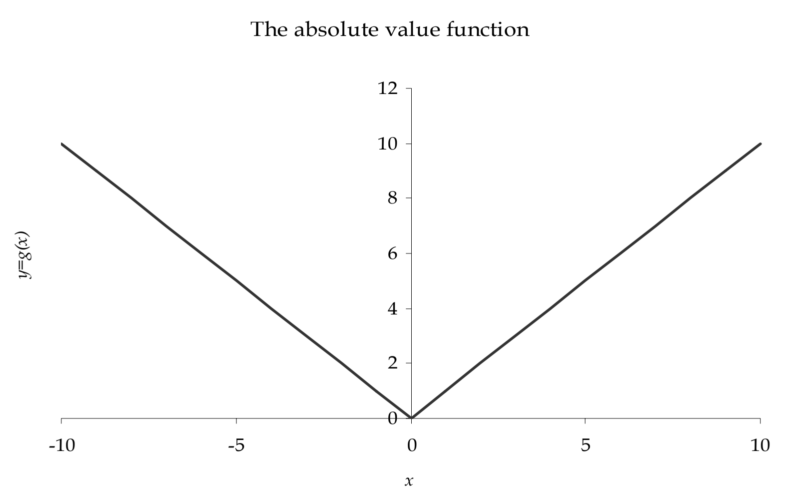



Let's pick an initial Δ=1. The line denoted *Rise* covers a vertical distance of $f(x+\Delta) - f(x) = f(3+1) - f(3) = f(4) - f(3) = 25$ units. The line denoted *Run* is equal to Δ=1. Dividing one by the other, the thick diagonal line of the right triangle has a slope equal to: $\frac{Rise}{Run} = \frac{f(x+\Delta) - f(x)}{\Delta} = \frac{f(4) - f(3)}{1} = 25$. But the derivative is defined as the **limit** of his slope as Δ approaches 0. It is easy to see that as $\Delta \to 0$, the diagonal line approaches the **tangent** to the function at *x*=3 (the dotted line). Thus, the slope of the tangent at *x*=3 will give us the derivative of the function at that point. It turns out that the slope of the tangent there is 22, which is the derivative of the function $y = 3x^2 + 4x + 1$ at the point *x*=3. What about at other points? Do we have to do this exercise at every point on the graph, i.e. for $x$ =-3, -1, 0, 2,… etc.? Not really. The above exercise was only for explaining the concept of the derivative. There are rules for **differentiating** (finding the derivative) of many, many functions[3]. The most basic rule of them all is that the derivative of $x^n$ is $n.x^{n-1}$. Let us use this rule in our simple example. The derivative of $y = 3x^2 + 4x + 1$ is 3(2$x$)+4= 6$x$+4. At the point $x$=3, this can be evaluated as 6(3)+4=22, as advertised in the last page!

Now, a function is **differentiable** if we can find its derivative at every point at which the function is defined. In plain English, this means that a function is differentiable at a given point, if we can draw an obvious tangent to it at that point. **Differentiability** turns out to be a very important property of functions, just like its first cousin, continuity. Again, to understand this concept, let's look at the example of a function that is not differentiable at a certain point.

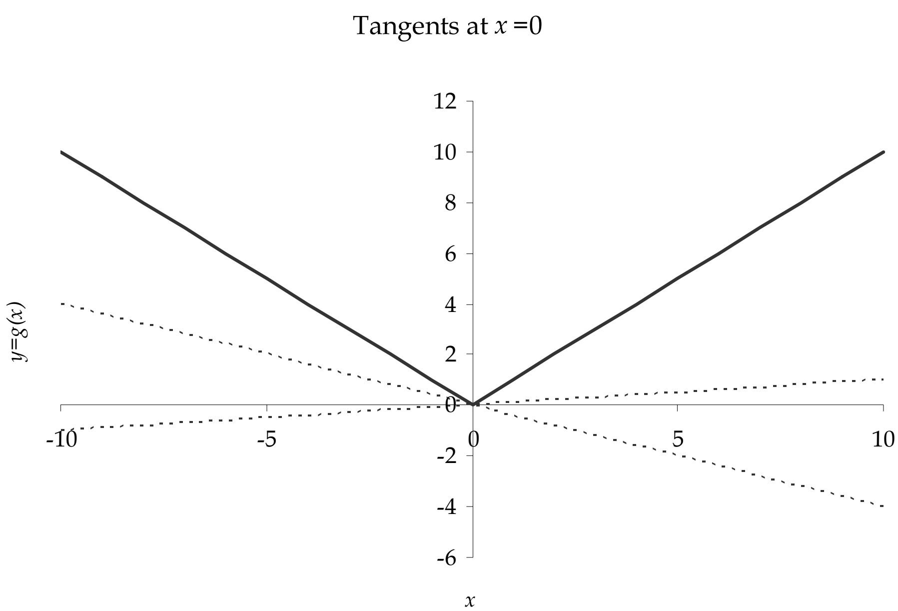
<u>Example 9</u>: Consider the **absolute value** function $y=g(x)=|x|$. This function returns the absolute value (magnitude) of any real number, ignoring its sign. So, it can be defined more clearly as: $y = g(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$

As usual, let's look at a graph of this function.




[3] Most basic calculus text books will give you the rules of differentiation

It should be obvious that there is no well-defined tangent that we can draw to this function at the point $x$=0. In fact, there are many, many tangents (an infinite number of them) that we could draw through the point $x$=0. A couple of possible tangents are shown in the following graph as dashed lines. Therefore, it is clear that this function is not differentiable at **the origin** –the point ($x$=0,$y$=0).




Note carefully: The function above is *continuous* at all points, although *not differentiable* at $x$=0. This leads us to a couple of observations that are often useful:

- ✓ If a function is differentiable at all points, then it is continuous at all points
- ✓ Even if a function is continuous at all points, it need not be differentiable at all points

<u>Higher Order Derivatives</u>

We have so far learned that if we differentiate a function once, we essentially plot its slope at all points at which the function is defined. For instance, in the extension to Example 5a, we found that the derivative of the function $f(x) = 3x^2 + 4x + 1$ is $f^{'}(x) = 6x + 4$. Technically, this is called the **first derivative** of the function *f(x)*. this means that we have differentiated the original function once. What if we differentiate $f^{'}(x) = 6x + 4$ w.r.t. x again? The answer to this computation is called the **second derivative** of *f(x)*. Now the derivative of $f^{'}(x) = 6x + 4$ is $f^{''}(x) = 6$. We can differentiate this one more time to get $f^{'''}(x) = 0$ and so on. So, for any function *f(x)*, by successive differentiation, we can obtain all **higher order derivatives** – first, second, third, fourth and so on …

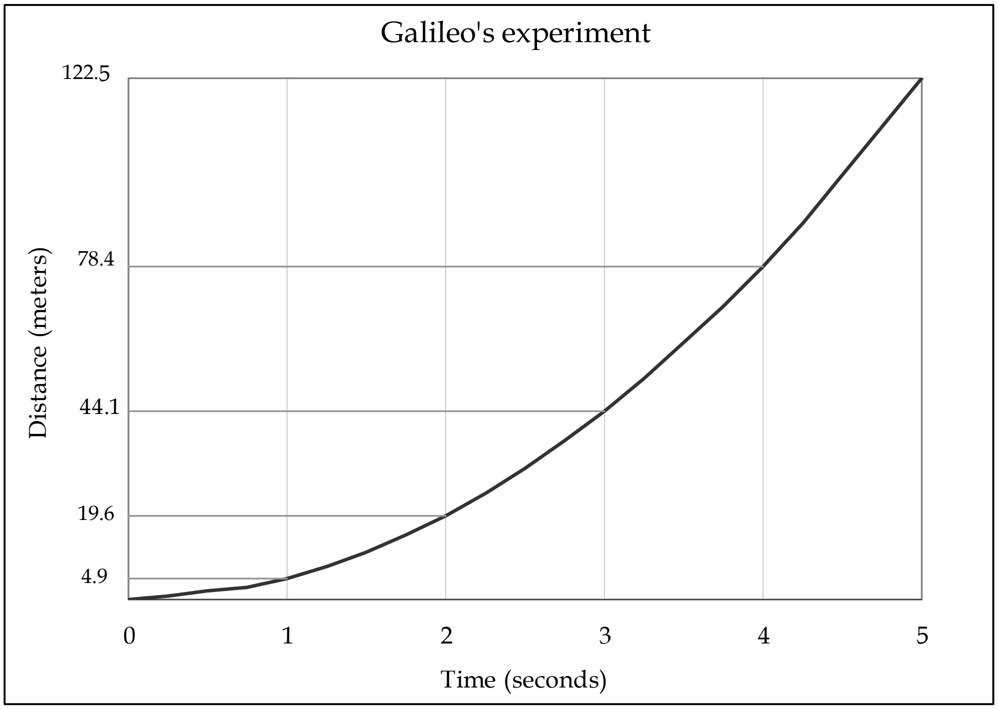
<u>Derivatives: A rate of change interpretation</u>

Derivatives are very nifty math devices as they can be interpreted as rates of change. Let us understand this interpretation by means of a simple example.

<u>Example 10: (Galileo's experiment)</u>

In the 1580s, Galileo Galilei, the famous Italian physicist (while studying at the University of Pisa) is supposed to have dropped a 10-pound ball and a 1-pound ball, both from the top of the famous leaning tower of Pisa, in order to prove that the two bodies, in spite of their difference in mass, fall at the same acceleration, thereby shattering the Aristotlean conception that the 10-pound ball will fall at a rate 10 times faster than the 1-pound ball. Although science historians doubt whether Galileo actually performed this experiment, we now know that his basic thesis was indeed correct. Both bodies do fall at the same acceleration due to the Earth's gravity which is denoted in physics by the letter *g*, and has a value of 9.8 meters/second$^2$.

If we drop a ball from the top of a tower 122.5 meters high, the distance traveled by the ball towards earth in *t* seconds is given by the familiar equation $S(t) = \frac{1}{2} g t^2$. This is a quadratic function, as we know. A graph of this function is shown on the next page.




As can be seen from the above graph, the ball takes 5 seconds to traverse the 122.5 meters and reach earth. It covers 4.9 meters in the first second, 14.7 meters (=19.6-4.9) in the second second, 24.5 meters (=44.1-19.6) in the third second, 34.3 meters (=78.4-44.1) in the fourth second, and the last 44.1 meters (=122.5-78.4) in the final, fifth second. That means the ball is gathering speed as time passes. There are two things to note from this analysis:

1) First, the rather obvious observation that the distance $S$ (function) traveled by the ball is increasing as time $t$ (variable) increases. Such a function is said to be an **increasing function**. Put another way, the *rate of change* in the function is positive at all times $t$. Mathematically speaking, this corresponds to a **positive first derivative**, i.e. $S(t)$ is increasing because $S'(t)$ is positive at all points t.
2) There is another, slightly more subtle observation: the rate of change is not uniform. The distance traveled per second, (or the rate of change of distance) a.k.a. the *velocity* of the ball is *increasing*. This means the function is **increasing at an increasing rate**! This corresponds to the fact that the first derivative itself is an increasing function, i.e. the **second derivative $S''(t)$ is also positive**.

Using the rules of differentiation, we have the following quantities:

Distance: $S(t) = \frac{1}{2} g t^2$

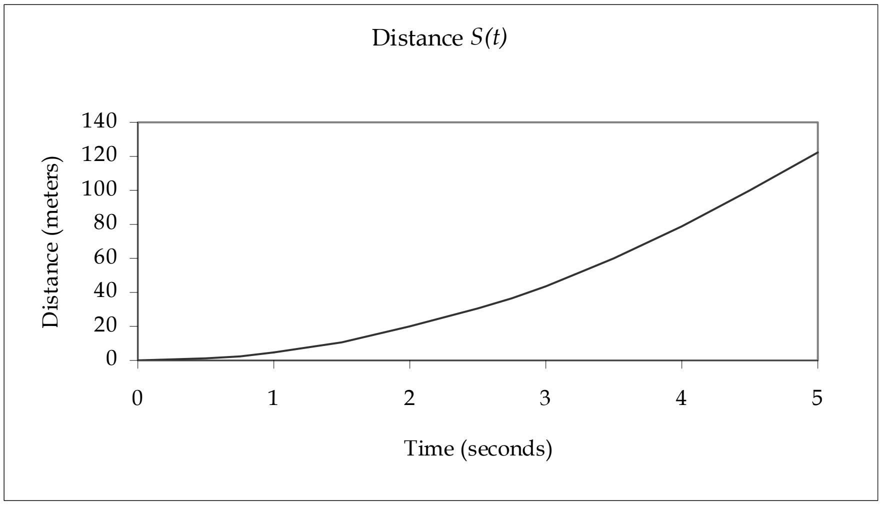
Velocity: $v(t) = S'(t) = gt > 0$ $\Rightarrow$ Distance is increasing with time

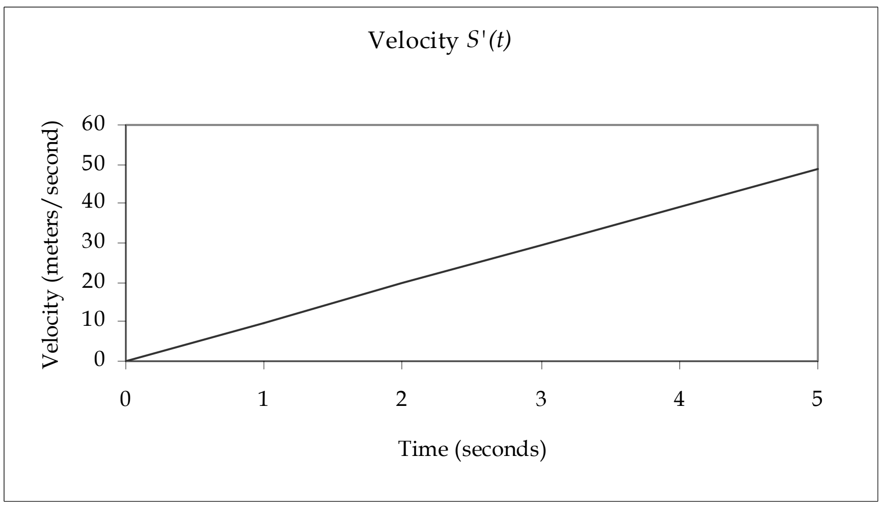
Acceleration: $v'(t) = S''(t) = g > 0$ and constant $\Rightarrow$ Velocity is increasing at a constant rate

