Class 1 Notes: Set Theory, Pre-Calculus, and Motivating Calculus

Myles D. Garvey, Ph.D

Summer, 2019

Contents

1 Introduction

In this set of notes, we will conduct a review of set theory, pre-calculus, and motivate some of the problems in calculus that are specific to business applications. Calculus sounds like a scary topic of study to many people. In actuality, it is one of the more simplistic topics in mathematics. Put aside the scary and funny looking symbols for the moment, and you will realize that all of calculus boils down to three subtopics: limits, derivatives, and integrals. The rest of calculus entails how to compute one of these three ideas.

In order to understand calculus, however, we must first establish a thorough review of *set theory*. Set theory **is** the language of mathematics. Without understanding sets, all else becomes moot. So we will begin our review of the fundamentals of set-theory. I recommend to students to read Chapter 0 of the book "Introductory Discrete Mathematics" by V.K. Balakrishnan. In my humble opinion, this is the most thorough review of set theory conducted within 34 pages. He also has a lot of practice problems with solutions in the back of the book to test your knowledge.

After we complete our review of set-theory, we will venture into a review of pre-calculus. Our focus will primarily be on the profile of different mathematical functions. We will see from our set-theory review that functions are defined in a very broad and general manner between two sets. However, when our sets are numerical, we often can define our function in the form of an equation. As such, we will review through the "toolbox" of different types of equations that we are likely to encounter in Calculus.

Once we complete our review of pre-calculus, we will motivate the most fundamental problem in calculus: the limit problem. Once we do this, we will move to motivating two other types of problems: rates of change and approximating areas. I would highly encourage to students that they read through Chapter 1 of Stewart's book, "Functions and Models". This chapter completes a very thorough review of pre-calculus.

Welcome to Calculus for Business!

2 Basic Set Theory

Set Theory allows us to express collections of ”things” in a cohesive framework. The theory was first developed in the early 20th century by Cantor. Cantor’s quest into set theory was primarily to study the notion of ”infinity”. This quest eventually drove him insane and much of his theory was dismissed by mathematicians of the day. A few decades later, set theory became well respected after it was discovered that nearly any field of study involving mathematics can easily be described using the notion of a ”set”. In this section we will cover some of the basic definitions and tools of set theory.

2.1 Sets, Elements and Complements

Before we can attempt to understand anything about functions, or use calculus on them for that matter, we must gain an understanding of sets. Sets are collections of ”stuff”. Think about an empty box for the moment. Suppose the box can fit anything you would like to put inside of it. You can put (almost) anything you want in this box. You can put physical things in this box, such as coffee mugs, trucks, buildings, people, etc, or you can put abstract things in this box such as numbers, truck routes or words. You can even put other boxes inside this box.

The only rules when it comes to this box is that you cannot put the box itself inside it’s own box (see Russell’s Paradox on Wikipedia (https://en.wikipedia.org/wiki/Russell’s_paradox) if you are interested as to why) and every element must be *distinct*. That is, you cannot repeat the same object twice in a box. If you happen to have two of the same *types* of objects, each of those objects still must be distinct. For example, you can have two trucks that look and behave exactly the same, but you must identify them uniquely if you wish to put both trucks in this box. In addition, every box that you define must be *well-defined*. You must specify what objects belong in your box and which ones do not belong in your box. This can be accomplished by either explicitly listing the objects in your box or by specifying some type of criteria such that when given an object the criteria will give a ”yes” or ”no” answer as to whether or not the given object belongs to the box. These ”boxes” are what we refer to as *sets*:

Definition 1 *A* set *is a collection of well-defined distinct objects. No object can appear twice in the set and no set can contain itself. We commonly denote sets using capital letters.*

The objects belonging to a set are commonly referred to as *elements*. We commonly use notation to indicate whether or not a given object is an element of a set or not. This notation is summarized in the definition below:

Definition 2 *An* element *is an object of a set. Let x be an object and A be a set. If x is an element of A then we write $x \in A$. If x is not an element of A, then we write $x \notin A$. Another term for* elements *that is commonly used is* members*.*

Example 1 *Let $A = \{a, b, c, d\}$ and $B = \{g, x, c, a\}$ be sets of letters. Then $a \in A$, $a \in B$, $b \in A$, $b \notin B$, $f \notin A$ and $f \notin B$. Notice how when given an object we are using the symbol ”$\notin$” to denote that an object is not in a set.*

Sets can be defined in a variety of ways. There are two primary ways to define a set. The first way is by explicitly listing the elements of the set within two curly braces. The second way is through the use of something known as *set-builder notation*. The example above lists sets using the explicit listing method. Set-builder notation first defines a variable under consideration, which is intended to represent any member of a set, and then specifies a condition (or criteria) that must be true about the object.

Definition 3 Set-Builder Notation *is a way to define a set by specifying criteria that must be true about elements of the set rather than explicitly listing every element of the set. Let x be a member of a set A. Let $P(x)$ be some type of criteria for belonging to A which gives a "yes" or "no" response. Then the set A can be defined as* $A = \{x | P(x) ='' \mathit{yes}''\}$

Example 2 *Let the set A be the collection of all numbers bigger than 4. Then we would have our criteria as $P(x) = x > 4$. If $x > 4$ then $P(x)$ will give an answer of "yes". For example, $P(5) = $ "yes" while $P(2) = $ "no". We can define the set A as $A = \{x | P(x) ='' \mathit{yes}''\} = \{x | x > 4\}$. That is, A is the set of all numbers x such that $P(x)$ gives an answer of "yes".*

Whenever we work with sets we typically must first define a *universal set*. That is we must start with some type of notion of "all" objects under consideration. There are mathematical reasons for doing this, but for more practical scenarios we typically start with a universe so we are clear as to what objects are under consideration.

Definition 4 *A* universe *is a set of all objects under consideration. All other types of sets we study can only have objects taken from the universe. We typically denote the universal set as U.*

Example 3 *Suppose we would like to study the probability of winning a die-toss game. The game is won by throwing a die and having it land on an even number. We can study this game using set-theory. First we must consider all possibilities. We know that by throwing a six-sided die that we can only have one of 6 possible outcomes. No other outcome is possible for throwing a die (it cannot land on an edge). Therefore, we would have $U = \{1, 2, 3, 4, 5, 6\}$. We can also design two other sets: outcomes of loosing and outcomes of winning: $W = \{2, 4, 6\}$, $L = \{1, 3, 5\}$*

Example 4 *A manager would like to determine which employees deserve a raise. There are 10 employees working for her. She is only willing to give 2 employees a raise. We can describe this situation using set-theory. The collection of all employees would be $U = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}\}$. If he chooses employees 3 and 5 to give a raise, then the set of employees that got a raise would be $R = \{e_3, e_5\}$. Those that did not get a raise could be denoted as $N = \{e_1, e_2, e_4, e_6, e_7, e_8, e_9, e_{10}\}$*

When given a collection that is assumed to be the universe (all possible objects under study), we can define smaller sets with elements that belong only from the universe. Sets with elements taken from another set are commonly referred to as subsets.

Definition 5 *A set A is said to be a* subset *of another set B if all elements belonging to A also belong to B. In this scenario, we denote the statement "A is a subset of B" as $A \subseteq B$. When starting from a universe U, then any set A consisting of elements selected from U is by definition a subset of U. That is, $A \subseteq U$. If there is an element in B that does not belong to A and $A \subseteq B$, then we refer to A as a* proper subset *of B. We denote this as $A \subset B$.*

Example 5 *Let $U = \{t_1, t_2, t_3, t_4, t_5\}$ be a collection of all trucks a company owns. Let $S = \{t_1, t_4, t_5\}$ be the collection of trucks that have a 17ft length. Let $T = \{t_1, t_5\}$ denote the set of trucks that only carry tv's in them. Then we can see that $T \subset S$ since all the elements of T also belong to S and by the same reasoning that $T \subset U$ and $S \subset U$. We notice that $S \not\subseteq T$ since t_5 is not in T.*

The part of this section will focus on the notion of *complements*. When given a universe and a set, we can identify all the elements that are not in the given set. Such a collection of objects is commonly referred to as the complement of a set:

Definition 6 *Let U be a universe and $A \subseteq U$ be a set in the universe. Then the* complement *of A, denoted as A^c, is the collection of all objects in the universe that do not belong to the set A. That is, $A^c = \{x | x \in U \text{ and } x \notin A\}$*

Example 6 *Let us look at the previous example involving the trucks. We have $S^c = \{t_2, t_3\}$ and $T^c = \{t_2, t_3, t_4\}$. Notice that the complements of these sets comprise of all the element* in the universe *that do not belong to the respective sets. We also reference a universe when trying to find the complement.*

Example 7 *The set of all real numbers is commonly denoted as $\mathbb{R}$. That is, $\mathbb{R} = \{x | -\infty < x < \infty\}$. Let $A = (-\infty, 4] = \{x | -\infty < x \leq 4\}$. Then $A^c = \{x | x \notin A\} = \{x | x \notin (-\infty, 4]\} = (4, \infty)$. Notice that our set A comprised of any number that was less than or equal to 4. Since the complement of a set is the collection of elements in the universe that do not belong to the set, and since the universe in this example is the entire number line $\mathbb{R}$, this means that any number not less than or equal to 4 is equivalent to saying the number is strictly larger than 4. Hence, any number larger than 4 is not in the set A and therefore by definition must be in the complement of A.*

Along with the universe set, which is defined based on the problem being described, we also have something known as the *null* or *empty* set. In set-theory this set is required to exist in order for our theory to be complete. That is, if we have some notion of "everything", we must also have the opposite.

Definition 7 *The* empty set *is the set consisting of nothing. We commonly denote this set using $\varnothing$. That is, $\varnothing = \{\} = \{x | x \notin U\}$.*

2.2 Unions and Intersections

In the previous section we introduced the basic concepts and definitions associated with set theory. Similar to numbers, sets can also be operated on. Just as we can add, subtract, multiply and divide number, we can do all the same (or rather more apropos, similar) operations on sets. Given two sets, we would like to "operate" on them to form a new set (similar to how when given two numbers, you would like to "operate" on them to get a new number [new not necessarily meaning different]). Unions, Intersections, Relative Complements and Cross (Cartesian) Products are all similar types of operations we can perform on two given sets. These are the most basic and common types of operations used in set theory.

We first begin with unions. Imagine you are given two sets: set $A = \{2, 4, 6\}$ and $B = \{1, 3, 5\}$ and would like to "mash" them together. If we think of forming a new set by taking all the elements in the first set and all the elements in the second set (without repeating duplicated elements), then we have in some sense "operated" on the two sets to form a new set, namely one that comprises of all the unique elements taken from both sets. Such an operation is called the *union* of two sets:

Definition 8 *Let A be a set an B be another set. Then the* union *of sets A and B defined to be the set comprising of all elements taken from both sets A and B. Mathematically we denote this as* $A \cup B = \{x | x \in A \text{ or } x \in B\}$

Notice in our mathematical definition we used the word "or". This word actually has a very specific meaning. When we use the word "or" in criteria statements for set-builder notation it means that *at least one* of the statements must hold true. This is different than our every day meaning or understanding of "or", which most of the time is understood to mean "only one of the two must be true". In set theory, when we write "or" we leave open the possibility that *both* statements could simultaneously be true.

Example 8 *Let* $A = \{1, 5, 3\}$ *and* $B = \{2, 6, 10\}$. *Then* $A \cup B = \{1, 2, 3, 5, 6, 10\}$. *We see that this is found by taking all of the elements from the set A and combining them together with all of the elements from set B. Taken together, we were able to form a new set, the union of A and B.*

Example 9 *Let* $A = \{2, 5, 8\}$ *and* $B = \{1, 4, 8\}$. *Then* $A \cup B = \{1, 2, 4, 5, 8\}$. *Notice how the number 8 appears in both sets but is only listed once in the union set. This is because 8 is the same element in both sets. Recalling our discussion earlier about the basics of sets, we cannot list the same element twice. A set needs to be only unique elements. 8 is a unique element. Even though it is in both sets, it only gets listed once in the final set. The same is true of any other elements that happen to overlap in both sets.*

Unions allow us to combine two sets together into a new third set. Another interesting operation on sets would be to determine the common elements of two sets. Such an operation is commonly referred to as an *intersection* of sets. This set finds the elements that are common to two given sets.

Definition 9 *Let A be a set an B be another set. Then the* intersection *of sets A and B is defined to be the set comprising of all elements that are common between sets A and B. Mathematically we denote this as* $A \cap B = \{x | x \in A \text{ and } x \in B\}$

Notice the distinction in the wording of the criteria for the intersection. For the union, we used "or". For the intersection, we used "and". Using "and" in criteria statements in Set-Builder Notation means that both statements must be true about an element simultaneously. In the definition of the intersection, we know that for an element to belong in the intersection is must be true that it belongs to both sets. This is different than saying it must belong to *either* set.

Example 10 *Let* $A = \{1, 5, 3\}$ *and* $B = \{2, 6, 10\}$. *Then* $A \cap B = \{\} = \varnothing$. *We see that this is found by taking all of the elements from the set A and finding common elements in set B. Taken together, we were able to form a new set, the intersection of A and B. In this example, we happen to notice that A and B have no common elements. Hence, the intersection would be empty. That is, the intersection would be just the empty set, since there are no common elements.*

Example 11 *Let* $A = \{2, 5, 8\}$ *and* $B = \{1, 4, 8\}$. *Then* $A \cap B = \{8\}$. *Notice that 8 is the only common element between the two sets. Once again we only list this once (since it is a unique and distinct element).*

2.2.1 Identities of Sets Involving Intersection and Unions

An important observation to note is that we can union and intersect sets with the universe and empty set. Let $A \subseteq U$. Then we notice that if we were to union the universe with the set A we should return back the universe. That is, $A \cup U = U$. Why is this? The universe contains every element under study. Since A is a subset of the universe, every element of A is already in U. Since every element of A is already in the universe, it must not contain any elements outside the universe. Hence, if we were to union any set with the universe, we get back the universe.

Conversely, if we were to find the intersection of A and U, we should expect to return A. That is, $A \cap U = A$. Why is that? Remember, the intersection between two sets comprises of all elements that are common to both sets. We know that by definition $A \subseteq U$. Hence, every element in A is also in U. But the only elements that U shares with A are only elements in A. If there were to exist another element in A that is not in U, then U is not a universe (by definition). Hence, for any set under consideration, anytime it is intersected with the universe, we just simply get back that set.

If we have a set A and union that with the empty set $\varnothing$, then we should expect to get back A. Think about it, if we union any set with a set that contains nothing, then we are not adding anything else other than what is in the original set. Therefore we must have $A \cup \varnothing = A$.

For the intersection of any set with the empty set, intuitively we must expect the resulting set to be $\varnothing$. Recall that the intersection only comprises of elements that are common in both sets. If we intersect any set with a set that contains nothing, then the only thing in common between the two sets is, well, nothing. Hence we expect to have $A \cap \varnothing = \varnothing$.

Summarizing these results we lead to the following theorem:

Theorem 1 *Set Universe and Empty Set Identities*
Given a universe U and a set $A \subseteq U$, we must have:

$$\begin{aligned} A \cup U &= U \\ A \cap U &= A \\ A \cup \varnothing &= A \\ A \cap \varnothing &= \varnothing \end{aligned} \tag{1}$$

These identities help us simplify complicated expressions of different sets involving intersections and unions. There are other identities, however, that help us further. These involve complements. If we were to union a set together with it's complement, we should expect to get the universe. Think about that intuitively. The complement of a set is any element in the universe that is not in A. A contains all the other elements in the universe. By using any set and it's complement, you have fully accounted for every element in the universe. Hence, we should expect to see $A \cup A^c = U$.

By definition of the complement of a set A, any element in A^c is not in A. Therefore, there should be no elements that are common in either of these sets. Hence, we should also expect to see $A \cap A^c = \varnothing$. We can summarize these results in the following:

Theorem 2 *Set and Complement Identities*
Given a universe U and a set $A \subseteq U$, we must have:

$$
\begin{aligned}
A \cup A^c &= U \\
A \cap A^c &= \varnothing
\end{aligned} \tag{2}
$$

Likewise, if we consider the set of everything (i.e the universe), then the opposite of "everything" is nothing, and vice versa. Therefore, the following is true:

Theorem 3 *Let U be the universe and $\emptyset$ be the empty set. Then:*

$$
\begin{aligned}
U^c &= \varnothing \\
\varnothing^c &= U
\end{aligned}
$$

Just like in algebra and arithmetic, the set operations of intersection and union are both associative and commutative. Likewise, we also have a distributive property to help us simplify expressions involving sets:

Theorem 4 *Let A, B, and C be sets. Then the following is true:*

$$
\begin{aligned}
(A \cup B) \cup C &= A \cup (B \cup C) \\
(A \cap B) \cap C &= A \cap (B \cap C) \\
A \cup B &= B \cup A \\
A \cap B &= B \cap A \\
A \cap (B \cup C) &= (A \cup B) \cap (A \cup C) \\
A \cup (B \cap C) &= (A \cap B) \cup (A \cap C)
\end{aligned}
$$

Example 12 *If we are given complicated expressions of sets, we can leverage the laws given above to simplify them. For example, suppose we are given the expression $(A^c \cup (A \cap (B \cup C) \cap B^c))^c$, then we can leverage each rule above to simplify this down:*

$$
\begin{aligned}
(A^c \cup (A \cap (B \cup C) \cap B^c))^c &= (A^c \cup (A \cap ((B \cap B^c)) \cup (C \cap B^c))))^c \\
&= (A^c \cup (A \cap (\varnothing \cup (C \cap B^c))))^c \\
&= (A^c \cup (A \cap (C \cap B^c)))^c \\
&= ((A^c \cup A) \cap (A^c \cup (C \cap B^c)))^c \\
&= (U \cap (A^c \cup (C \cap B^c)))^c \\
&= (A^c \cup (C \cap B^c))^c
\end{aligned}
$$

2.3 Demorgan's Law

In the previous section we introduced intersections and unions. They proved to be valuable operations for forming new sets from two other given sets. Such operations will show up again when we discuss probability theory. The theory discussed so far will vastly help in understanding operations on events of random variables, when we reach that point in the course. As it turns out, there is a relationship between unions, intersections and complements that helps in simplifying complicated expressions involving sets. We will not discuss the proof of this relationship, but you can draw out Venn Diagrams to confirm for yourself that these relations do indeed work.