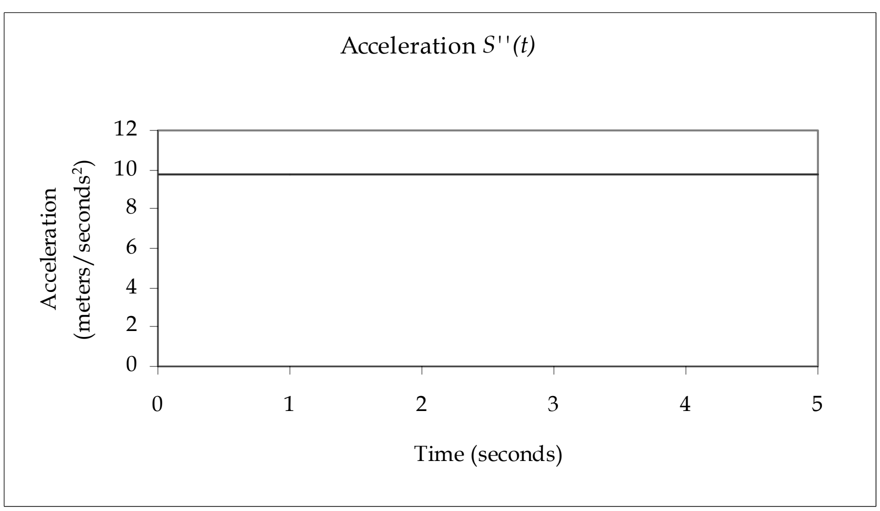
It is this interpretation of derivatives as successive rates of change that give calculus much of its elegance and content. (It also gives us an insight into why Isaac Newton had to practically invent the entire domain of calculus to describe his laws of motion)[4].

It is useful to see the distance, velocity, and acceleration, all as functions of time in one picture, as in the next page.

[4] Note that Gottfried Wilhelm Leibnitz (1646-1716) is also (rightly) jointly credited along with Newton for the invention of calculus


Distance S(t)
Distance (meters)
140
120
100
80
60
40
20
0
0
1
2
3
4
5
Time (seconds)


Velocity S'(t)
Velocity (meters/second)
60
50
40
30
20
10
0
0
1
2
3
4
5
Time (seconds)


Acceleration S''(t)
Acceleration (meters/seconds2)
12
10
8
6
4
2
0
0
1
2
3
4
5
Time (seconds)

<u>Derivatives and Curvature</u>

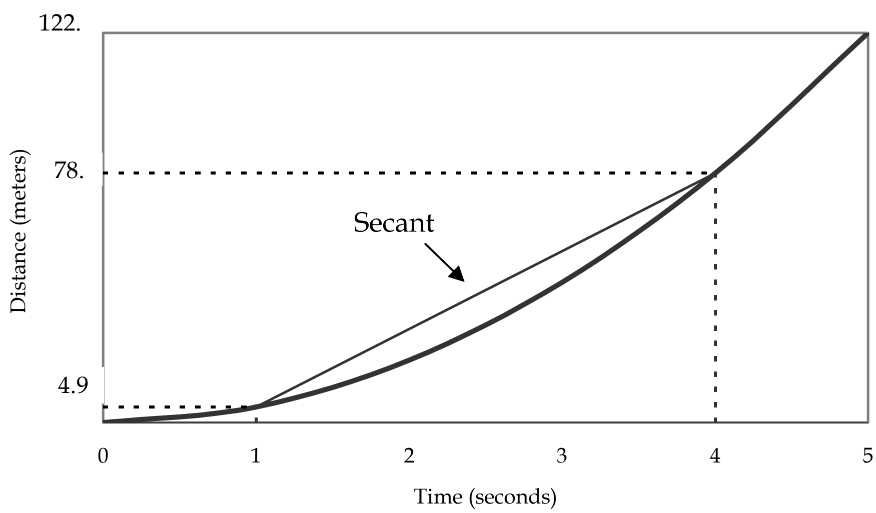
Remember convexity and concavity of functions? It turns out that knowledge of the first and second derivatives of a function enables us to tell which functions are convex and which are concave, without even graphing them.

*Convexity*: Let $y=f(x)$ be a function. Choose two points $x_1$ and $x_2$ at which the function is defined. Then the function is said to be **convex** if the following property holds.

$$f(\overline{x}) \leq \lambda f(x_1) + (1-\lambda) f(x_2), \text{ where } \overline{x} = \lambda x_1 + (1-\lambda) x_2;\ \lambda \in [0,1] \qquad \ldots (3)$$

Geometrically, this simply means that the straight line joining the values of the function at $x_1$ and $x_2$ is always *above* the graph of the function between $x_1$ and $x_2$. The following picture helps illustrate this property.

Galileo's experiment: A Convex function




You will recognize this function as the same one from Example 10. Here, I have picked $x_1$=1, $x_2$=4. The straight line joining the function values at 1 and 4 is called the secant line. (As an aside, in the limit, as we reduce the points $x_1$ and $x_2$ to zero, the **secant** gradually becomes the **tangent** at the terminal point).

Notice that the secant always lies above the curve (function) at very point in between these two points. And this happens no matter which $x_1$ and $x_2$ we pick! This is the defining property of a convex function. As we vary the value of $\lambda$ from 0 to 1, the value of $\overline{x} = \lambda x_1 + (1-\lambda)x_2$ traces every point in between $x_1$ and $x_2$. Therefore, we should read $\overline{x} = \lambda x_1 + (1-\lambda)x_2$ as "every point between $x_1$ and $x_2$".

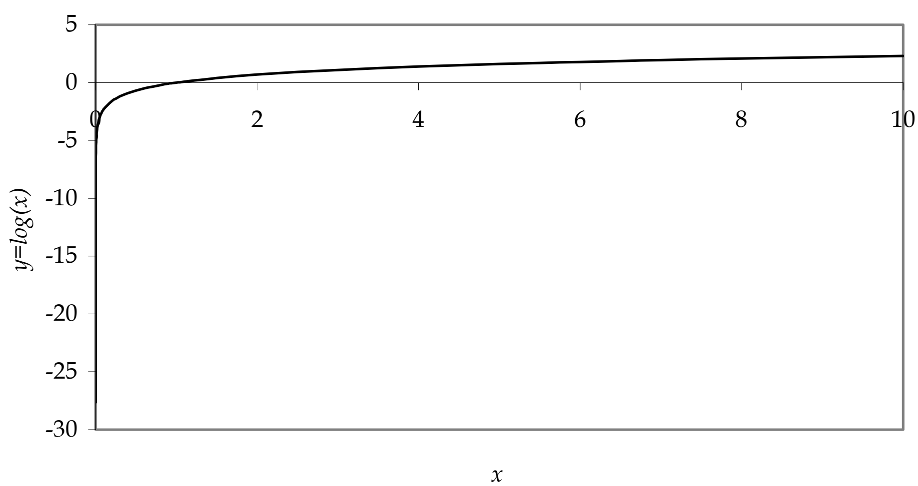
As I said before, derivatives give us more insight into the shape of a function., without this elaborate process of graphing the function. Convex functions have the property that $f''(x) > 0$ i.e. the second derivative is *positive*. You should now go back and check that this is indeed true for Example 10 involving Galileo's experiment.

*Concavity*: As you might have guessed by now, concavity implies the opposite property for a function. Formally, a function *y=f(x)*, (again picking two points $x_1$ and $x_2$ at which it is defined) is said to be **concave** if the following property holds.

$$f(\overline{x}) \geq \lambda f(x_1) + (1-\lambda)f(x_2), \text{ where } \overline{x} = \lambda x_1 + (1-\lambda)x_2;\ \lambda \in [0,1] \qquad \dots (4)$$

Geometrically, again, this means the opposite of convexity. For concave functions, the straight line joining the values of the function at $x_1$ and $x_2$ (secant line) is always *below* the graph of the function between $x_1$ and $x_2$. Let us draw the picture of the (natural) logarithm function from Example 6b to illustrate concavity.

The (natural) logarithm function: A concave function




Concavity, as you might expect, means that $f''(x) < 0$, i.e. the second derivative is *negative*.

To summarize:

$f'(x) > 0$: (Strictly) Increasing function

$f'(x) < 0$: (Strictly) Decreasing function

$f''(x) > 0$: (Strictly) Convex function … (5)

$f''(x) < 0$: (Strictly) Concave function

<u>Note</u>: For a linear function, $f''(x) = 0$, which means that it is *neither convex nor concave*.

- **MULTIVARIATE CALCULUS**

Consider the function $f(x, y) = x^2 - y^2$. As we discussed (in Example 7) above, this is a function of *two* variables. It turns out we can readily extend the concepts of continuity, differentiability, derivative and function curvature to this *multivariate* case. We only need to modify our notation somewhat. I will briefly summarize below what we need here.

<u>Differentiation</u>

We start studying multivariate calculus by looking at the change in the function changing one variable at a time, keeping all the other variables constant.

Let's say there us a multivariate function of *n* variables, $y = f(x_1, x_2, \cdots, x_n)$. Then, we define the **partial (first order) derivative** of the function *f* w.r.t. $x_i$, the $i^{\text{th}}$ variable as:

$$f_i = \frac{\partial y}{\partial x_i} = \lim_{\Delta \to 0} \frac{f(x_1, x_{2,\dots,} x_i + \Delta, \cdots, x_n) - f(x_1, x_{2,\dots,} x_i, \cdots, x_n)}{\Delta} \quad \dots \quad (6)$$

It should be clear from the above definition that we are considering the effect on the function of a change only in the $i^{\text{th}}$ variable, treating all other variables constant.

Higher order derivatives are defined in the same way. For example, $f_{ii} = \frac{\partial^2 y}{\partial x_i^2}$ is the **second order partial derivative** of the function w.r.t. the $i^{th}$ variable $x_i$. We also could define the **cross partial** derivative $f_{ij} = \frac{\partial^2 y}{\partial x_i \partial x_j}$, as the result of differentiating the function once with respect to the $i^{th}$ variable $x_i$, and then by the $j^{th}$ variable $x_j$.

Example 11: Consider a simple production function that you would find in any basic microeconomics text book. The quantity of output produced by a firm $Q$ depends on two inputs capital ($k$), and labor ($l$) according to the function:

$$Q(k,l) = k^{1/2} l^{1/2}$$

This particular production function is called the **Cobb-Douglas production function**.

In this simple world, starting with 100 units of capital and 25 units of labor, output is: $Q(100,25) = 100^{1/2} 25^{1/2} = 10 \times 5 = 50$ units

Using the rules of differentiation, we can compute partial derivatives with respect to both variables as follows:

$$Q_k = \frac{\partial Q}{\partial k} = \frac{1}{2} k^{-1/2} l^{1/2}$$ (remember to treat $l$ as constant), and

$$Q_l = \frac{\partial Q}{\partial l} = \frac{1}{2} k^{1/2} l^{-1/2}$$ (remember to treat $k$ as constant)

Evaluating these derivatives at $k$=100, and $l$=25, we have:

$$\left.\frac{\partial Q}{\partial k}\right|_{k=100,l=25} = \frac{1}{2} 100^{-1/2} 25^{1/2} = 0.25$$, and $$\left.\frac{\partial Q}{\partial l}\right|_{k=100,l=25} = \frac{1}{2} 100^{1/2} 25^{-1/2} = 1$$

The **first-order partial derivative** $\frac{\partial Q}{\partial k}$ ($\frac{\partial Q}{\partial l}$) is interpreted as the **marginal[5] product of capital** (**labor**) i.e. holding everything else constant, *at the margin* (given where we

[5] Marginal quantities are of great importance in economic reasoning, and indeed in real life situations. At 3 AM before final exam day, students deciding whether to go to sleep trade off the

are), what will be the impact of a unit increase in capital (labor)? We can use these partial derivatives to evaluate the impact of say, 10 more units of capital, or 2 more units of labor. How? Let us see…

We can write

$\Delta Q = Q(k+\Delta k, l) - Q(k,l) \simeq \frac{\partial Q}{\partial k}.\Delta k$, and

$$\Delta Q = Q(k, l+\Delta l) - Q(k,l) \simeq \frac{\partial Q}{\partial l}.\Delta l \qquad \ldots \qquad (7)$$

which are shorthand for saying that we can make use of the marginal products for estimating the change in the function due to a $\Delta k$ units change of capital (holding the labor input constant) or a $\Delta l$ units change in labor (holding the capital input constant).

Given that we currently have 100 units of capital and 25 units of labor, what is the marginal impact of an increase in 10 units of capital? Using the above formula, we have: $\Delta Q \simeq \left.\frac{\partial Q}{\partial k}\right|_{k=100,l=25}.\Delta k = 0.25 \times 10 = 2.5 \text{ units}$. How good is this approximation? We know that 10 more units of capital take us to 110 units of capital, which means production will be: $Q(110,25) = 110^{1/2}25^{1/2} = 52.44 \text{ units}$, an increase of 2.44 units. As you can see from comparing 2.44 units with 2.50 units, the approximation is quite good.