Theorem 5 *De Morgan's Law*
Given a set A and a set B, the following are true:

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c \end{aligned} \tag{3}$$

Example 13 *Let* $A = \{1, 5, 8\}$, $B = \{5, 11, 52\}$ *and* $U = \{1, 5, 7, 8, 11, 50, 52\}$. *We would like to find the complement of the union of these two sets. This can be done in one of two ways. First, we will calculate the union:* $A \cup B = \{1, 5, 8, 11, 52\}$. *Now the complement of this set is* $(A \cup B)^c = \{7, 50\}$. *The alternative way to calculate this is to first find the complements of A and B. We have* $A^c = \{7, 11, 50, 52\}$ *and* $B^c = \{1, 7, 8, 50\}$. *The intersection of these two would be* $A^c \cap B^c = \{7, 50\}$. *Hence, we have demonstrated that* $(A \cup B)^c = A^c \cap B^c$.

2.4 Quantifiers

Sometimes in our mathematical models we will use a language of logic to help simplify and describe our models. When we would like to make statements about multiple objects in a set we commonly employ something known as the *universal* quantifier. Doing so greatly simplifies our mathematics and drives home the message we are trying to effectively communicate. In addition, sometimes we would like to make a statement about the existence of certain objects. These tend to be called *existential* quantifiers. We will motivate the definition of these symbols with an example.

Suppose you have a set of 5 trucks denoted as $T = \{t_1, t_2, t_3, t_4, t_5\}$. In addition, each truck has a total capacity of the number of boxes, of a fixed size, that it could hold. We can denote these capacities as $C = \{c_1, c_2, c_3, c_4, c_5\}$. We also would like to take into account the number of products that will be in the truck. We can denote the number of products each respective truck is carrying as $X = \{x_1, x_2, x_3, x_4, x_5\}$. We know that these variables are constrained by the capacity of each truck. We can state this using the following inequalities:

$$\begin{aligned} x_1 &\le c_1 \\ \text{and } x_2 &\le c_2 \\ \text{and } x_3 &\le c_3 \\ \text{and } x_4 &\le c_4 \\ \text{and } x_5 &\le c_5 \end{aligned}$$

Now, in a simple example like this, this may not seem too bad. What if, however, you had over 100 trucks? 1000 trucks? Are you really going to waste time writing all of these inequalities? This will just take too long. Not to mention, it will not really summarize the key points of your mathematical model correctly. Instead, we want to shrink this inequality down to one line. Let i be *any* truck possible in the fleet. That is, i is any number between 1 and 5, or, using our newfound terminology, let $i \in \{1,2,3,4,5\}$. This previous statement says that i can be any number in the set $\{1,2,3,4,5\}$. Yet, we don't know *which* number i is referring to. We do know, however, that regardless of what the value of i is, we want the following statement to be true: $x_i \leq c_i$

Let us clarify this. First, assign a value to the variable i, where the only possible numbers i can be set equal to are the values in the set $\{1,2,3,4,5\}$. Then, replace i everywhere in the inequality $x_i \leq c_i$. We have now made a statement about a *specific* i. What if, however, we want to make this statement about *all* of the i assigned to values in the set $\{1,2,3,4,5\}$? We could say "for all of the elements i in the set $\{1,2,3,4,5\}$, the inequality $x_i \leq c_i$ must be true ". This is too long, however. It turns out that we can shorten this sentence down using a symbol known as the *universal quantifier*.

Definition 10 *The* universal quantifier *refers to a statement about* all *elements of a given set. This is commonly denoted as* $\forall$.

So for the example above, we can make the statement:
$x_i \leq c_i \; \forall i \in \{1,2,3,4,5\}$

This line of mathematics says "for every single member denoted as i in the set {1,2,3,4,5}, it must be true that after assigning a value from this set to i that the inequality $x_i \leq c_i$ must be true.

Quantifiers such as this help us summarize statements about elements in sets. Rather than having to list out all 5 inequalities, we only needed to list one in a more general form, which allowed us to make a statement not just solely about one of the trucks, but all of the trucks.

There is another quantifier that helps us make general statements such as the one presented above. Suppose that we would like to place a condition in our model so that a statement is true for *at least one* element in our set but not necessarily all of them (although we allow for this to be a possibility). For example, suppose we would like to have at least one of the trucks carry a load exactly equal to it's capacity. We might not care specifically about which truck, but rather at least one truck must carry a load that is exactly equal to it's capacity. We can make this statement the long way:

$$\begin{aligned} x_1 &= c_1 \\ \text{or } x_2 &= c_2 \\ \text{or } x_3 &= c_3 \\ \text{or } x_4 &= c_4 \\ \text{or } x_5 &= c_5 \end{aligned}$$

Note that the equalities listed doesn't mention that *all* of the trucks need to equal their capacity, but only at least one, yet we don't know (nor do we care) which specific one equals it's capacity. We can simplify this statement down, once again, to only one line, using something known as the *existential quantifier*:

Definition 11 *The* existential quantifier *refers to statement about* at least one *of the elements of a given set. This is commonly denoted as* $\exists$. *Note, that it is possible for a statement to be true for more than one of the elements of a given set. However, it is not required. The only requirement is that there is at least one element for which a given statement is true.*

So, we reduce our earlier statement to:
$\exists i \in \{1, 2, 3, 4, 5\}$, $x_i = c_i$

In English, this says ”there exists at least one element i in the set $\{1, 2, 3, 4, 5\}$ such that the statement $x_i = c_i$ is true. When we design mathematical models, we will commonly employ quantifiers in our model description to shorten down the summary and deliver the main reason for a particular statement without having to resort to make a statement about every or at least one element in a given set.

2.5 Cartesian Products of Sets

Now that we have covered the basics of set theory we can discuss some more advanced topics such as looking at possible relationships between sets. Before we can discuss relations and functions, we must first discuss yet again another operation that can be performed on two sets. Imagine for the moment that a company has three trucks in their fleet, denoted by the set $T = \{t_1, t_2, t_3\}$. In addition, there are four different routes that the truck can take, denoted by $R = \{r_1, r_2, r_3, r_4\}$. The goal is to assign trucks to different routes. Obviously, this is an example of a function, since we most likely would only have one truck following one route, but two trucks can possibly take the same route. However, for this example, we are not necessarily concerned with finding the assignment of the trucks to routes but merely looking to ask the question: what are all the possible assignments of trucks to routes? We can visualize all the possible assignments in the figure below:

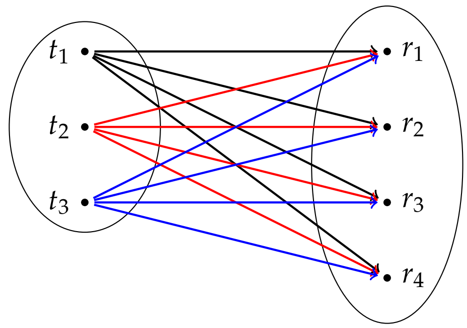


Figure 1: All possible assignments of trucks to routes

We notice that for every truck, we have an arrow pointing to every route. Each arrow represents some type of *assignment* from a single truck in the set of trucks to a single route in the set of routes. While we have a way to represent these assignments visually (using arrows), we are interested in finding a way to represent these assignments using mathematics. If we have two sets, A and B, we can denote an assignment from an element $a \in A$ to an element $b \in B$ using the ordered pair (a, b). Therefore, if we are looking to find all the *possible* assignments from set A to set B, we can list out every single possible ordered pair, where the first element in the pair is an element of the first set and the second element in the pair is an element in the second set. Therefore, for our example, the set of every possible assignment of trucks to routes can be mathematically represented as follows:

$$\{(t_1,r_1),(t_1,r_2),(t_1,r_3),(t_1,r_4),(t_2,r_1),(t_2,r_2),(t_2,r_3),(t_2,r_4),(t_3,r_1),(t_3,r_2),$$
$$(t_3,r_3),(t_3,r_4),(t_4,r_1),(t_4,r_2),(t_4,r_3),(t_4,r_4)\}$$

Notice what we have done in our little example here. We have taken two sets and found every possible assignment from one set to another. In this example, all the possible assignments from the set of trucks to the set of routes. The resulting set represents all the possible assignments from elements of the first set to elements of the second set. This type of set is commonly referred to as a *Cartesian product*.