In fact, the smaller the value of $\Delta k$, the better the approximation, and vice versa. Let us say we want to measure the impact of 44 more units of capital. The marginal product formula (7) gives us $\Delta Q \simeq \left.\frac{\partial Q}{\partial k}\right|_{k=100,l=25}.\Delta k = 0.25 \times 44 = 11 \text{ units}$. But we know that the actual production with 44 more units of capital will be:

$$Q(144,25) = 144^{1/2}25^{1/2} = 60 \text{ units}, \text{ a change of 10 units}$$

---

*marginal value* of one more hour of studying (a bit more cramming) versus the *marginal cost* of the hour of study (fatigue from sleep deprivation). Note that at this point, the tradeoff is strictly between *marginal* quantities and not between *total* quantities (total hours studied in the last week or semester versus total time slept in the last week or semester).

Here, the approximation isn't as good: 11 units compared with 10. Why is this happening? Recall that the partial derivative is calculated at an initial value of *k*=100. Recall also that the derivative is the slope of the function. The farther away we get from this number, the worse the approximation using the derivative, since the slope is changing more and more as we move away farther and farther from *k*=100. This is to be expected, as we are using a **linear approximation** to a **non-linear function**.

Obviously, a similar exercise could be conducted for understanding the impact on production of an increase in labor inputs. We can see that the marginal impact on output of $\Delta l = 2$, i.e. two more units of labor is given by:

$$\Delta Q \simeq \left.\frac{\partial Q}{\partial l}\right|_{k=100,l=25} .\Delta l = 1 \times 2 = \text{ 2 units}$$

Holding capital at *k*=100 units, and increasing labor by $\Delta l = 2$ units to 27 units, we have $Q(100,27) = 100^{1/2} 27^{1/2} = 51.96 \text{ units}$, an increase of 1.96 units which compares very well with the approximate estimate of 2 units from above.

The next obvious question to ask is: What will happen if *both* capital and labor inputs change? It turns out we can extend the analysis of the last page to estimate the change in production due to a *simultaneous* change in both inputs. To do this we define the **total derivative** or **total differential** of a function with respect to changes in more than one variable.

$$\Delta Q = Q(k + \Delta k, l + \Delta l) - Q(k,l) \simeq \frac{\partial Q}{\partial k}.\Delta k + \frac{\partial Q}{\partial l}.\Delta l \qquad \dots \qquad (8)$$

This equation says that the total change in the (production) function of varying *k* and *l* simultaneously can be approximated by the sum of the estimates from the one-variable-only changes.

Following the example from above, what if capital increased by 10 units ***and*** labor increased by 2 units? Equation (8) tells us that we can estimate this increase by:

$$\Delta Q \simeq \left.\frac{\partial Q}{\partial k}\right|_{k=100,l=25} .\Delta k + \left.\frac{\partial Q}{\partial l}\right|_{k=100,l=25} .\Delta l = (0.25 \times 10) + (1 \times 2) = 4.5 \text{ units}$$

What is the actual increase? We can compute this by using the production function as : $Q(110,27) = 110^{1/2}27^{1/2} = 54.4977 \text{ units}$, an increase of 4.4977 units, which is a very, very good approximation (compare 4.5 units to 4.4977 units).

Note also that these linear approximations using one-variable and total differentials are not special properties of this Cobb-Douglas function, but are generally applicable to a wide range of functions of more than one variable. More generally, the total differential for a function of *n* variables, $f(x_1, x_2, \cdots, x_n)$ is given by a straightforward generalization of equation (8):

$$\Delta f = f(x_1 + \Delta x_1, \cdots, x_n + \Delta x_n) - f(x_1, \cdots, x_n) \simeq \frac{\partial f}{\partial x_1}.\Delta x_1 + \cdots + \frac{\partial f}{\partial x_n}.\Delta x_n \quad \ldots \quad (9)$$

<u>Curvature</u>

How do we extend the concepts of convexity and concavity to the multivariate case? Simply by using second order partial derivatives. Consider again the general definition of a multivariate function: $y = f(x_1, x_2, \cdots, x_n)$. In this case, the shape of the function is an n-dimensional surface, and not a simple curve. We saw this earlier (in Example 7) with a function Hence, it makes sense to talk of convexity and concavity *with respect to a single variable, treating all other variables as fixed*. Let's continue Example 7 here:

<u>Example 7 (extended)</u>: The function is $f(x,y) = x^2 - y^2$

Let us use the rules of differentiation and find all the partial derivatives of this function.

First order partial derivatives: $f_x = \frac{\partial f}{\partial x} = 2x;\ f_y = \frac{\partial f}{\partial y} = -2y$

$$f_{xx} = \frac{\partial^2 f}{\partial x^2} = 2;\ f_{yy} = \frac{\partial^2 f}{\partial y^2} = -2$$

Second order partial derivatives: $f_{xy} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}(-2y) = 0$

$$f_{yx} = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}(2x) = 0$$

All derivatives of the third order and higher are zero here, a result not surprising, as the function is **of degree two** (quadratic) in both *x* and *y*.

Let us restrict our attention to this function in the range $x \geq 0$, $y \geq 0$, and note a few facts:

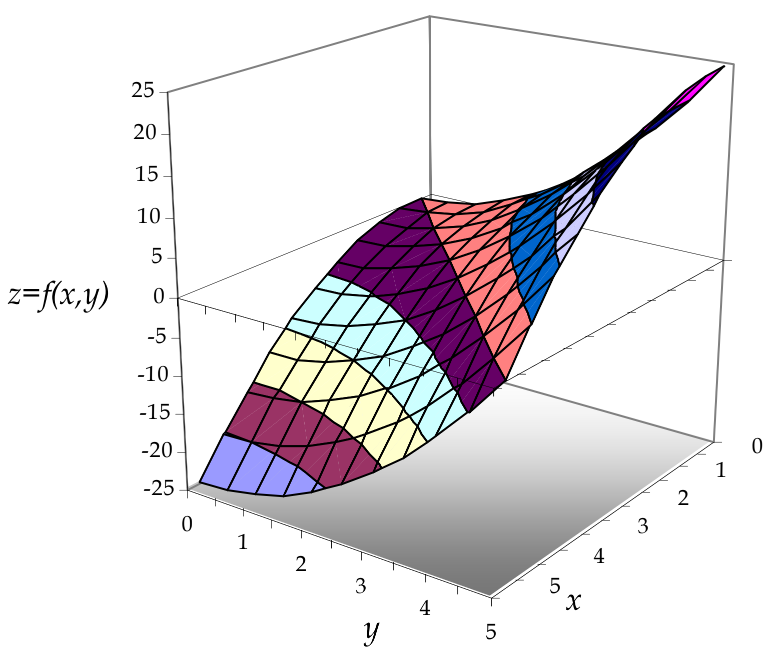
a) Over this range of *x* and *y*, we can see that $f_x = \frac{\partial f}{\partial x} = 2x \geq 0; f_y = \frac{\partial f}{\partial y} = -2y \leq 0$.

Using the sign of the first partial derivative, we can say that this function is (weakly) *increasing in x*, and (weakly) *decreasing in y*.

b) Over this range of *x* and *y*, $f_{xx} = \frac{\partial^2 f}{\partial x^2} = 2 > 0; f_{yy} = \frac{\partial^2 f}{\partial y^2} = -2 < 0$. Using the signs of these second partial derivatives, we can say that this function is (strictly) *convex in x*, and (strictly) *concave in y*.

Pictures will again tell the story better:

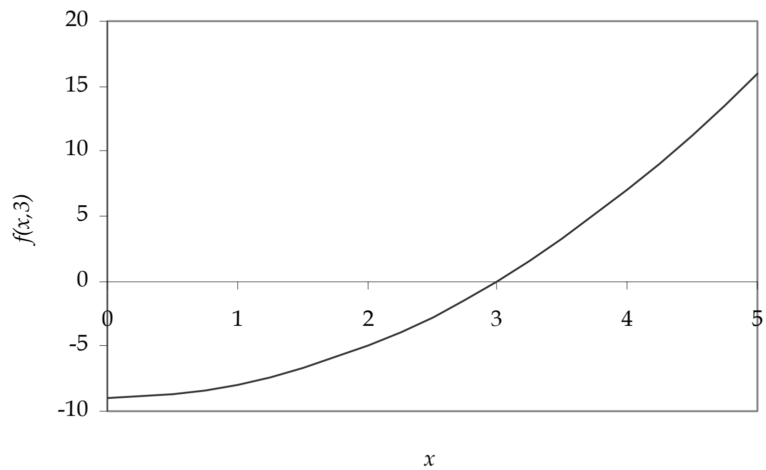
Example 7 (extended)



Above, we see a picture of the function itself. This is the same graph as was shown in Example 7, except now we have restricted the range to values of the function for $x \geq 0$, $y \geq 0$ (a.ka. the **first quadrant**, in the language of mathematics).

Let us fix the value of $y$ at $y$=3, and observe how the function varies with $x$ alone.

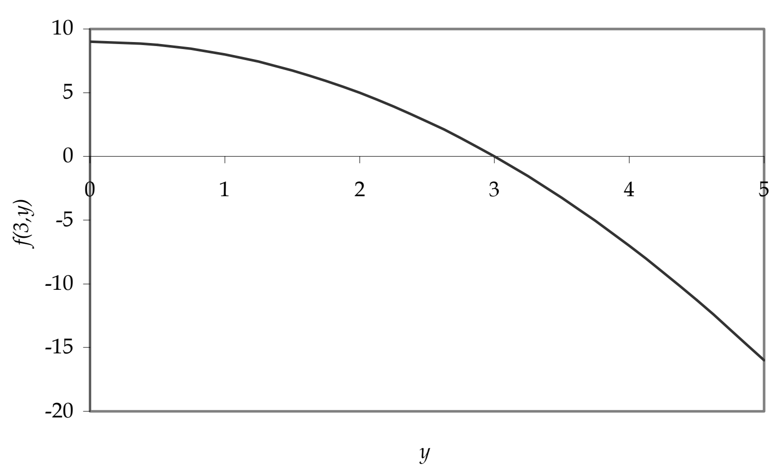
*f(x,y)* as a function of x alone




Clearly, this is an *increasing* and *convex* function in the variable $x$, as the signs of the partial derivatives calculated above revealed.

Now, let us fix the value of $x$ at $x$=3, and observe how the function varies with $y$ alone.

*f(x,y)* as a function of y alone



Clearly, this is a *decreasing* and *concave* function in the variable $y$, as the signs of the partial derivatives calculated above revealed.

The moral of the story from the above pictures is that, when one talks about a function of more than one variable as being increasing, or concave, one needs to specify increasing *in which variable*, concave *in which variable* and so on.

Note one final fact, before we leave this topic. From the above calculations:

$$f_{xy} = \frac{\partial^2 f}{\partial x \partial y} = f_{yx} = \frac{\partial^2 f}{\partial y \partial x} = 0$$

The cross-partials are both zero, which means that the slope of the function with respect to $x$ does not depend on $y$, and vice versa. Aside from the fact that the values of the cross-partials are zero, the important thing is that *they are equal*. In other words,

$$f_{xy} = \frac{\partial^2 f}{\partial x \partial y} = f_{yx} = \frac{\partial^2 f}{\partial y \partial x} \qquad \ldots \qquad (9)$$

That is, it does not matter whether we differentiate by $x$ first and then by $y$, or by $y$ first, and then by $x$. We always get the same answer. This important (and seemingly obvious) fact is very important in calculus, and goes by the name of **Young's theorem**. For instance, you should be able to verify that in Example 11, the cross-partial derivatives of the Cobb-Douglas production function are both equal to:

$$\frac{\partial^2 Q}{\partial k \partial l} = \frac{\partial^2 Q}{\partial l \partial k} = \frac{1}{4} k^{-1/2} l^{-1/2}$$

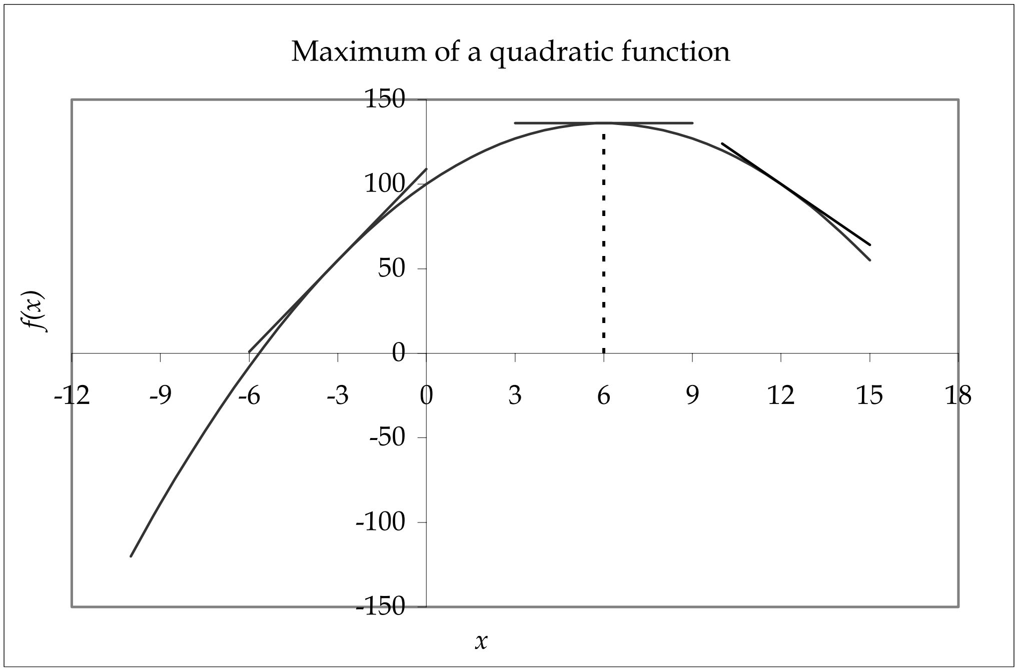
- **UNCONSTRAINED OPTIMIZATION**

Frequently, we wish to find the **maximum** and **minimum** values of a given function. Economists need such things since they routinely set up problems where they need to maximize profit or utility, or minimize cost in an endless variety of problems. It turns out calculus is the principal tool we use to solve such problems. As always, there are neat geometrical interpretations for every thing we can and do say in the language of calculus.