Definition 12 *Let A and B be two sets. Then the Cartesian product between set A and set B is the set of all assignments of elements from set A to elements in set B. In other words, it the set of all possible ordered pairs between the two sets. We denote this mathematically as:*
$A \times B = \{(a, b) | a \in A \text{ and } b \in B\}$

Example 14 *Let $A = \{1, 4, 6, 8\}$ and $B = \{5, 7, 8, 9\}$. Then the cartesian product of the sets would be:*

$$A \times B = \{(1,5),(1,7),(1,8),(1,9),(4,5),(4,7),(4,8),(4,9),(6,5),(6,7),(6,8),$$
$$(6,9),(8,5),(8,7),(8,8),(8,9)\}$$

Notice that this is different than asking "what is the Cartesian product between B and A?" If we wanted to find this, we would need to match every element of B to an element in A:

$$B \times A = \{(5,1),(5,4),(5,6),(5,8),(7,1),(7,4),(7,6),(7,8),(8,1),(8,4),(8,6),$$
$$(8,8),(9,1),(9,4),(9,6),(9,8)\}$$

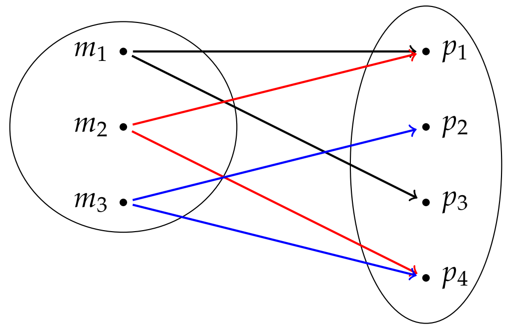


Figure 2: Computer models using different parts

We clearly, in this example, can see that $A \times B \neq B \times A$

2.6 Relations and Functions

In the previous section we discussed yet another operation between two sets, the Cartesian product. This operation gave us the set of all possible assignments between one set and another. What if, however, we wanted to focus on a very specific type of assignment? In nature and practical applications, very rarely do we actually have all members of one set assigned to all members of another set. That is, *relations* between members are not always fully exhaustive such as what the Cartesian product provides. In this section we will discuss the theory behind *relations* and *functions*, which allow us to express associations between members of two sets without having to fully exhaustively list every possible association that the Cartesian product provides.

We will motivate our definition of a *relation* with an example. Suppose that a computer manufacture produces three different models of computers, which we will represent with the set $M = \{m_1, m_2, m_3\}$. The different models use parts, which we will represent with the set $P = \{p_1, p_2, p_3, p_4\}$. Model 1 uses parts 1 and 3, Model 2 uses parts 1 and 4 and Model 3 uses parts 2 and 4. We can represent these associations between models and parts using the diagram in Figure 2.

Notice how this is similar to our discussion in the previous section on Cartesian products. Just as we had done there, we can represent these model-part associations using ordered pairs. Let the first element in the pair denote the model and the second element in the pair denote the part. Then the associations that the manufacture uses for production of each model to each part can be represented mathematically as the following set:
$R = \{(m_1, p_1), (m_1, p_3), (m_2, p_1), (m_2, p_4), (m_3, p_2), (m_3, p_4)\}$

This set completely describes the relationships between the models and the parts mentioned verbally earlier. Also notice, however, that the set R happens to be a subset of the Cartesian product between the two sets, that is, $R \subseteq M \times P$. The set R is commonly referred to as a *relation*. In general, any subset of the Cartesian product between two subsets is called a relation.

Definition 13 *Let A and B be sets. Then a* relation *R between A and B is a set of ordered pairs with elements in the first position of the pair taken from the set A and elements of the second position of the pair taken from the set B. Put simply, a relation between A and B is any subset R of $A \times B$.*

While relations may be of interest in solving some practical problems, they are less frequent than a special type of relation: the function. Often in the real-world we encounter this special

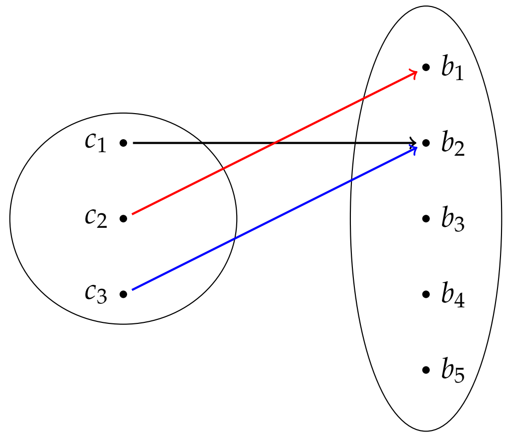


Figure 3: Consumers and their Brand Preference

type of relation. Hence we will shift focus from the general theory on relations to the basics of functions.

Let us motivate the definition of a function with another example. Suppose that we have 3 consumers that would like to purchase a new tv. All consumers in the market are heavily brand loyal and once they purchase a brand they tend not to change their brand. Let the set $C = \{c_1, c_2, c_3\}$ denote our consumers and assume there are only 5 brands of televisions, denoted by the set $B = \{b_1, b_2, b_3, b_4, b_5\}$. Assume consumer 1 prefers brand 2, consumer 2 prefers brand 1 and consumer 3 prefers brand 2. The first observation we should take note of is that this once again is a relation. We can describe this relation mathematically as the set of ordered pairs $R = \{(c_1, b_2), (c_2, b_1), (c_3, b_2)\}$. We once again can also visualize these associations using the map diagram in Figure 3

There is a very interesting property of this relation in this example. We make two observations. First, every single element in the set A appears in at least one of the ordered pairs in R. In addition, every element in A appears in an ordered pair *exactly once*. That is, there are no repeats of any of the elements from A. Such relations are commonly referred to as *functions*. We will properly define them below:

Definition 14 *A* function f *between two sets* A *and* B*, commonly written as* $f : A \to B$*, is a relation (i.e.* $f \subset A \times B$*) between the elements of these sets such that*

1. $\forall a \in A \; \exists b \in B$ *such that* $(a, b) \in f$
2. $\forall (a_1, b_1) \in f$ *and* $(a_2, b_2) \in f$ *if* $a_1 = a_2$ *then* $b_1 = b_2$

In other words:

1. *All elements in* A *must also appear as a first element in an ordered pair in the relation* f
2. *For any element in* A *there can only be one ordered pair in* f *with that element in the first position*

In the mapping diagram for f *we should only see one arrow exiting from every element in* A *.*

When we define functions, we usually first define which sets the function is between. After, we define the *structure* of the function itself. This structure can be defined in multiple ways, as we will see in the next section.

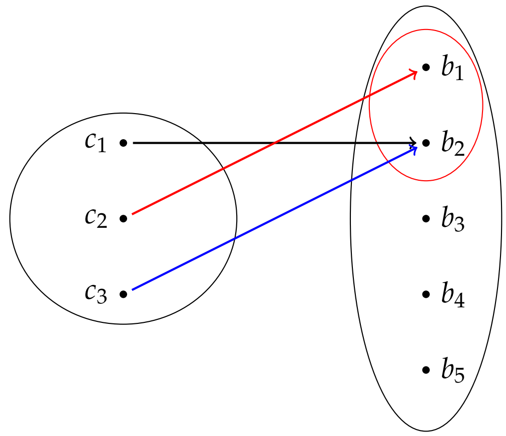


Figure 4: Consumers and their Brand Preference with Range in the Red Circle

Definition 15 *If we have two sets A and B and a function $f \subset A \times B$, then we can denote the function using the following notation:*
$f\colon A \to B$

The arrow indicates that elements of A are being mapped (or assigned) to elements in B. A is commonly referred to as the domain *and B is commonly referred to as the* codomain. *We also can denote the elements in B for which elements in A are mapped to using a specific notation as well. If $(a, b) \in f$, then we can write $b = f(a)$. The notation $f(a)$ is called the* image *of $a \in A$ and represents the element in B to which a is mapped.*

We might notice another interesting observation from the previous example as well. We notice that not every element in the codomain has an arrow pointing to it. We hence can define another type of set when it comes to functions:

Definition 16 *The* range *of the function f is the set of all images of $a \in A$ under the function f. That is, the range of the function is the set*
$\{f(a)|a \in A\}$

We can identify the range visually (at least in an example this small) from the previous example's mapping diagram by observing all the elements in B that have arrows being pointed to. In Figure 4 we see that the red circle indicates the range, that is, the set of all images. Typically, in the study of functions, we usually concern outselves over the domain and the range.

Example 15 *Suppose we have the following sets:*
$A = \{1, 3, 5, 6, 7\}$
$B = \{4, 8, 10, 13, 19, 21\}$
$C = \{-14, -2, 8, 17\}$

In addition, assume we have the following relations:
$R_1\colon A \to C$
$R_2\colon C \to B$
$R_3\colon B \to A$

each of which are defined as

$R_1 = \{(1, -14), (3, 17)\}$
$R_2 = \{(-14, 8), (-2, 8), (8, 21), (17, 13)\}$
$R_3 = \{(4, 3), (8, 3), (8, 5), (10, 3), (13, 7), (19, 5), (21, 7)\}$

We notice that R_1 and R_3 are not *functions. In R_1, not every element of A is used as a first element. In R_3 we notice that the element $8 \in B$ is listed twice with different mapped elements in A, namely 3 and 5. We do notice, however, that R_2 is a function. Every element in C is listed in the relation and is only listed once. By definition, it's domain is C, it's codomain is B and it's range is found by observing all the elements* mapped to. *In this case, the elements in B that are mapped to from C are $\{8, 21, 13\}$. Therefore, this set is the range of R_2.*

2.7 Three Fundamental Types of Functions

Functions sometimes have special properties which help aid us in other mathematical tasks. These three properties are referred to as surjectivness, injectiveness, and bijectiveness. Recall that all functions are defined by first defining a *domain* (the A) and then defining the *codomain* (the B). Further recall the definition of the *range* of a function, which is the set $\{f(x) | \forall x \in A\}$. It is not always true, as we have seen like in the example illustrated in Figure 3, that $f(A) = B$. When it is, however, we say that the function is *surjective*:

Definition 17 *A function $f : A \to B$ is referred to as a surjection if $\forall y \in B, \exists x \in A$ s.t $f(x) = y$, or $(x, y) \in f$*

Another property of functions is that of injectivness. When we have all elements in the domain map to one unique element in the range that no other value in the domain is mapped to, then we say that the function is *injective*:

Definition 18 *A function $f : A \to B$ is referred to as an injection if $\forall x_1, x_2 \in A$, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, or put differently, if $y_1 \neq y_2$, then $\forall x \in A$, we cannot have $(x, y_1) \in f \wedge (x, y_2) \in f$.*

When both of these properties are true about a function, then we shorten our vocabulary and just simply say that the function is *bijective*:

Definition 19 *If the function $f : A \to B$ is both injective and surjective, then it is said to be bijective.*

We can see in Figures 5 that the function f is surjective but not bijective, in Figure 6 that the function f is injective but not surjective, and in Figure 7 that the function f is bijective.

2.8 Cardinality

Cardinality is the size of a set. That is, it measures how much "stuff" is in the set. A set is said to be *finite* if we can explicitly write down all the elements on a sheet of paper when given a large enough sized paper and stop at some point. That is, the elements in a finite set come to an "end", and we can write them all out explicitly when given enough time and space. For example, all whole numbers from 0 to 1,000,000 is a finite set. This is a very large set, but it is still finite, since we can write them all on a sheet of paper and eventually come to an end.

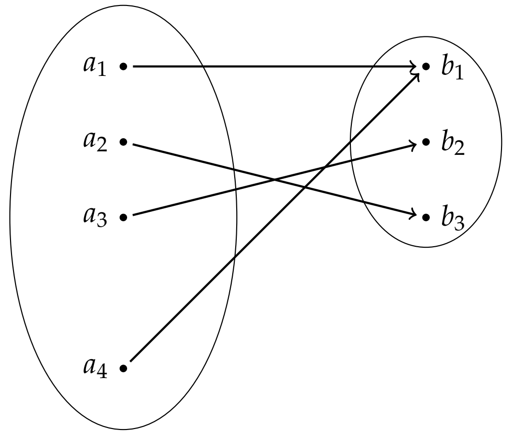


Figure 5: The function $f : A \to B$ (the entire function is represented by the set of all arrows in the figure) is defined as $f = \{(a_1, b_1), (a_2, b_3), (a_3, b_2), (a_4, b_1), \}$. We see that this is a function, since every $a \in A$ appears exactly once as the first element in the tuples in the definition of f. We also see that this is surjective, since every $b_i \in B$ has a corresponding $a_j \in A$ for which the function maps. However, we see that this function is not injective, since $(a_1, b_1) \in f$ and $(a_4, b_1) \in f$, which violates the definition of a function being injective. Since it is not injective, it therefore is also not bijective.

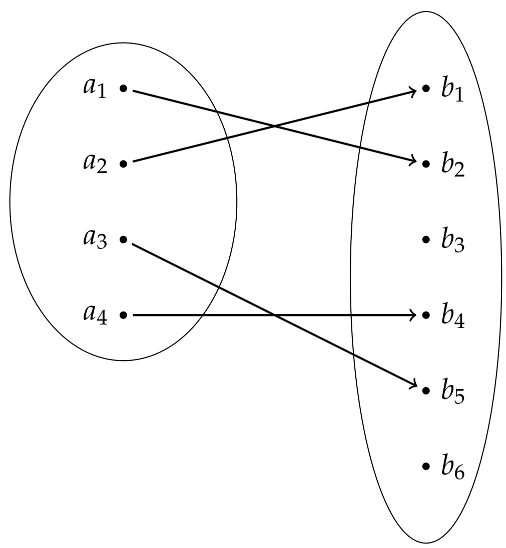


Figure 6: The function $f : A \to B$ (the entire function is represented by the set of all arrows in the figure) is defined as $f = \{(a_1, b_2), (a_2, b_1), (a_3, b_5), (a_4, b_4), \}$. We see that this is a function, since every $a \in A$ appears exactly once as the first element in the tuples in the definition of f. We also see that this is injective, since for every $a_i, a_j \in A$ such that $a_i \neq a_j$, we see that $f(a_i) \neq f(a_j)$. However, we notice a violation the definition of a function being surjective, since elements b_3, b_6 are not in the definition of f. That is, there does not exist an $a_i \in A$ so that $b_3 = f(a_i)$ and $b_6 = f(a_i)$. Since it is not surjective, it therefore is also not bijective.

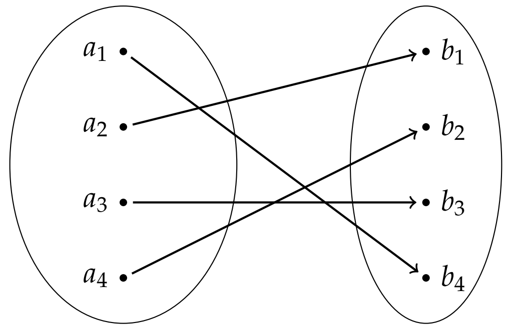


Figure 7: The function $f : A \to B$ (the entire function is represented by the set of all arrows in the figure) is defined as $f = \{(a_1, b_4), (a_2, b_1), (a_3, b_3), (a_4, b_2), \}$. We see that this is a function, since every $a \in A$ appears exactly once as the first element in the tuples in the definition of f. We also see that this is injective, since for every $a_i, a_j \in A$ such that $a_i \neq a_j$, we see that $f(a_i) \neq f(a_j)$. We also see that for all elements in B, we can find an element in A to match it to using the function. Therefore, this function is also surjective. Hence, this function is bijective, since it is both surjective and injective

.

The cardinality of a finite set is just simply equal to the number of elements in the set. More generally, if we have a finite set A with n elements in it, then it's cardinality is n, and we denote this by $|A| = n$. Some examples of cardinalities:

$$\begin{aligned}
|\emptyset| &= 0 \\
|\{a_0\}| &= 1 \\
|\{a_0, a_1\}| &= 2 \\
|\{a_0, a_1, a_2\}| &= 3 \\
&\vdots \\
|\{a_0, a_1, \dots, a_{n-1}\}| &= n
\end{aligned}$$

In set theory, we consider two sets to be "the same" if they have the same size. For example, the sets $A = \{1, b, joe\}$ and $B = \{0, k, m\}$ are considered to be "equivalent" (not equal, but just equivalent with respect to size) since $|A| = |B|$. There is a very useful theorem that we can use (I will leave it to you as an exercise to prove on your own) to show that two sets are the same "size" (heretofore referred to as cardinality):

Theorem 6 *If there exists a bijective function $f : A \to B$, then $|A| = |B|$.*

You can see that this theorem is true visually by observing our earlier figure depicting a bijection. However, while it is easy to see that two finite sets which have a bijection between them must be the same size, and while cardinality itself is easy to understand when working with finite sets, the same cannot be said of non-finite sets (notice my emphasis on the lack of using the word "infinite", this is intentional for reasons which will soon become clear). A set that has no end is referred to as a non-finite set. A question that comes to mind is, if a set A is non-finite, does this mean that $|A| = \infty$? In short, and shockingly, the answer is no. We will illustrate this by reconstructing the number systems (Natural Numbers, Integers, Rational Numbers, Real Numbers) in the next section.

3 The Number Systems and Their Cardinality

3.1 The Natural Numbers

Throughout many centuries, practitioners and mathematicians alike have tried to understand (1) what actually is a number, and (2) can we "account" for all numbers? This quest began with the idea of counting. We know that the number 1 represents a count of there being only 1 "thing". The number 5 represents a count of there being only 5 "things". The number 0 (debates of it's existence led to the start of much bloodshed) represents that we have no "things". A question is, can we account for all possible "counts"? This too is a hotly debated topic, but for our purposes, we will assume the answer is yes.

Definition 20 *The set of natural numbers, denoted by* $\mathbb{N}$*, is the collection of all possible numbers where each element represents the cardinality of a finite set, namely* $\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, \dots\}$.

This collection represents our ability to count "things" as well as list "things". If we can list or count something, then we should be able to find a subset $A \subset \mathbb{N}$ such that our set or list of "things" can be put into a bijection with A. For example, if we list all employees in our company $E = \{bob, susan, megan\}$, there are a total of 3 employees, so we can create a bijection between the subset $A = \{0, 1, 2\} \subset \mathbb{N}$ and the list of our employees E. For example, one such bijection is $f : A \to E$, $f = \{(0, bob), (1, susan), (2, megan)\}$.

3.1.1 Cardinality of the Natural Numbers

A question that stumped many mathematicians was a simple one (simple questions are the often the most difficult to resolve): what is the cardinality of $\mathbb{N}$? We may speculate that it is ∞, however, this would cause a lot of problems (which we will soon see). Instead, we will just simply define, rather than "find", the cardinality of $\mathbb{N}$ as $\aleph_0$. That is, $|\mathbb{N}| = \aleph_0$.