Let's start with a simple quadratic function, which will allow us to understand optimization in some detail.

<u>Example 12</u>: $f(x) = 12x - x^2 + 100$

A picture of this function has the following familiar shape.




Looking at the picture, it is clear that the maximum value of this function is achieved at $x$=6. Substituting the value of $x$=6 into the function, we have the value of $y = f(x)$ at this point equal to $y$=136. This much is obvious to anyone who knows how to read a graph. At least two questions come up at this juncture:

1) Do we have to draw a picture every time, or is there a more general method one can follow to identify the maximum of a function?
2) What about the minimum of a function, as opposed to the maximum?

Let us now proceed to answer both these questions.

A general methodology

Note that our example function is increasing upto a point ($x$=6), and decreasing thereafter. This can be seen from the shape of the function itself. We can also use the first derivative of the function to verify this observation. According to our earlier understanding, this means that the *first derivative*, or *slope* of the function is positive up to $x$=6, and negative thereafter. We can apply the basic rules of differentiation to find the first derivative of this simple function: $f'(x) = \frac{df(x)}{dx} = 12 - 2x$. This is the equation of the slope of this function at every point $x$.

For example, at $x$=-3, this function has a slope of 12-2(-3)=18. At $x$=12, the same equation yields a slope of 12-2(12)=-12. These slopes are shown in the above figure as *tangents* to the function at $x$=-3 and $x$=12 respectively. Now intuitively, one can sense that in going from a value of 18 to -12, the slope has to somewhere pass through a value of zero. Indeed, graphically, we can see that at the maximum of the function (i.e. when $x$=6), the slope is zero (the tangent is perfectly horizontal). Such points are the ones that will be the maxima or minima of a function. Since they are so important, they are given the name **critical points** or **points of inflection** of a function.

Definition: A point $x_0$ is said to be a critical point or a point of inflection of a function $f(x)$, if the first derivative of the function evaluated at that point, $f'(x_0)$ is zero.

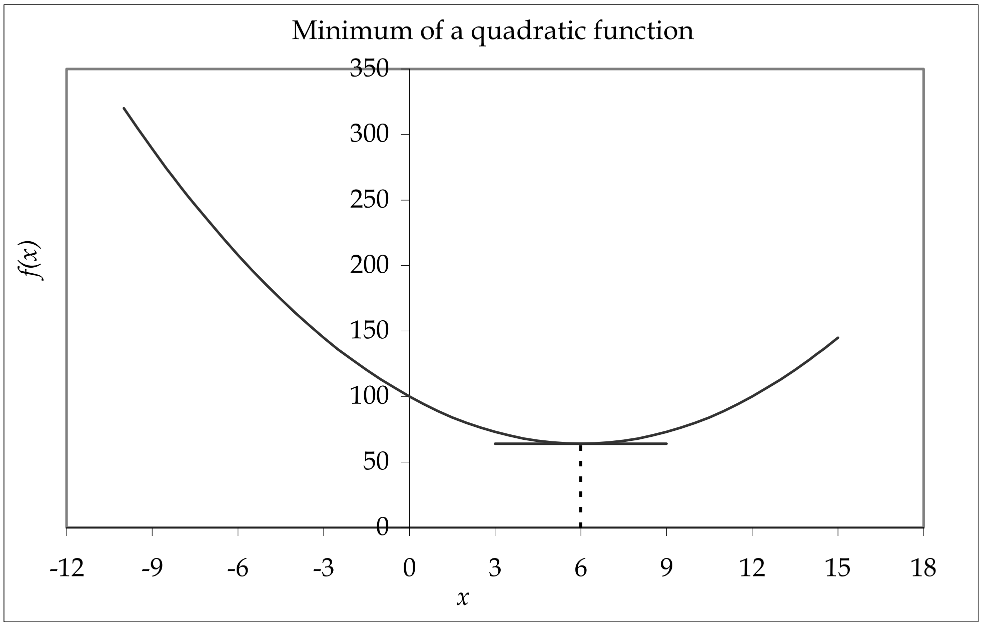
In our example, setting the slope of the function to zero means solving the equation, $f'(x_0) = 12 - 2x_0 = 0$, which yields $x_0 = 6$ as a critical point of the given function. This is the first piece of the optimization puzzle. But this is not the entire story. In this example, the critical point happens to be the point where the function attains its maximum value. But this is not true in general. There are two problems:

Problem A: Without drawing a picture, it is not obvious if the critical point we identified is a maximum or a minimum.

Problem B: There might be more than one critical point for a given function.

Problem A is illustrated with the following graph of another function:

<u>Example 13</u>: $f(x) = -12x + x^2 + 100$




The first derivative of this function is $f'(x) = -12 + 2x$, which when set equal to zero, gives us *x*=6. By our earlier definition, this suffices to claim that *x*=6 is a critical point of the function. But looking at the graph, it is obvious that here, the critical point is a *minimum* of the given function.

So, we now know that a necessary condition for a point to qualify as a minimum or a maximum of a function is that the first derivative of the function evaluated at that point should be zero. But we need to do more work in order to figure out if the critical point is a maximum or a minimum – it turns out that we need the sign of the second derivative.

Remember that if a function has a positive second derivative, i.e. $f''(x) > 0$, it is a convex function (bowl shaped). This means that the critical point of such a function must be a **minimum**. Conversely, if $f''(x) < 0$, i.e. for a concave function that looks like an inverted bowl, the critical point must be a **maximum**. The function in

Example 12 has a second derivative equal to $f''(x) = \frac{d^2 f(x)}{dx^2} = -2 < 0$. Therefore, at $x$=6, we have a *maximum* for that function, as we identified from the graph. By the same logic, for the function in Example 13, the second derivative is $f''(x) = \frac{d^2 f(x)}{dx^2} = 2 > 0$. Therefore, at $x$=6, we have a *minimum* for that function, as can be seen from its graph.

This discussion is usually summarized in math textbooks in the following fashion: For a given function *f(x)* to attain its maximum (minimum) value at a point $x_0$, we need $x_0$ to satisfy:

i) First Order Condition (FOC): $f'(x)\big|_{x=x_0} = 0$

ii) Second Order Condition (SOC):

$$f''(x)\big|_{x=x_0} < 0 \text{ for a max; } f''(x)\big|_{x=x_0} > 0 \text{ for a min} \qquad \ldots \qquad (10)$$

Problem B is tricky to deal with. There can be (and there are) functions with more than one critical point. What then? Consider yet another example.

<u>Example 14</u>: $f(x) = x^4 - 4x^3 + 4x^2 + 4$

Let us try and find critical points of this function. We follow the usual procedure and set the first derivative of this function equal to zero.

$$f'(x) = 4x^3 - 12x^2 + 8x = 0$$

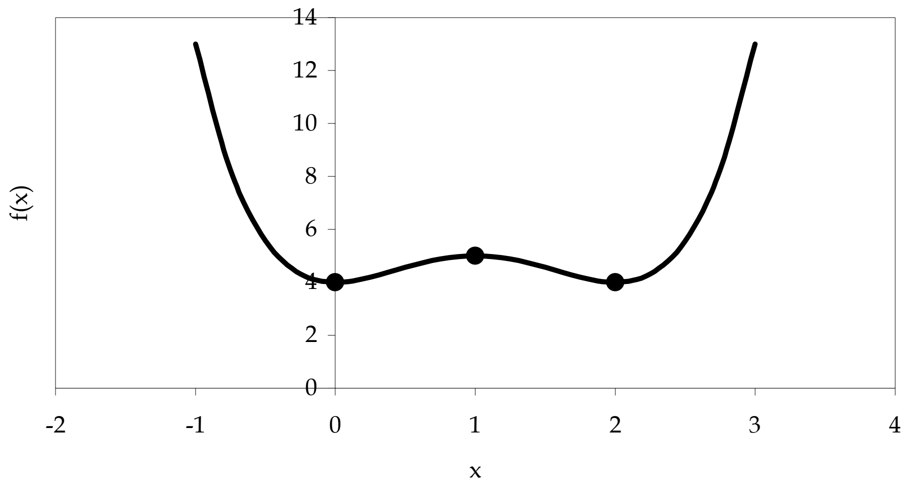
To solve this equation, we note that the above equation can be factorized as follows, which helps us solve the equation for its three roots (since this is a cubic equation).

$$f'(x) = 4x(x-1)(x-2) = 0$$

The three solutions (roots) to this equation are $x$=0, $x$=1, $x$=2.

Which (if any) are the minima, the maxima, and which are neither? At this point, a graph may be useful in seeing the big picture (pardon the pun!).

A Strange Function




The figure confirms what we already know: that *x*=0, *x*=1, *x*=2 are the critical points of the above function. It also gives us a lot more information. *x*=0 and *x*=2 are both minima of the given function. What about the point *x*=1? And how can we tell without looking at the graph?

It is necessary at this point to introduce the idea of a **local optimum** and a **global optimum**. A function has a *local* maximum (minimum) at $x = x_0$ if it achieves its maximum (minimum) value in the neighborhood of all points *x* around $x_0$. The function has a global maximum (minimum) if it achieves its maximum (minimum) value over all points *x* at which it is defined.

Strictly speaking, the FOC and SOC conditions we derived are conditions to obtain local maxima and minima. To see this, let's evaluate the second derivative of this function: $f''(x) = 12x^2 - 24x + 8$ at our three critical points, we have: $f''(0) = 8$, $f''(1) = -4$, $f''(2) = 8$. According to our earlier conditions, we confirm that we have local minima at *x*=0 and *x*=2, and a local maximum at *x*=1. Now, looking at the graph itself, it is obvious that *x*=0 and *x*=2 also yield global mimima, as the function is always of higher value than at these points. But *x*=1 is clearly a local maximum point. At points $x \to \infty$, we can see that the function gets arbitrarily large, and so $x$=1 cannot be a global maximum point.

So how do we find global maximum and minimum points using calculus? The short answer is: it's not easy. At least in economics and finance, we get around this problem by employing functions that have desirable conditions such as:

a) $f$ has only one critical point in its domain

b) $f'' > 0$ or $f'' < 0$ for all values of x at which the function is defined

In other words, we use only well-behaved functions where it is guaranteed that the FOC and SOC (that yield local optimum points) give us a global minimum or maximum.

All the math we showed above for one variable can be extended very simply to the multivariate case. As it turns out, for the FOCs, we have to set all first order *partial* derivatives to zero, and for the SOCs, we have to examine the signs of all the second order *partial* derivatives *including the cross-partials*. In the interest of brevity, I will not deal with the multivariate case here.

- **CONSTRAINED OPTIMIZATION**

All the optimization we dealt with in the previous section is called unconstrained as we tried to find the maximum or minimum of a function (this function is called the **objective function**), without any constraints. But in many economic problems, as in real life, we have several constraints under which we operate. An example will make this distinction clearer.

Example 15: Consider the standard maximization problem of a consumer who gets utility (happiness) from consuming two goods: *pizza* ($p$) and *beer* ($b$). The consumer's utility function is defined as: $U(p,b) = 100.p.b$

Notice that if there were no constraints, the consumer would like to consume as many pizzas and as much beer as he can to maximize his utility. In other words, the consumer's utility increases arbitrarily as $p \to \infty; b \to \infty$. That means the consumer is never satisfied at any level of consumption (by the way, this is called **non-satiation** in economic jargon).

On the other hand if you take the first derivative with respect to the variables *p* and *b*, and set them equal to zero (FOCs), you will see that the critical points are of the form *p*=0 or *b*=0. We know that if *p*=0 or *b*=0, the consumer's utility is zero, which means the FOCs yield us minimum points. So, the question of *maximizing* the consumer's utility is vacuous in an unconstrained world.

However, in the real world, our consumer is faced with constraints. And they make the utility maximization problem interesting. Let us say pizzas sell for $10 apiece and bottles of beer sell for $2 apiece. Further, assume our consumer has a total of $400 with him (where he got this money is not our concern here). So, somehow, he has to adjust his consumption of pizzas and beer such that he gets maximum happiness. These facts can be written in the form of a *budget constraint* as follows:

$$10.p + 2.b = 400$$ [6]

Now the utility maximization problem becomes more interesting. It is usual to write such a problem in the following form:

$$\underset{p,b}{Max}\ 100.p.b$$

$$\text{subject to } 10.p + 2.b = 400$$

This pithy formulation tells us everything we need to know about this problem. The variables under the *Max* are the variables over which we want to maximize the utility function, i.e. *p* and *b* are the **decision variables** in this problem. The utility function is called the **objective function**, and the problem is subject to a **constraint** (in this case, a budget constraint).

[6] Strictly speaking, we should have written: $10.p + 2.b \leq 400$ to allow for the possibility that the consumer does not spend all his money. But in this simple world, we know that the consumer would like to spend all his money, and get more utility. In mathematical jargon, we know that the constraint is **binding**.

More generally, a maximization problem is written as:

$$\begin{aligned} &\underset{x_1, x_2, \cdots x_n}{Max} \; f(x_1, x_2, \cdots x_n) \\ &\text{subject to } g_1(x_1, x_2, \cdots x_n) = 0 \\ &\qquad\qquad g_2(x_1, x_2, \cdots x_n) = 0 \\ &\qquad\qquad \vdots \\ &\qquad\qquad g_m(x_1, x_2, \cdots x_n) = 0 \end{aligned} \qquad \ldots \qquad (11)$$

Here $f(.)$ is the *objective function*, and $g_1(.)\ldots g_m(.)$ are the *m constraints*

So, how do we solve this problem? In this case, we have a very simple problem, so we can solve for $b$ from the constraint as $b = 200 - 5.p$, and plug this into the utility function to give us a function only of one variable, i.e. $p$. In this method, by eliminating $b$, we convert this constrained optimization problem into an unconstrained optimization problem. But we need a more general method, which will work for this case, as well as for more complex cases. This general method which we will study next is called the **Lagrange Multiplier Method**. I will present a cookbook-type recipe for this method, and present the intuition with some pictures.

The Lagrange Multiplier Method

Step 1: Form the so-called **Lagrangian function**, or simply **Lagrangian**:

$$\begin{aligned} \mathcal{L}(x_1, x_2, \cdots x_n, \lambda_1, \lambda_2, \cdots \lambda_m) = {} & f(x_1, x_2, \cdots x_n) - \lambda_1 . g_1(x_1, x_2, \cdots x_n) \\ & - \lambda_2 . g_2(x_1, x_2, \cdots x_n) \cdots - \lambda_m . g_m(x_1, x_2, \cdots x_n) \end{aligned} \qquad \ldots(12)$$

For our utility maximization problem, we have only one constraint, so $m$=1. Hence, the Lagrangian is:

$$\mathcal{L}(p, b, \lambda) = 100pb - \lambda.(10p + 2b - 400)$$

Step 2: Find the critical points of the Lagrangian function. In other words, set *all* the first derivatives of the Lagrangian function to zero:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0 \cdots \frac{\partial \mathcal{L}}{\partial x_n} = 0; \; \frac{\partial \mathcal{L}}{\partial \lambda_1} = 0 \cdots \frac{\partial \mathcal{L}}{\partial \lambda_m} = 0 \qquad \ldots \qquad (13)$$

Note that this is a system of $n$+$m$ equations we need to solve for the $n$+$m$ unknowns $x_1, x_2, \cdots x_n, \lambda_1, \lambda_2, \cdots$ and $\lambda_m$.

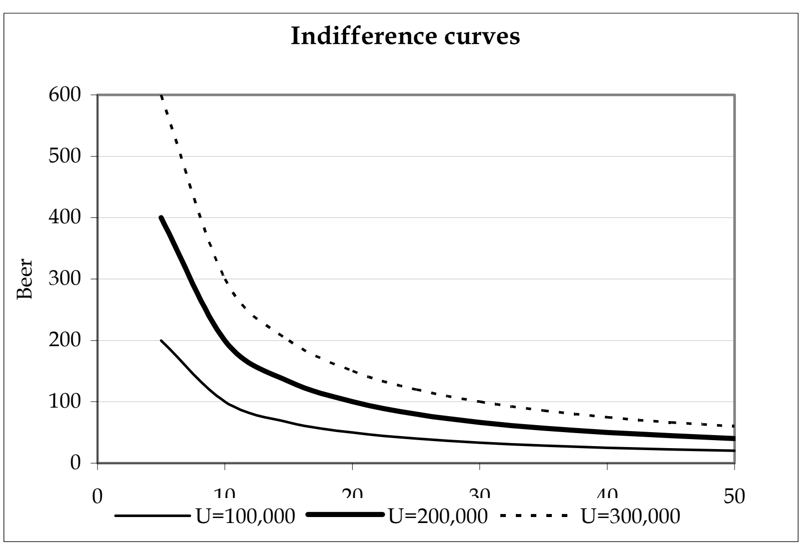
For our problem, we have two variables and one constraint. So we will have 3 equations in 3 unknowns:

$$\frac{\partial \mathcal{L}}{\partial p} = 0 \;\Rightarrow 100b^* - 10\lambda^* = 0$$
$$\frac{\partial \mathcal{L}}{\partial b} = 0 \;\Rightarrow 100p^* - 2\lambda^* = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \;\Rightarrow 10p^* + 2b^* = 400$$

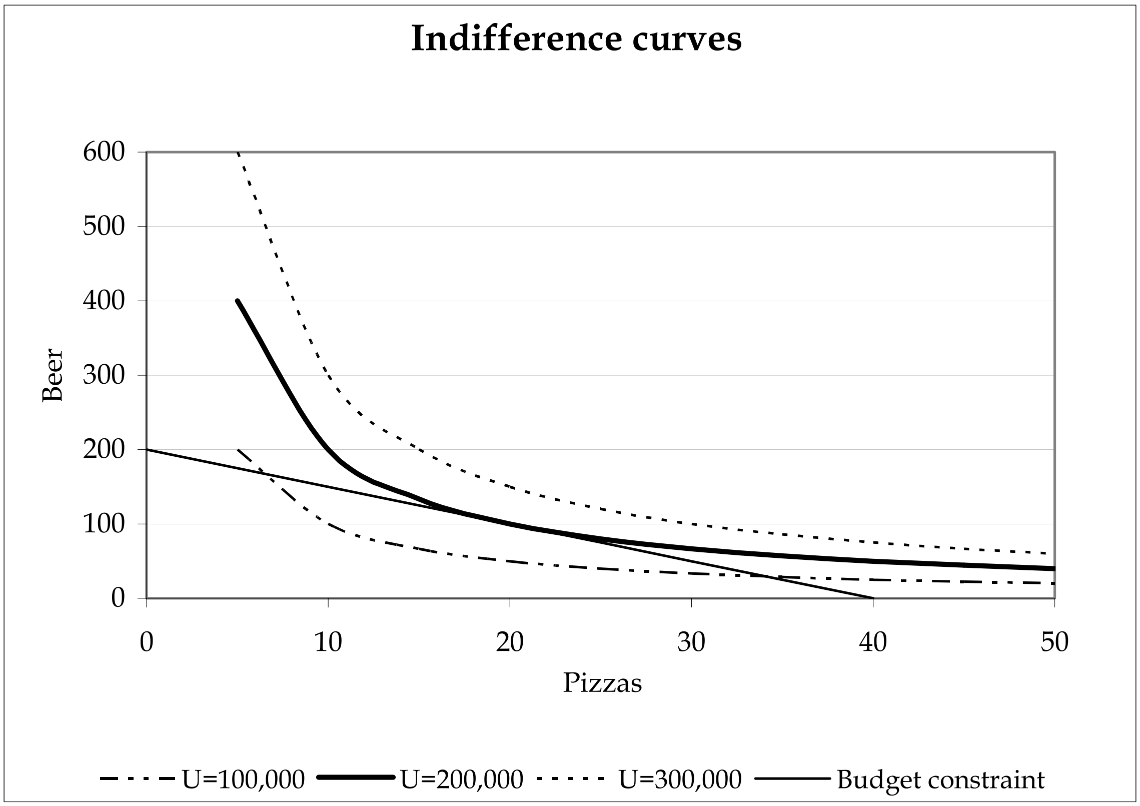
I have superscripted each unknown variable in the above system with an asterisk to indicate that these first order conditions must be satisfied at the optimum consumption levels of pizza, $p^*$ and beer $b^*$.

Step 3: Solve the $n+m$ equations from Step 2. For our example, solving this system of three equations yields $p^* = 20; b^* = 100; \lambda^* = 1000$

Now, we know that the constrained maximum of utility for our consumer is obtained when he chooses to consume 20 pizzas and 100 bottles of beer. It is useful to see all this graphically. Three-dimensional graphs are cumbersome, so we make use of two-dimensional graphs of the utility called **indifference curves**. Along the same indifference curve are combinations of pizzas and beer that give the same utility.



Notice that the higher utility corresponds to a northeastern migration of the indifference curve. Geometrically, we want to be on the most northeastern indifference curve, given our budget constraint. The highest utility achievable given our budget is the maximum value from our utility maximization exercise. This can be seen by superimposing our (linear) budget constraint on the above graph, as I have shown below.




As you can see from the above, the budget constraint is exactly tangent to the maximum achievable utility indifference curve (with U=200,000), at the point where consumption is 20 pizzas and 100 beers, which is the maximum we obtained from the Lagrange Multiplier method! And the maximum utility obtained is given by the utility function: $U(p^*, b^*) = 100.p^*.b^* = 100.20.100 = 200,000 \text{ units}$, as we can see from the graph above.

We need to clear up one tiny issue before leaving the topic of constrained optimization. What is the intuition behind the Lagrange multiplier $\lambda$? We made use of it as a mathematical device as part of the optimization recipe. Can we say something about its economic meaning? In particular, what does it mean to say that, at the optimum, $\lambda^* = 1000$?

It turns out that the Lagrange multiplier in our problem has a neat interpretation as *the marginal effect on the optimal utility of relaxing the budget constraint*. In a general problem, the Lagrange multiplier measures the marginal change in the optimal value of the maximized function due to a small relaxation of the budget constraint. Let us see this in numbers.

Suppose that instead of having $400 in his pocket our consumer had $401. This is tantamount to relaxing the budget constraint. (He is a bit less constrained than he was before because he has $1 more now). Redoing the optimization calculations (and you should do this yourself), we can see that the optimal values of beer and pizza are now: $p^* = 20.05; b^* = 100.25$.

Utility is now $100.p^*.b^* = 100.(20.05).(100.25) = 201,001.25 \text{ units}$, an improvement of 1,001.25 units from before. This is the true *marginal change in optimal utility* due to relaxing the budget constraint by $1. Using the $\lambda^* = 1000$ number from before, we can approximate this as $\Delta \text{U}^* = \lambda^*.(\$1) = 1000 \text{ units}$, which compares very well to the actual marginal increase in utility of 1001.25 units.

- **UNCERTAINTY**

A large class of problems in economics and finance involves uncertainty. This is simply the *Forrest Gump* idea that "Life is like a box of chocolates. You never know what you are going to get next". We use the concept of **probability** to denote the degree of uncertainty, as in "There is a 60% probability that it will rain tomorrow".

Since probability theory is used extensively in financial economics, we study it here in a special section.

Probability

In our earlier language, **probability** can be looked at as a function that assigns to subsets of an **event space** a number between 0 and 1 called the probability of that subset's occurrence, i.e. $P:\Omega \mapsto [0,1]$. If this sounds mystifying, let us look at a couple of simple examples.

Example 16a: Consider tossing a fair coin. The event space consists of all possible outcomes of the experiment: here Ω={Head, Tail}. Now the probability function assigns to each element and subset of the event space, some number between 0 and 1. For instance, P{outcome=Head}=1/2; P{outcome=Tail}=1/2; P{outcome = Head *or* tail}=1, P{outcome=Head *and* Tail}=0.

Example 16b: Consider now rolling a fair die. Now the event space is larger: Ω={1,2,3,4,5,6}. Again, the probability function assigns to each element and subset of the event space, some number between 0 and 1.

Here, P{outcome=1}=1/6 …P{outcome=6}=1/6

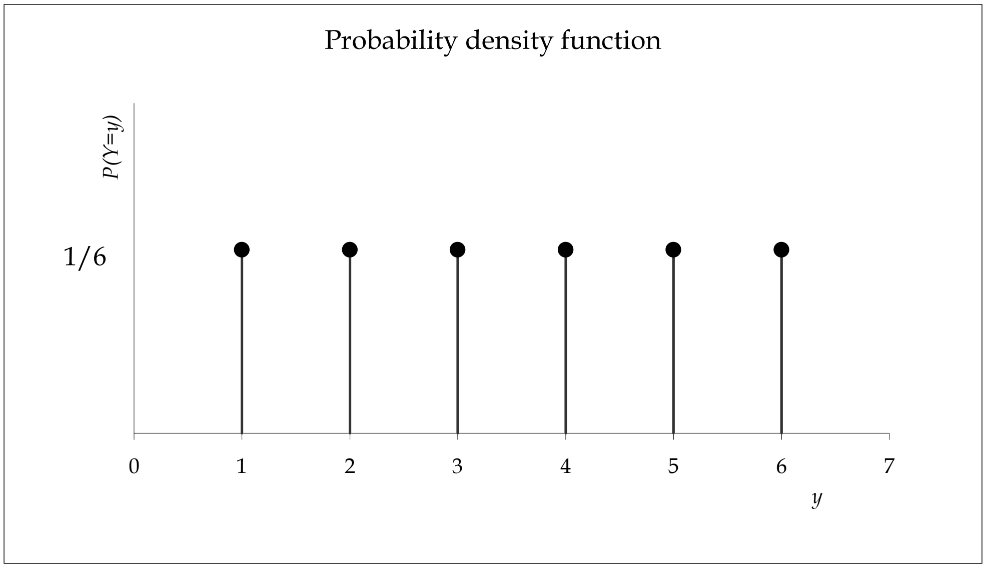
Random Variables

In the die-rolling experiment above, before the die is rolled, the outcome is said to be **random** as the outcome is uncertain. Let's use $Y$ to denote the random outcome. Then, $Y$ is said to be a **random variable** (hereafter, r.v.). Thus, by definition, the random variable could take on one of several values. The probability summarizes our beliefs regarding the different values the r.v. could take on.

Discrete random variables: Density and distribution functions

Let's continue the example of the die-rolling experiment. Such a random variable which can take on only a discrete set of values (here 1, 2, 3, 4, 5 and 6) is called a **discrete random variable**.