Here is a fun puzzle to think about. Define the set of even numbers as $\mathbb{E} = \{0, 2, 4, 6, 8, \dots\}$ and the set of odd numbers as $\mathbb{O} = \{1, 3, 5, 7, 9, \dots\}$. We notice that in both of these sets, there are numbers missing. The even numbers are missing the odd numbers, and vice versa. Naturally (no pun intended), we may think that $|\mathbb{O}| < |\mathbb{N}|$ and $|\mathbb{E}| < |\mathbb{N}|$, right? Afterall, we did take numbers "out" of the set of natural numbers. Well, this is where our intuition fools us.

Recall we said that if we can find a bijective function between two sets that the two sets must be the same cardinality. Consider the function $f : \mathbb{N} \to \mathbb{E}$ defined by $f(n) = 2n$. Taking the first few elements of the natural numbers, we see that

$$\begin{aligned} f &= \{(0, f(0)), (1, f(1)), (2, f(2)), \dots\} \\ &= \{(0, 2(0)), (1, 2(1)), (2, 2(2)), \dots\} \\ &= \{(0, 0), (1, 2), (2, 4), \dots\} \end{aligned}$$

Is the function $f(n) = 2n$ bijective? Lets see. First, we need to prove it is surjective. Let $n \in \mathbb{E}$. By definition, n is an even number and all even numbers can be written as a factor of 2: $n = 2k$. Since n is divisible by 2, and since $n \in \mathbb{E}$, we must have $k \in \mathbb{N}$. Therefore, the function is surjective. What about injective? Let $n_1, n_2 \in \mathbb{E}$ and assume that $n_1 = n_2$. Since it is surjective, we know there must exist a corresponding $k_1, k_2 \in \mathbb{N}$ such that $n_1 = f(k_1) = 2k_1$ and $n_2 = f(k_2) = 2k_2$. Since we assumed that $n_1 = n_2$, we must have $2k_1 = 2k_2$. Dividing both

sides by 2 leaves us with $k_1 = k_2$, which fits our definition of injective. Therefore, the function is bijective. This is very psychedelic indeed! We just proved that $|\mathbb{N}| = |\mathbb{E}|$, despite us "missing" numbers from $\mathbb{E}$! A similar proof will show that $|\mathbb{N}| = |\mathbb{O}|$ (do this as an exercise!). Therefore, we have shown that:
$\aleph_0 = |\mathbb{N}| = |\mathbb{O}| = |\mathbb{E}|$

3.2 The Integers

The invention of the integers grew out of a need to have numbers represent an action of "taking away" something. Natural numbers are easy to understand: they represent counts. But what if we had a deficit of something? We are lacking 3 customers than yesterday (-3), we took a debt out for 5 coins of gold on something, and now have (-5) gold coins, etc. As such, mathematicians realized that we are missing numbers, and so they constructed *the integers* to "fill in" this gap:

Definition 21 *The negative of a natural number n is defined as $-n$, such that $n + (-n) = 0$. The set of integers, denoted as $\mathbb{Z}$, is the collection of natural numbers and the negatives of the natural numbers, namely $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.*

3.2.1 Cardinality of the Integers

Now, $\mathbb{Z}$ clearly has "more" numbers in it than the natural numbers $\mathbb{N}$, right? We saw our intuition fail before, but can it fail now yet again? The answer, yes, our intuition has failed, again! These are indeed the same size! I know, we are heading down a very strange rabbit hole here. Despite the integers having "more" numbers in them than the natural numbers, they are the same size sets. Let us construct a bijection between the sets to prove this. Let the function $f : \mathbb{N} \to \mathbb{Z}$ defined by:

$$f(n) = \begin{cases} 0 & n = 0 \\ \frac{n+1}{2} & n \in \mathbb{O} \\ -\frac{n}{2} & n \in \mathbb{E} \end{cases}$$

For the first few natural numbers, we have:

$$\begin{aligned} f &= \{(0, f(0)), (1, f(1)), (2, f(2)), (3, f(3)), (4, f(4)), \ldots\} \\ &= \{(0,0), (1, \frac{1+1}{2}), (2, -\frac{2}{2}), (3, \frac{3+1}{2}), (4, -\frac{4}{2}), \ldots\} \\ &= \{(0,0), (1,1), (2,-1), (3,2), (4,-2), \ldots\} \end{aligned}$$

I will leave it to you as an exercise to prove that this is indeed a bijection. So now we have established that:
$\aleph_0 = |\mathbb{N}| = |\mathbb{O}| = |\mathbb{E}| = |\mathbb{Z}|$

3.3 The Rational Numbers

Part of the history of mathematics involves the invention of rational numbers. Mathematicians of the day realized that numbers were still "missing" from the standard natural numbers and

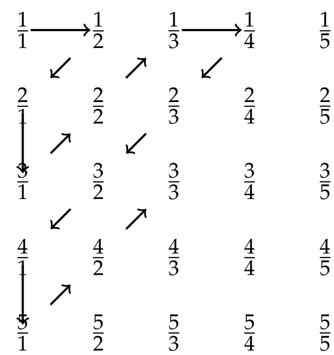


Figure 8: How to "count" the rational numbers. Lay out all possible fractions in the grid with the numerators as the row number and the denominators as the column number. Start the count at $\frac{1}{1}$. Move from $\frac{1}{1}$ to $\frac{1}{2}$. Now count down the diagonal ending at $\frac{2}{1}$. Move down to $\frac{3}{1}$, count up this diagonal to $\frac{1}{3}$. Move across to $\frac{1}{4}$, count down to this diagonal to $\frac{4}{1}$, and so on and so forth. This puts the natural numbers into a bijection with the rational numbers.

integers. We know that we can "count" things, in two directions (gains vs losses), but what if we took a single "thing" and "broke it"? If we have one slice of bread, we can break the bread. The unit of a "slice" has not changed, and yet, we have more "things" of a smaller size than we had before (split a slice of bread into two, eat the first part, we have less "slices" remaining than we had before, but not 0 slices). If we split the bread into two, and count that there is now more bread, this is somewhat paradoxical. So, this observation led mathematicians to the discovery (or invention, whichever viewpoint you have on that one!) to *rational numbers*.

Definition 22 *The collection of rational numbers, denoted as $\mathbb{Q}$, is the set of all numbers that can be expressed as a fraction of two integers. We denote this set by:* $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}\}$

3.3.1 Cardinality of the Rational Numbers

Surely, the rational numbers MUST have more "stuff" in it than the integers! Well, I think at this point our intuition has now shifted to believing that all of these non-finite sets are the same "size", despite "adding" and "removing" elements from them. So yes, our adjusted intuition is now correct, it is in fact true that the set of rational numbers is the same size as the set of integers and natural numbers. I will forgo an official function definition, but I illustrate the function in Figure 8. If we list all possible fractions as we do in the figure below, then we can "count" each fraction by moving diagonally. We can then assign our function from a natural number as the fraction in our diagonal count. If we follow the arrows in the diagram, then we know how to "count" the rational numbers. The function, as a result of this counting, would look like $f = \{(0, \frac{1}{1}), (1, \frac{1}{2}), (2, \frac{2}{1}), (3, \frac{2}{2}), (4, \frac{1}{3}), (5, \frac{1}{4}), (6, \frac{2}{3}), (7, \frac{3}{2}), (8, \frac{4}{1}), \dots\}$. So now we have established that:
$\aleph_0 = |\mathbb{N}| = |\mathbb{O}| = |\mathbb{E}| = |\mathbb{Z}| = |\mathbb{Q}|$

3.4 The Irrational and Real Numbers

Our intuition has failed us, and even made us change it up a bit. Just when we think that we have accounted for all the possible numbers, there are still numbers that are "missing". Take $\sqrt{2}$ for example. If we have already accounted for $\sqrt{2}$, then we should be able to find two integers $a, b \in \mathbb{Z}$ such that $\sqrt{2} = \frac{a}{b}$. To illustrate why this is not true, suppose that a and b are *relatively prime*. This means that if we take the fraction of $\frac{a}{b}$, it cannot be reduced any more, and so a and b do **not** share any common factors.

Multiply both sides of $\sqrt{2} = \frac{a}{b}$ by b and we obtain $b\sqrt{2} = a$. Now square both sides, which leads us to $2b^2 = a^2$. Therefore, by definition of an even number, a^2 must be an even number. Since the square of an odd number is always odd, we then know that a itself must be even since we just have shown that a^2 is even. Therefore, by definition, we can write $a = 2k_1$. Let's make this substitution, and we get: $2b^2 = a^2 = (2k_1)^2 = 4k_1^2$. Simplifying this expression leads us to $b^2 = 2k_1^2$. By the same reasoning as before for a, we find that b^2 is even and so therefore b must also be even. But this would mean that a and b share a common factor, namely 2. Therefore we reach a contradiction, since we assumed that a and b do not have any common factors! Hence, $\sqrt{2}$ cannot be written as a fraction of two integers, it is itself not an integer, and therefore it is a brand new number we discovered!