If we plot the different values of the r.v. $Y$ could take on the horizontal axis and the probability that the r.v. will take each value on the vertical axis, we obtain the following plot:

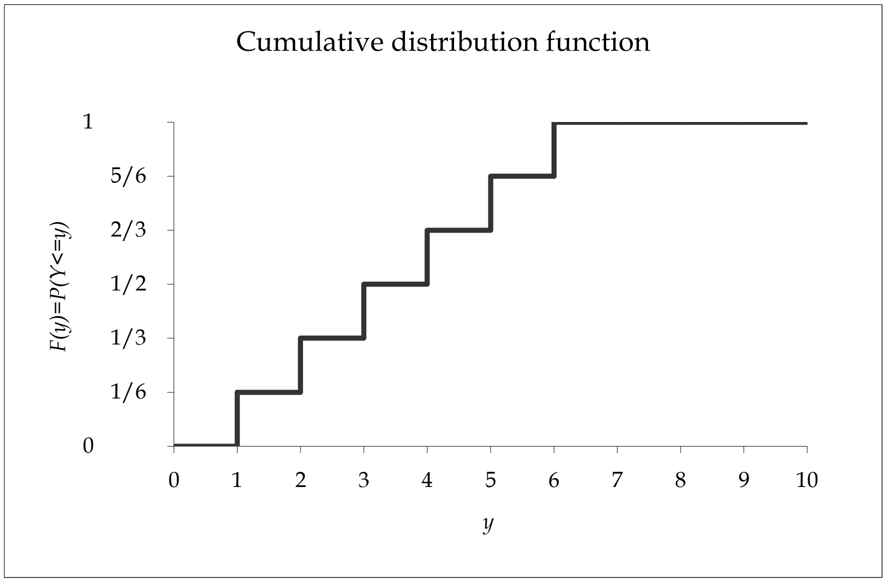



The obvious thing this plot tells us is that the probability that $Y$ could take the values 1, 2, 3, 4, 5 or 6 is 1/6 each. Note two other things carefully:

1) The probability of $Y$ taking values such as $y$=1.5 or $y$=2.8 is *zero*. This is another way of saying that when you roll a die, 1.5 is not an option.
2) When you add up the probabilities of $Y$ taking on each possible value (i.e. 1 thru 6), you get 1. So, it is as if in the above graph, we have taken a lump of probability of quantity 1, and distributed it among the 6 possible points.

The above plot is a picture of the **probability density function (pdf)** of the random variable Y. In other words, for any r.v., the pdf informs us about the concentration or density of probability at various values that the r.v. can take.

The **distribution function,** also known as the **cumulative distribution function (cdf),** of an r.v. is another handy device. It is defined as $F_Y(y) = P(Y \leq y)$. The cdf asks the question: given a point $y$, what is the probability that the r.v. $Y$ takes on values less than or equal to $y$? For our die-rolling example above, the distribution function looks like the following:




Note a few things about cdf s here:

1) A lot of times we write $F_Y(y) = P(Y \le y)$ as simply $F(y)$, assuming that it is known which r.v. we are talking about.
2) $F(y)$ is an increasing function
3) $F(y) \to 0$ as $y \to -\infty$; $F(y) \to 1$ as $y \to \infty$
4) $P(a < Y \le b) = P(Y \le b) - P(Y \le a) = F(b) - F(a)$. For example, if we wanted to find the probability that our r.v. $Y$ takes a value between 2 and 4 (not including 2), we would find it as: $P(2 < Y \le 4) = F(4) - F(2) = 2/3 - 1/3 = 1/3$. This makes sense: In our simple experiment, $Y$ taking a value of between 2 and 4, not including 2, means the possibilities $y$=3 and $y$=4. Indeed the probability of these two values is $1/6 + 1/6 = 1/3$.

There are many useful discrete r.v.s, and we shall study one more in the next example.

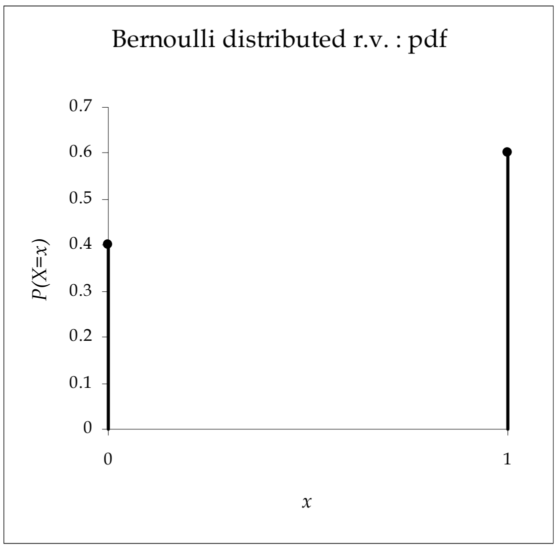
Example 17: A Bernoulli distributed r.v.

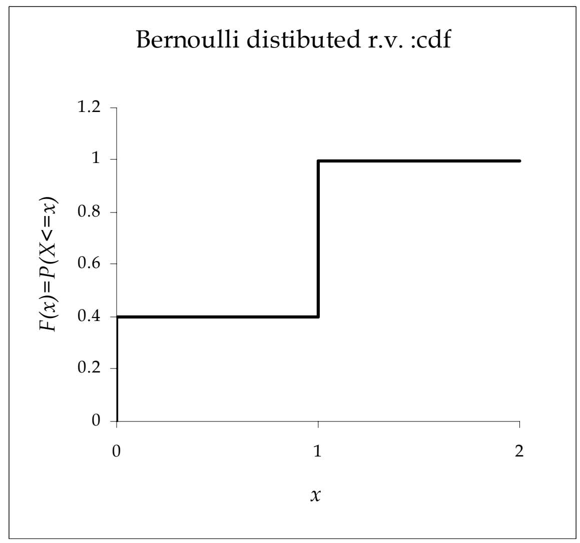
A r.v. $X$ is said to have the **Bernoulli distribution** with parameter $\pi$ if:

1) $X$ can take only one of two values: 0 or 1

2) $P(X=1)= \pi$; $P(X=0)=1-\pi$

I have shown *density* and *distribution* function plots for a Bernoulli distributed r.v. with $\pi=0.6$ as shown below:




Continuous random variables: Density and distribution functions

Not all random variables are discrete. Many are continuous, i.e. can take on a continuum of real values. For instance, the temperature in this room a minute later could be viewed as a continuous r.v., as it could take values of 55.1 degrees, 55.15 degrees … and so on. (We are limited only by our thermometers in measurement).

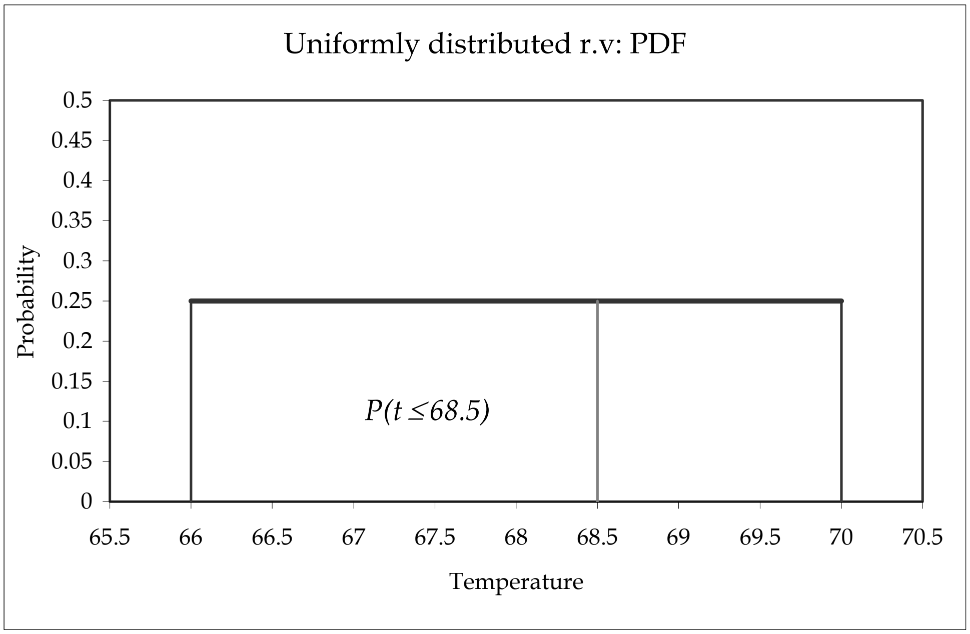
Next, we generalize the concepts of pdf and cdf from the discrete to the continuous case, and note some differences with the discrete case.

Example 18: A uniformly distributed r.v.

Let us say the folks in the Freeman School make a good effort to keep the temperature in this room at 68°F. However, due to day-to-day conditions, let us say the temperature varies between 66°F and 70°F. Note that the temperature can be considered as a continuous r.v., say $T$ – a *random* variable, as we do not know exactly which value it will be tomorrow, and *continuous* because any value between 66°F and 70°F is possible. This can be represented by the following *probability density function*:

$$f(t) = \begin{cases} 1/4, & 66 \le t \le 70 \\ 0 & \text{elsewhere} \end{cases}$$

Here is a graph of the above function.




This is obviously a continuously distributed r.v., as the graph above indicates. It is defined only over the range [66,70]. Note a couple of things here:

1) It does not make any sense to ask the question: What is the probability that the temperature will be exactly 68.5°F? Indeed, *the probability of a continuous r.v. taking on any point value is zero*. This is a significant difference between discrete and continuous r.v.s. In the discrete case, the probability of the r.v. taking on a point value may be non-zero. For instance, in the die-rolling experiment, the probability that the outcome is 3 is 1/6, while the probability that the outcome is 3.25 is 0.
2) With continuous r.v.s, we can only ask questions like: What is the probability that the temperature is less than or equal to 68.5°F? Or, what is the probability that the temperature is greater than 68.5°F? This probability is given by the area

under the graph in the specified range. For example, to determine the probability of *t* below 68.5, we need to consider the area below the graph, and to the left of the point 68.5. It is particularly easy to find the value in this simple case: this area is a rectangle, whose area is the length times width = (68.5-66)×0.25=0.625, or 62.5%. Conversely, the probability that the temperature is greater than 68.5 is (70-68.5)×0.25=0.375, or 37.5%

3) Now, we know that the probability that *t* will be between 66 and 70 is 1, since this is the entire range of possible values for *t*. This means that the total area under the graph, between 66 and 70 has to add up to 1! Is this true? Sure enough, (70-66)×0.25=1.0, as we surmised. How did we achieve this? Because, the height of the rectangle was cleverly set at 0.25.

4) In general, a uniformly distributed r.v. can be defined over any interval [*a*,*b*], in which case, we write the pdf as:

$$f(t) = \begin{cases} 1/(b-a), & a \le t \le b \\ 0 & \text{elsewhere} \end{cases}$$

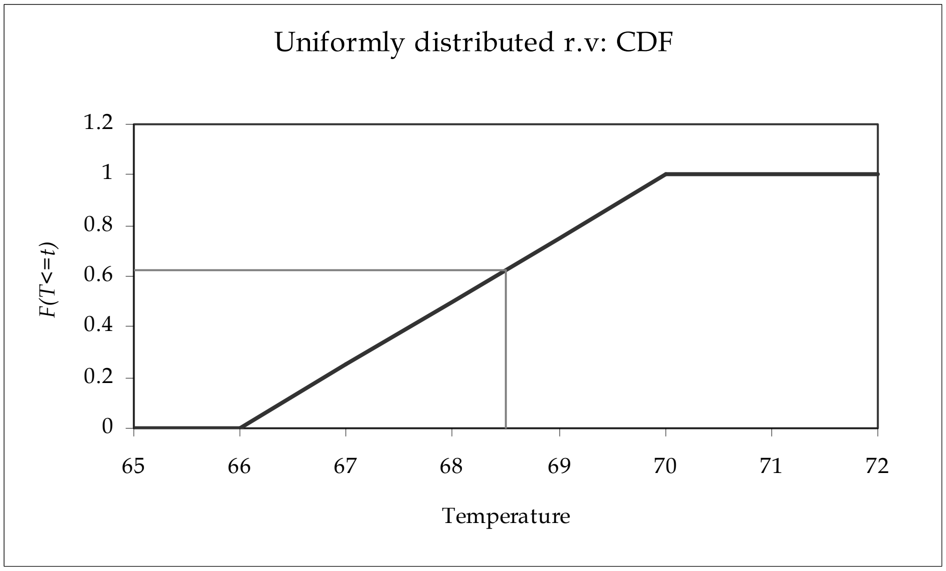
This ensures that the total probability under the graph is $\frac{1}{(b-a)} \times (b-a) = 1$.

What about the cdf of this uniform r.v.? Recall that the cdf is defined as $F_Y(y) = P(Y \le y)$, and is the answer to the question: given a point *y*, what is the probability that the r.v. *Y* takes on values less than or equal to *y*? We have just learned that the probability that a continuous random variable *Y* takes a value below a certain *y* can be found as the area under the probability density function graph to the left of the point *y*. In the language of calculus, summation of areas under a graph is accomplished by the technique of **integration** (which is the opposite of differentiation, which we looked at before). Hence, you will see the following definition in a text book on probability.

$$F(Y \le y) = \int_{-\infty}^{y} f(y)dy \qquad \dots \qquad (14)$$

This formula merely says that if we sum up all the area under the curve *f(y)* between -∞ and the point *y*, you will obtain the cumulative probability that the continuous r.v. will take a value below *y*.

For our temperature random variable *T*, we can find this probability for every point *t*, and plot this in the form of the following graph.




Note again that, as with discrete r.v.s:

1) $F(.)$ is an increasing function
2) $F(t) \to 0$ as $t \to -\infty$; $F(t) \to 1$ as $t \to \infty$
3) $P(a < T \leq b) = P(T \leq b) - P(T \leq a) = F(b) - F(a)$. For example, if we wanted to find the probability that our r.v. T takes values below 68.5, we simply read it off this graph as 0.625, as shown in the above graph.