Definition 23 *The irrational numbers, denoted as $\mathbb{I}$, is the collection of all numbers that cannot be written as a fraction of two integers, namely $\mathbb{I} = \{x | \text{there does not exist } a, b \in \mathbb{Z} \text{ s.t. } x = \frac{a}{b}\}$*

When we combine the irrational and rational numbers, we form a new set called the *real numbers*:

Definition 24 *The real numbers, denoted as $\mathbb{R}$, is the collection of all irrational and rational numbers, namely $\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$*

3.4.1 Cardinality of the Irrational and Real Numbers

Is the size of this set of new numbers the same as the natural numbers? If they are, then we have effectively shown that $\aleph_0 = \infty$. That is, there is one and only one infinity. However, our newfound intuition is actually incorrect, and our original intuition is correct. That is, the irrational numbers (those numbers that are not rational) is actually "bigger" than the rational numbers. How do we know?

We will prove this using a proof technique that is known as "proof by contradiction". The approach assumes the opposite of our conclusion in question and tries to find a contradiction in something that we know is true. Let us assume that the size of the irrationals is the same size as $\mathbb{N}$. Then we should be able to "list" every single number on the number line (but in a non-finite manner). Assume we begin to list all of the numbers:

0.0144522
0.5667321
0.5622256
0.4003335
$\vdots$

Let us construct a new number as follows. We construct the number digit by digit by going down the supposed list of all the numbers listed here. Let us construct the first digit in our new number. The way we will construct this is by choosing a different number in the digit we are on than the same digit in the number in the list. For the first digit, we see in our list that it is 0. So let us select a different number, say 4. So our new number starts with 0.4. Now we know that our new number cannot be accounted for by the first number, since the digits are different (our number has a digit of 4, the one on the list is 0).

So we cross off the first number. Now let's move to the second number on the list. We see that the second digit in this number is 6, so let us pick a different number to use in the construction of our new number, say 3. So our new number is 0.43. We know that this number cannot be equal to the first two numbers in the list, since the first digit is different and the second digit is different. Let us continue this with the third digit in the third number on the list, which is 2. Pick a number that is not 2, say 9. Our new number is 0.439. We know for a fact that this number will never equal the first, second, or third number on our list, and therefore, this number is not in the list. If we continue this construction of a number in the same way throughout all the number in our list, we find that it is impossible to have this number accounted for in the list. This is a contradiction, since we assumed that our list contained all the possible numbers, rational and irrational. Therefore, we conclude that there must exist a set of all numbers where the size of the set is different than $\aleph_0$.

This is SHOCKING. This means, that we do not just have only one type of "infinity", but two! In actuality, there are an infinite number of types of infinity (don't stress, we're not going that far down the rabbit hole!). The type of infinity that we can't "count", which was illustrated with this problem of counting all possible numbers, is what is known as *uncountable*. The set of *real numbers* is the union of rational and irrational numbers, $\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$. We have just proven that $|\mathbb{R}| \neq |\mathbb{N}|$. This means, that despite the natural numbers and the real numbers both being non-finite, they are different kinds of non-finite. The natural numbers are referred to as a *countable set* (that is, we can list and count the elements in the set). The real numbers are referred to as a *uncountable set* (that is, we **cannot** list and count all the elements in the set).

3.5 Addition over Uncountable and Countable Sets

When we perform arithmetic, we typically do so with *sets* of numbers. Think about it. If I told you to add the ages of 5 people in the room, then what you are really doing is *counting*. Think about the 2nd grade (I know, this is tough!). When you first learned addition, you learned that it was a short hand way to count. Saying 3+2 is the same as counting 1,1,1,1,1. Addition is a short hand way to count the two numbers in total.

When I tell you to "add" the ages of 5 people in the room, you are effectively counting the total number of years of life among all the people. The reason you can compute this computation is because the set of ages of five people is *countable*. Suppose we had 30,24,13,45,20 as the ages in the room. We can represent these ages as a set, namely $A = \{30, 24, 13, 45, 20\}$. When I tell you to add these together, then you are iterating over every element in the set and adding their ages together: 30+24+13+45+20. You can do this, because it is *countable*.

Now suppose I told you to add together the number 1 for all of the numbers in the interval $(0, 1)$. Obviously, we can't do this with our typical way, since the set $(0, 1)$ is an uncountable set. In contrast to our previous example, we cannot say something like $1 + 1 + \dots$, because explicitly writing all of the numbers in our sum that fully enumerates the interval $(0, 1)$ would mean that we could "count" the numbers in $(0,1)$, which is a contradiction, since $(0,1)$ is uncountable. So then, how can we "add" together all of these numbers in the interval $(0,1)$?

The answer lies in *approximation*. Think about the operation of multiplication. Again, when we learn this in the 3rd grade, we learn that if we add the same number a certain number of times, then we can find this via the operation of multiplication. Abstracting this idea to adding the same number over an uncountable number of times, then our operation of addition turns to multiplication. So adding the number 1 for all of the numbers in the interval $(0,1)$ amounts to adding 1 a total of 1 times, which of course leads us to 1. To see how we could have approximated this, suppose we split the interval into 10 parts, each part comprising of length $\frac{1}{10}$. Then when we add two of these parts, we're adding the number 1 a total of $\frac{1}{10}$th of a time together with another $\frac{1}{10}$th of a time and so on and so forth.

So if we approximate the total number of 1's over the uncountable interval $(0,1)$ with 10 intervals, we are adding the numbers
$A_{10} = \{\frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1), \frac{1}{10}(1)\}$ together. We can see that doing this for any n leads us to $\{\frac{1}{n}(1), \ldots, \frac{1}{n}(1)\}$. For each n, the sum is 1. So we can see that if we take n to ∞, we just get a sequence of numbers that converge to 1.

What we have illustrated here is that if we want to add the same number an uncountably infinite number of times, then the way in which we do this is by creating successive approximations of countable sums, looking at the sequence of approximations, and seeing where the sequence converges. We will revisit this idea when we discuss integrals in a later lecture.

4 Functions

In the previous section we discussed basic set-theory. We will find the tools there very useful in future topics that we will discuss. The mathematics we will employ is often very lengthy and involves many different cases, sometimes an infinite number of cases. Set-Theory allows us to summarize complicated statements about collections of numbers in a concise form. In more advanced management science courses, set-theory also allows us to prove certain theorems when designing models. This is a bit outside the scope of this course and we will not use set theory for these purposes.

Now that we have a fundamental understanding of the theory of sets, relations and functions, we can shift our focus to study functions from a different perspective. You should have learned in algebra and pre-calculus that functions can be defined on the real numbers $\mathbb{R}$. In this section, we will summarize some of these topics. We first will discuss how functions can be represented. We saw in the previous section that functions are simply mappings between elements of two sets. How this mapping is expressed, however, can vary depending on the application. In our course, functions are commonly expressed using equations.

With that said, there are many different types of functions which can be expressed as an equation. We will offer a review of these different types of function. After, we will discuss a way in which we can reverse the mapping to express the relationships between sets but in the opposite direction. Such relations are commonly referred to as *inverse functions*. Last we will discuss how functions between two sets of numbers on the number line can be graphed. Graphing is an important tool towards visualizing the behavior of different types of functions. They can, however, be deceiving since the scale at which these functions are mapped are usually determined by the individual drawing the graph.

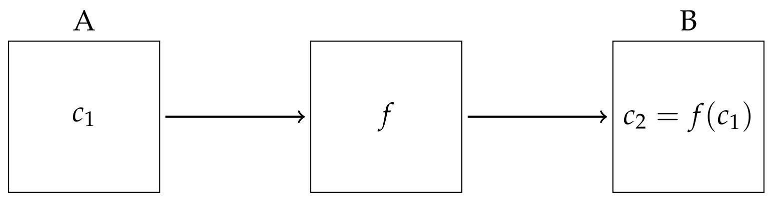


Figure 9: Function Acting as a ”Machine”

4.1 Different Representations of Functions

We had seen in the previous section that functions are special types of mappings of elements between two sets. These types of mappings provide very powerful applications to real world problems. Many times in the real world we have scenarios that have the notion of ”inputs” being sent through some type of operation and as a result produce a single ”output”. For example, if we sell a product at a certain price and purchase a product at a certain cost, then when we input the ”number of products purchased ” and the ”number of products sold” we output a ”profit”. Every unique level of demand and order quantity produce a single output. That is, it is impossible for us to make two different profits ordering and selling the same quantity of products.

Functions allow us to model events in the real-world. Typically they are expressed in terms of operations on numbers. Functions need not only be expressed using the mapping diagrams and explicit set notation as we had seen in the previous section. Sometimes it is more useful for us to express functions as equations. If we change our viewpoint from ”mapping” elements to ”manipulating” elements then we can gain some more intuitive ideas of how functions operate on elements of a domain.