Finally, note that the *continuous uniform* r.v. in Example 18 is the continuous counterpart to the *discrete uniform* r.v. of Example 16b (the die-rolling experiment).

- **Expectation and Variance**

Statisticians are frequently interested in finding the **expectation** and **variance** of a random variable. Let us study these two quantities in some detail.

Formally, the **expectation** or simply, **mean** or **average** of a random variable is defined as: $E(X)=\int_{-\infty}^{\infty} x.f(x)dx$ for a continuous r.v. … (15a)

and the corresponding definition:

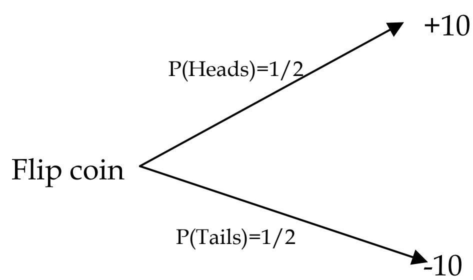
$$E(X)=\sum_x x.P(X=x) \text{ for a discrete r.v.} \quad \ldots \quad (15b)$$

It is easiest to understand this quantity in the context of a simple discrete r.v.

Example 19: A simple gamble

Say you are faced with the following gamble. You flip a coin and if Heads shows up, you get \$10, and if Tails shows up, you lose \$10. This can be summarized very effectively by the following **tree**.

**Gamble 1:**



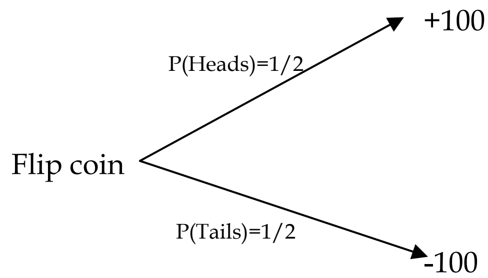

The outcome of this gamble is a discrete random variable, say X, which takes values +10 and -10 with probability of ½ each. Now, the expectation of this random variable, using equation (15b) is:

$$E(X)=\left(\frac{1}{2}\times 10\right)+\left(\frac{1}{2}\times -10\right)=\$0$$

In other words, the *mean* or *average* of this gamble (also known as the **expected value**) is simply the probability weighted average of all possible outcomes. What does it mean to say that the expected value of this gamble is zero? It means that if you play this game a (very large) number of times, you *expect* to have no money in your pocket at the very end. In one trial, you will either win or lose, but after a million trials, given that you are flipping a fair coin, you expect to win \$10 in about half the trials and lose \$10 about half the time, leaving you with nothing at the end of it all.

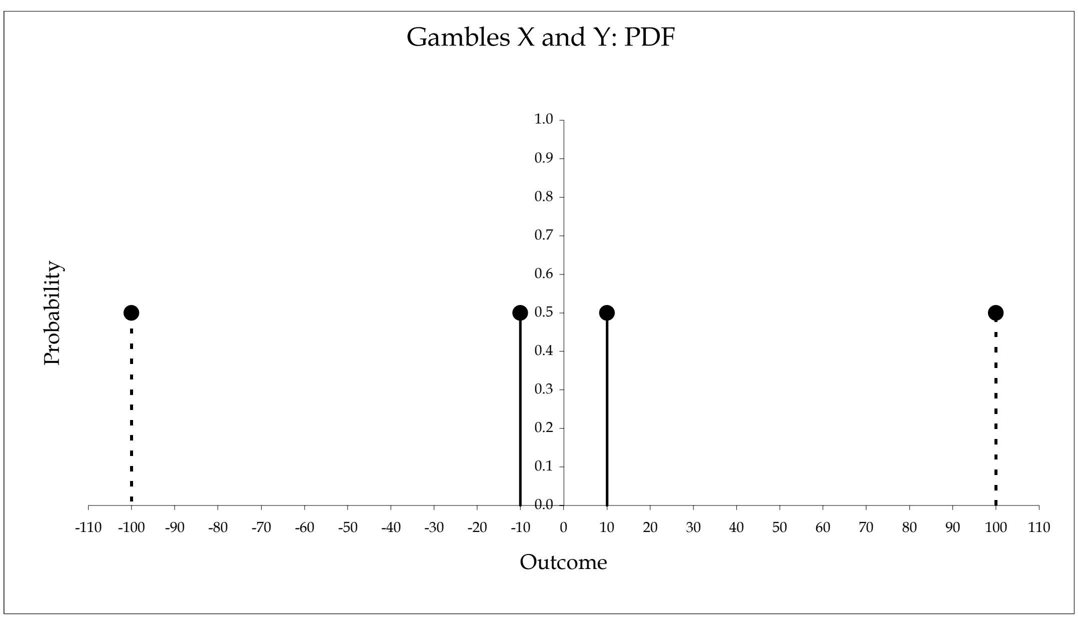
Now, consider another similar gamble:

**Gamble 2:**




Let us denote the outcome of this gamble by a random variable $Y$. One can verify easily that the expected value of this gamble is also zero. In other words, $E(Y)=0$

Suppose you are a player at a casino offering these two gambles. You ask a simple question. Which of these two gambles is *more risky*? Well, the expected value of the two gambles is identical. Does that mean each are equally risky? Clearly not! Intuitively speaking, we sense that Gamble 2 is somehow riskier than Gamble 1, as you gain or lose a lot more in the former as compared to the latter. At this point, it may help to look at the pdf s of the r.v.s associated with both these gambles.




The dark lines are for Gamble 1 (r.v. $X$), and dashed lines are for Gamble 2 (r.v. $Y$)

One can see that our intuitive feeling about the risk of these gambles, the notion of gaining or losing a lot more than in the other gamble is represented as Gamble 2's pdf being much more "spread out " about the average (zero, in this case). We need a quantity that will tell us how "spread out" the pdf of a given r.v. is about its average value. It turns out this quantity is the *variance* of a random variable.

Formally, the **variance** of a random variable is defined as:

$$Var(X) = \int_{-\infty}^{\infty} \left[x - E(X)\right]^2 .f(x)dx \text{ for a continuous r.v.} \quad \ldots \quad (16a)$$

and the corresponding definition:

$$Var(X) = \sum_x \left[x - E(X)\right]^2 .P(X = x) \text{ for a discrete r.v.} \quad \ldots \quad (16b)$$

Let us apply this definition to both our gambles:

For Gamble 1, denoted by random variable $X$:

$E(X)=0$

$$Var(X) = \left[\frac{1}{2}(10-0)^2\right] + \left[\frac{1}{2}(-10-0)^2\right] = 100 \text{ squared dollars}$$

For Gamble 2, denoted by random variable $Y$:

$E(Y)=0$

$$Var(Y) = \left[\frac{1}{2}(100-0)^2\right] + \left[\frac{1}{2}(-100-0)^2\right] = 10{,}000 \text{ squared dollars}$$

By this measure, Gamble 2 is 100 times riskier than Gamble 1, which squares with our intuition. One can immediately see the use of this quantity in finance and economics. *Variance gives us a way of comparing the risk of different investments*. When people say stocks are riskier than bonds, they usually mean that the variance of stock returns is likely to be higher than the variance of bond returns. In other words, stock returns are more **volatile** than bond returns.

The only problem with variance is its units. Since we squared the deviations from the average during the variance calculation, we come up with a quantity measured in squared dollars. Clearly, squared dollars are cumbersome to work with and we need a workaround. Why not take the square root of the variance? Doing this gives us a quantity called the *standard deviation*.

Formally, the **standard deviation** of a random variable, discrete or continuous, is defined as:

$$\text{Standard Deviation}(X) = \sqrt{Variance} \quad \ldots \quad (17)$$

For our example gambles:

Gamble 1: Standard deviation($X$)= $\sqrt{100} = \$10$

Gamble 2: Standard deviation($Y$)= $\sqrt{10,000} = \$100$

Even though we have worked through only the simplest of examples, the expectation, variance and standard deviation of more complex discrete r.v.s, and indeed, continuous r.v.s can be found using formulas (15), (16) and (17) quite easily. The interpretation of expected value as the average, and variance and standard deviation as the volatility of a random variable is true for any r.v. in general.

We are now ready to get introduced to a very important distribution: The Normal Distribution, or the so-called Bell curve. This is far and away the most important distribution in probability and statistics, so read the next section well.

- **THE NORMAL DISTRIBUTION**

A (continuous) random variable X is said to be normally distributed with parameters $\mu$ and $\sigma$ if it has the following pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in (-\infty, +\infty) \quad \ldots \quad (18)$$

Do not let this complicated formula intimidate you; rest assured we will *almost never* manipulate such complicated functions.

The corresponding cdf is obviously given by

$$F(X \le x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad \ldots \quad (19)$$

Mathematicians have known at least for a couple of hundred years now that the integral in (19) has no **closed-form** solution, which means we cannot actually express it in the form of a function of $x$, $\mu$ and $\sigma$. But this does not mean it is useless. We can evaluate it at any point numerically. Tables of cumulative probability for different points $x$ under the Normal distribution are usually found in the back of any statistics text book.

What are the parameters $\mu$ and $\sigma$? It turns out that $\mu$ is the *expected value* or *average* or *mean* of this r.v., and $\sigma$ is the *standard deviation* or square root of the *variance* of this r.v.

Often, we will find it much more convenient to work with a special normal distributed r.v. with zero mean, and unit variance, i.e. with $\mu = 0$ and $\sigma = 1$. Such an r.v. is said to be distributed according to a **Standard Normal Distribution**.

Formally, a standard normal r.v. $Z$ has the following pdf:

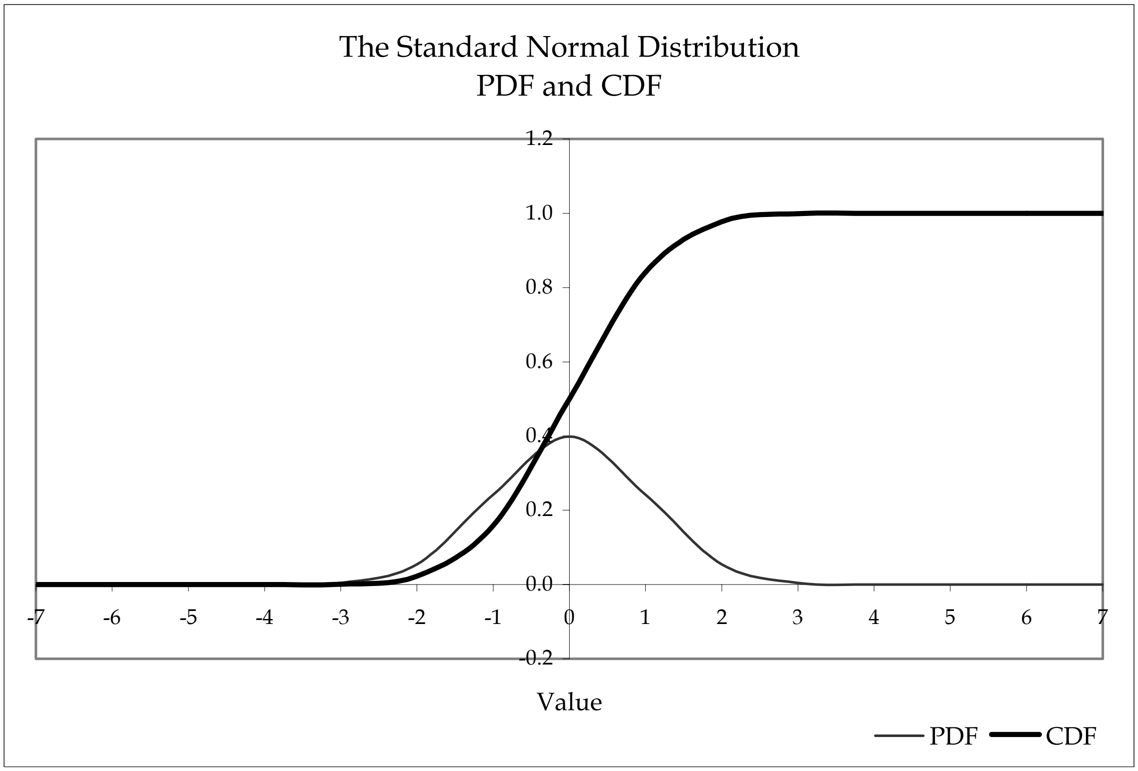
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, z \in (-\infty, +\infty) \quad \ldots \quad (20)$$

and its cdf is given by:

$$\Phi(Z \le z) = \int_{-\infty}^{z} \phi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \quad \ldots \quad (21)$$

The function $\Phi(Z \le z)$ is sometimes written as $\Phi(z)$ or $N(z)$, and plays an especially important role. The Black-Scholes formula for option pricing, you will remember, contains a couple of terms involving this function.

The next page shows graphs of the pdf and cdf of the standard normal distribution, i.e. equations (20) and (21) above, on the same plot. As expected, the pdf is the familiar bell-shaped curve, while the cdf goes smoothly from 0 to 1 as we traverse values from -∞ though +∞.




- **BASIC FINANCE MATH**

Let's switch gears a bit and explore the mathematics of finance, using the basic math we learned in earlier sections.

<u>Compound Interest</u>

All financial calculations involve the *time value of money*. The basic idea of interest is very simple, and quite familiar. If you start with \$100 in your bank account that pays 5% a year, and do not withdraw any amount during the year, then at the end of one year, you will have your original \$100, called the **principal** amount, plus **interest** of $0.05\times$ \$100 = \$5, i.e. you will have a total of \$105. If you do not withdraw any money in the second year, you will have at the end of the second year, an amount equal to \$105 + \$105 $\times 0.05$= \$110.25. In the second year, you earned interest of \$110.25-\$105=\$5.25. Why \$5.25 and not \$5.00 as in the first year? The interest rate is still the same (5%), but you earned interest on a starting amount of \$105, not \$100.

To summarize, in our simple example, if we denote principal as *P* and the yearly rate of interest as *r*, at the end of *t* years, you will have a **future value** of :

$$\boxed{FV = P(1+r)^t} \quad \ldots \quad (22)$$

This is the most basic formula in finance, and is called the **compounding** formula.

<u>Compounding at various intervals</u>

In the above example, we assumed that your bank pays interest only once every year. Now, your bank might decide (as most banks do) to calculate and pay you interest every quarter. In the jargon of finance, you now have an account on which interest is compounded every quarter. The **compounding interval** is 3 months, and the **compounding frequency** is 4 times per year. What do we do now? Note that the **annual interest rate** (a.k.a. the **annual percentage rate** or **APR**) is still the same: 5%. Now, over the first year your money will grow four times as shown in the table below.

| Period (quarter) | Annual rate | Periodic rate | Begin Amount | Interest | End Amount |
|---|---|---|---|---|---|
| 1 | 5% | 1.25% | 100.00 | 1.25 | 101.25 |
| 2 | 5% | 1.25% | 101.25 | 1.27 | 102.52 |
| 3 | 5% | 1.25% | 102.52 | 1.28 | 103.80 |
| 4 | 5% | 1.25% | 103.80 | 1.30 | **105.09** |

Notice that you end up with $105.09, slightly more than the $105 you would if the money were compounded annually. This makes sense: *As money is compounded more frequently, you earn more interest, which leads to a greater future value*.

What about our formula? Obviously, we need to modify it to account for more frequent compounding. Let us introduce one more variable, *m* to represent the number of compounding periods *per year*. Then, formula (22) becomes:

$$\boxed{FV = P\left(1+\frac{r}{m}\right)^{t.m}} \quad \ldots \quad (23)$$

Let us see what we will have at the end of two years, with quarterly compounding. Plug in $P$=\$100, $r$=0.05 per year, $t$=2 years, $m$=4. Formula (23) gives us the answer: \$110.45, which we naturally expect to be greater than the \$110.25 under **annual compounding**.

Continuous compounding

One can imagine dividing the year into smaller and smaller periods, thereby compounding monthly, weekly, daily, hourly, or even every minute and second. As we divide the year into smaller and smaller intervals, $m$ grows larger and larger. For example, daily compounding implies $m$=360, compounding every minute implies $m$=360×24×60=518400. In the limit, we could imagine compounding every instant, as $m$ grows to a very, very big number. This is the idea of **continuous compounding**. To determine the effect of continuous compounding on formula (23), we need to quantify what we mean by "$m$ grows to a very, very big number". Mathematically, this is accomplished by finding the **limit** of the expression in (23) as "$m$ goes to infinity", and is represented as:

$$FV \text{ (continuous compounding)} = \lim_{m\to\infty}\left[P\left(1+\frac{r}{m}\right)^{t.m}\right]$$

How do we evaluate this? A bit of basic calculus comes to our rescue. We know that as m grows large, the following is true:

$\lim_{m\to\infty}\left(1+\frac{r}{m}\right)^{m} = e^{r}$, where $e$=2.71828…, the **base of the natural logarithm**

This means our expression in (23) becomes:

$$FV = \lim_{m\to\infty}\left[P\left(1+\frac{r}{m}\right)^{t.m}\right] = P\left[\lim_{m\to\infty}\left(1+\frac{r}{m}\right)^{m}\right]^{t}$$, which means

$$\boxed{FV \text{ (continuous compounding)} = Pe^{rt}} \quad \ldots \quad (24)$$

How good is this approximation? In the table below, I have used our example numbers: $P$=\$100, $r$=5% per year, $t$=2 years. I have shown future values after 5, 10, 25, and 50 years, with $m$=1, 2, 360 and ∞.

| Years | m=1 | m=2 | m=360 | m→∞ |
|---|---|---|---|---|
| 0 | 100.0000 | 100.0000 | 100.0000 | 100.0000 |
| 5 | 127.6282 | 128.0085 | 128.4003 | 128.4025 |
| 10 | 162.8895 | 163.8616 | 164.8664 | 164.8721 |
| 25 | 338.6355 | 343.7109 | 349.0040 | 349.0343 |
| 50 | 1146.7400 | 1181.3716 | 1218.0379 | 1218.2494 |

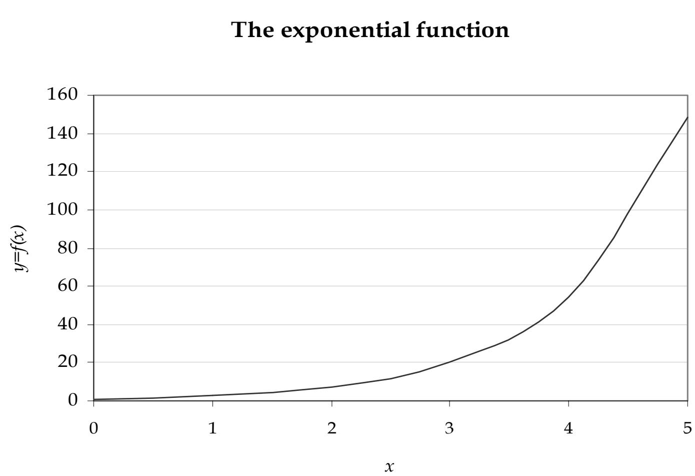
You can see that beyond m=360 is quite close to continuous compounding. (Credit cards compound daily the interest on your balances. Beware!).

- **THE EXPONENTIAL AND LOGARITHMIC FUNCTIONS**

As can be imagined from the discussion in the previous section, the exponential and logarithmic functions are used extensively in finance. Hence, it behooves us to understand these well. This section will also serve to collect all necessary properties of these functions.

Exponential function

Example 6a introduced us to the shape of the exponential function defined as: $y = e^x$ , which looks like the following:



The exponential function has the following five basic properties, which we will use at various times:

E.1: $e^r.e^s = e^{r+s}$

E.2: $e^{-r} = \dfrac{1}{e^r}$

E.3: $\dfrac{e^r}{e^s} = e^{r-s}$

E.4: $(e^r)^s = e^{rs}$, and

E.5: $e^0 = 1$

These are all algebraic properties of exponents that work for any number, not just *e*. For example, they also work for 10.

$$10^2.10^3 = 10^5; 10^{-2} = \frac{1}{10^2}; \frac{10^2}{10^3} = 10^{2-3}; (10^2)^3 = 10^6; 10^0 = 1$$

<u>Logarithm function</u>

Let's start with logarithms to the base 10 (since we are used to counting in tens).

*<u>Definition</u>: The logarithm of any number x to the base 10 is defined as the power to which one must raise 10 to yield x.*

For example, take 100. We are looking for the number *y* such that $10^y = 100$. Obviously, *y*=2 satisfies this equation. So we say: the logarithm of 100 to the base 10 is 2. Similarly the logarithm of 1000 to the base 10 is 3.

Let's now deal with a couple of special cases:

*x*=1: We are looking for a number *y* such that $10^y = 1$. Obviously, *y*=0.

*x*=0: We are looking for a number *y* such that $10^y = 0$. This is not so easy. One can see however, that as *y* becomes a really really large negative number, say -1,000,000, the value of $10^{-1,000,000} = \dfrac{1}{10^{1,000,000}} \to 0$ (Note that we have made use of property E.2. from above). So it makes sense to define the logarithm of 0 to any base as -∞.

To understand the use of logarithms, consider the well-known **Richter scale**[7] used for measuring the magnitude of earthquakes. Although the scale has no upper bound, it is typical to measure an earthquake on the scale by assigning a number between 1 and 9. The Richter scale is *logarithmic* (with base 10), which means the seismic waves of a magnitude 6 earthquake are ten times greater than those of a magnitude 5 earthquake.

Similar to base 10 logarithms, we can take logarithms to the base *e*. Conceptually, this is identical to base 10 logarithms, except that we use *e* in the place of 10, i.e. *the logarithm of any number x to the base e is defined as the power to which one must raise e to yield x, and is written as y=ln(x)*[8]. This means that if we can write $e^y = x$, then *y* is said to be the logarithm of *x* to the base *e*.

Notice that *the logarithm function is the inverse function of the exponential function*. That is, if we raise *e* to the *y* th power, we obtain *x*. If we take the logarithm of *x* to the base *e*, we can get back to *y*. Corresponding to properties E.1. through E.5 of the exponential function, we have five properties of the natural logarithm function.

L.1: $\ln(r.s) = \ln(r) + \ln(s)$

L.2: $\ln(1/s) = -\ln(s)$

L.3.: $\ln(r/s) = \ln(r) - \ln(s)$

L.4: $\ln(r^s) = s.\ln(r)$

L.5: $\ln(1) = 0$

To understand logarithms better, it is worthwhile for you to prove properties L.1 through L.5. using properties E.1 through E.5.

---

[7] So named after Charles F. Richter (1900-1985), renowned American geophysicist and seismologist, who devised the scale.

[8] *ln*(*x*) is read as *natural logarithm of x* or *logarithm of x to the base e*, while *log*(*x*) usually stands for logarithm of *x* to the base 10.

We shall understand the advantages of using logarithms in finance in the next section.

- **MEASURING RETURN**

Consider any financial asset, say a share of stock. We are frequently interested in measuring the return on this asset. If the price of an asset is $P_0$ at time $t$=0, and $P_1$ at time $t$=1, the familiar definition of return during this period (between $t$=0 and $t$=1) is:

$$R_1 = \frac{P_1 - P_0}{P_0} = \frac{P_1}{P_0} - 1 \qquad \ldots \qquad (25)$$

In general, between any time t-1 and t, this return can be written as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1 \qquad \ldots \qquad (26)$$

This is known as the **simple net return** between $t$-1 and $t$.

The **simple gross return** is defined as $1+R_t$.

Similarly, the **return over the most recent k periods** (a multiperiod return) can be written as:

$$\begin{aligned} 1 + R_t(k) &= (1+R_t).(1+R_{t-1}).\cdots(1+R_{t-k+1}) \\ &= \frac{P_t}{P_{t-1}}.\frac{P_{t-1}}{P_{t-2}}.\cdots.\frac{P_{t-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}} \end{aligned} \qquad \ldots \qquad (27)$$

Finally, the **annualized return** over the most recent k periods is written as:

$$\begin{aligned} \text{Annualized } R_t(k) &= (1+R_t(k))^{1/k} - 1 \\ &= [(1+R_t).(1+R_{t-1}).\cdots(1+R_{t-k+1})]^{1/k} - 1 \end{aligned} \qquad \ldots \qquad (28)$$

As we discussed a couple of sections ago, the continuously compounded rate of return between time $t$-1 and $t$ is given by $r_t$, where $r_t$ satisfies:

$$P_t = P_{t-1}.e^{r_t} \qquad \ldots \qquad (29)$$

Using our knowledge of logarithms, we see that we can solve for $r_t$ as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(1+R_t) \qquad \ldots \qquad (30)$$

What about multiperiod returns? The continuously compounded return over the most recent k periods is given by:

$$\begin{aligned} r_t(k) = \ln[1+R_t(k)] &= \ln[(1+R_t).(1+R_{t-1}).\cdots(1+R_{t-k+1})] \\ &= \ln(1+R_t)+\ln(1+R_{t-1})+\cdots+\ln(1+R_{t-k+1}) \qquad \ldots \qquad (31) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$

Finally, the annualized continuously compounded return over the k-periods can be obtained as the simple average of the compounded returns for each period.

Continuously compounded annualized return is:

$$\ln[1+R_t(k)]^{1/k} = \frac{r_t + r_{t-1} + \cdots + r_{t-k+1}}{k} \qquad \ldots \qquad (32)$$

Equations (31) and (32) show the convenience of continuously compounded returns: The continuously compounded return over any number of periods is simply the sum of the continuously compounded returns for each period. *A multiplicative formula is converted to a convenient additive formula!* As an additional advantage, it turns out that at a more advanced level of finance, additive processes are much easier to deal with compared to multiplicative processes.

Another advantage of continuously compounded returns can be explained by the following example:

<u>Example 20</u>[9]: On March 4, 1999, the NASDAQ composite index closed at 2292.89. On March 10, 2000, the index closed at 5048.62. On January 2, 2001, the index closed at 2291.86, essentially at the same level as in March 1999.

The simple net return between March 1999 and March 2000 can be calculated using the formula above to be: $\frac{5048.62}{2292.89} - 1 = 120.19\%$, and the simple net return between March 2000 and January 2001 is: $\frac{2291.86}{5048.62} - 1 = -54.60\%$

[9] This example is from "Derivatives Markets" by Robert L. MacDonald, Addison Wesley, 2003

The thing to note here is that even though the index went up and came back down to the same level, the increase and decrease are *not symmetric*.

Observe the contrast with continuously compounded returns below:
The continuously compounded return between March 1999 and March 2000 is given by: $\ln\left(\frac{5048.62}{2292.89}\right) = 78.93\%$ and that between March 2000 and January 2001 is given by: $\ln\left(\frac{2291.86}{5048.62}\right) = -78.97\%$. Now, the increase and decrease is *symmetric*.

Finally, note one other important difference. *Simple returns can never be less than 100%, while continuously compounded returns can be less than 100%.*
To see this, consider the return calculations if the index level as of January 2001 were 100 (an extreme case to make our point) instead of 2291.86.

The simple return between March 2000 and January 2001 is: $\frac{100}{5048.62} - 1 = -98.02\%$

The continuously compounded return is: $\ln\left(\frac{100}{5048.62}\right) = -392.17\%$

---

This concludes the math (p)review for this course. I have deliberately covered more mathematical ground than is necessary, just to be sure you can use this note for more courses than one. Hopefully, you will find it helpful to use this note as a handy reference.