We can certainly think of functions as how we defined them: as mappings. Taking elements from one set of numbers (or more generally ”things”) and assigning them to elements from another set of numbers (again, or ”things”) is a common practice in the real-world, as we have demonstrated with the examples above. We can think of functions, however, in a different yet equivalent way. Since functions, by definition, only map one and every element from the domain to a single element in the codomain, we can think of functions as ”inputting” something into a ”machine” which manipulates the input to produce a single ”output”. This is equivalent to the notion of ”mapping” or ”assigning” elements, but it is more natural to think about how functions work in this manner. We can see this process illustrated in Figure 9

Generally, the ”manipulation” (or mapping) can be defined in a variety of different ways:

1. Describe the manipulation of every possible element in the domain verbally

2. Describe the manipulation of every possible element in the domain with a set of ordered pairs

3. Describe the manipulation of every possible element in the domain with an equation or multiple equations

4. Describe the manipulation of every possible element in the domain with a mapping diagram

5. Describe the manipulation of every possible element in the domain with a graph describing the inputs and outputs

In more practical scenarios we tend to describe functions as equations. Doing so allows us to perform algebra on the inputs, easily graph functions and easily calculate outputs (rather than have to refer to a table of inputs to outputs, which can grow quickly if the domain is large). For example, we could write:
$f\colon \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ such that
$f(x) = x^2$

We then can calculate an output when provided an input $x \in \mathbb{R}$ by replacing every occurance of x in the function equation with the input and then perform arithmetic to solve for the output. For example, we can express and calculate the output of this function when the input is 4:
$f(4) = (4^2) = 4 \cdot 4 = 16$

The key point here is that whenever an input is provided (which sometimes could be another equation itself), we need to replace every occurance of the input in the function equation with the provided input. For example, suppose the input to the function were another function expressed by the equation $g(x) = e^x - \sigma$. We would then have the output expressed as:
$f(g(x)) = f(e^x - \sigma) = (e^x - \sigma)^2 = e^{2x} - 2e^x\sigma + \sigma^2$

The ability to express functions as equations offers us a very powerful tool not only to represent events in the real-world but allows us to easily manipulate different functions algebraically. Since functions can be expressed as algebraic equations, and since we have different types of equations, we therefore can classify different types of functions based on how the equation itself looks. In the following sections we will review through a few different types of functions that we will likely encounter in this course. This list is by no means a complete enumeration of all the different types of functions. There exist many many different types. The other types that fall outside this list, however, rarely appear in management science applications and therefore we will omit a discussion on them.

4.2 Types of Functions

4.2.1 Constant Functions

By far the most trivial type of function is the constant function. The constant function is defined as taking all the possible inputs in the domain and providing as an output the same number.

Definition 25 *Let $A, B \subseteq \mathbb{R}$, let $a \in B$ and f a function defined on $f\colon A \to B$. Then f is said to be a* constant function *if it's mapping is defined as:*
$f(x) = a$

Since every input in the domain is being mapped to the same output in the codomain, then the graph of any constant function is just a horizontal line. This can be seen in Figure 10

Example 16 *Let the constant function f be defined as $f(x) = 6$. Then we have the following outputs for the following inputs:*
$f(4) = 6$, $f(1) = 6$, $f(1) = 6$
As we can see, no matter what the input is, the output is always 6.

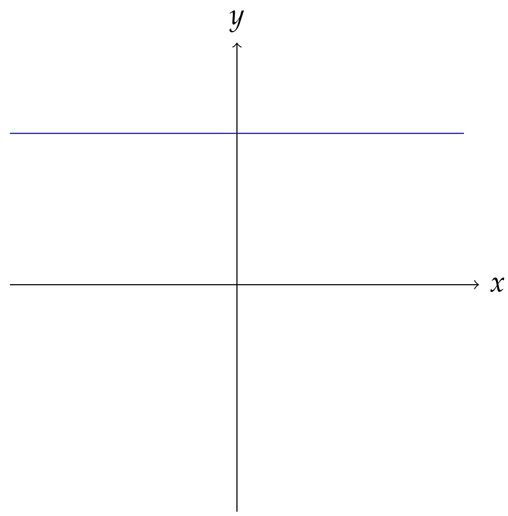


Figure 10: A constant function. The blue line is at the value of $f(x) = a$

4.2.2 Linear Functions

The next of simple types of functions is the class of functions known as the *linear* functions. They are called linear because when they are graphed the resulting figure is a straight line. Linear functions have two primary components to them: the y-intercept and the slope. Any straight line can be defined by first identifying these two numbers.

Definition 26 *A* y-intercept *is the value of a linear function when the input is equal to 0. Visually, this is the point where the straight line intercepts the vertical axis in the coordinate system. From a practical perspective, the y-intercept represents the "starting value" in a real-world representation.*

Definition 27 *A* slope *is the value that represents the change in the output when the input is changed by a value of 1. This usually represents a marginal quantity in a real-world application.*

Definition 28 *A* linear function *is a function f defined by the following equation:*
$f(x) = a_0 + a_1 x$

where a_0 is the y-intercept and a_1 is the slope

We can see from Figure 11 that the blue line represents the graph of the function $f(x) = a_0 + a_1 x$. The definition of the slope a_1 represents how much the y-value changes when the x-value changes from a point by 1. We can find a_0 by picking two points and calculating the x and y differences between the points. We see this illustrated in Figure 11 where $\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$. We therefore would have $a_1 = \frac{\Delta y}{\Delta x}$. The y-intercept can be found by finding $f(0)$ since $f(0) = a_0 + a_1(0) = a_0$. Hence, the equation of a line is completely determined by the values given for the slope and intercept. As illustrated in the Figure, we also can see that the equation of a line can be found by specifying two points $(x_1, y_1), (x_2, y_2)$, finding the slope and then solving for the intercept using one of these two points and the slope.

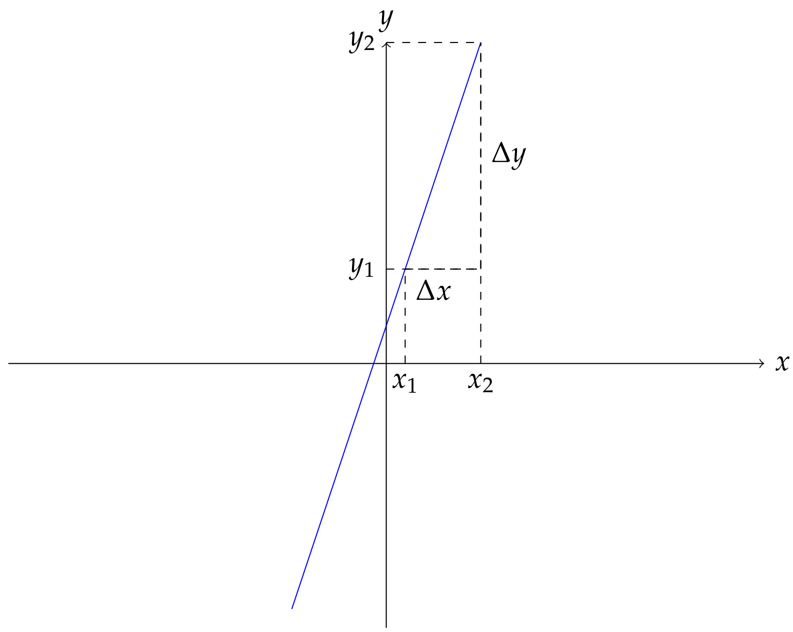


Figure 11: A Linear Function.

Example 17 *Find the equation of a line that goes through the points* $(1, 2)$ *and* $(5, 6)$.
We first can find the slope:
$a_1 = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{6-2}{5-1} = \frac{4}{4} = 1$
Now we can find the y-intercept by using the general equation of the line and plugging in the values for one of the points given:
$y = a_0 + a_1 x$
$(2) = a_0 + (1)(1)$
$2 - 1 = a_0$
$a_0 = 1$
Hence the equation of the line that goes through these two points would be
$y = a_0 + a_1 x = 1 + (1)x = 1 + x$

Linear Functions are an important topic for our field of study. As we will see later, we can generalize these to involve more than one variable. When we get to linear regression models, we typically have the goal in mind to find the equation of a line that "best fits" a given set of (x, y) values from a data set. We will see that "best fit" is often a subjective term, since many different ideas of "best fit" exist. The important point here is that linear functions play a very significant role in understanding relations between variables within a given set of data that was observed.

4.2.3 Polynomial Functions

In the previous two sub-sections we discussed two types of common functions that are found in practice. The constant function is very useful for describing certain events where the inputs don't any effect on the outputs (other than being assigned to a single number). The linear function is quite useful in statistical applications. The classical Linear Regression Model allows Management Scientists (as well as others) to describe potential causal relationships between factors using a linear equation. This type of association is quite frequent in nature, hence it's importance. One of the most useful types of functions not only in application but also in theory is the *polynomial function*. This function is a generalization of the previous two types of functions discussed where each is just a special case of a more general polynomial type of function.

If we take a look at the basic equation for a linear function, we notice that if the slope $a_1 = 0$, we get as a result the constant function. Thus, the constant function is a special case of a linear function. Similar to how we set one of the numbers equal to 0, we can do the same for a more general equation in order to get the equation of a linear function. This general equation is of the form defined below.

Definition 29 *A* polynomial function *is a function that has the following equation:*
$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$
where every $a_i, i \in \{1, \ldots, n\}$ *is called a* coefficient

The highest exponent that appears in a polynomial function is quite often a topic of interest. This number often allows us to determine the number of *roots* a polynomial has. Root analysis is important in the theory of optimization, as we will see in future topics.

Definition 30 *A* root *of a function* $f\colon A \to B$ *is a value* $a \in A$ *such that* $f(a) = 0$*. That is, any root of* $f(x)$ *is any value in the domain such that* $f(x) = 0$*.*

Definition 31 *The* degree *of a polynomial function is the highest value exponent in the function. That is, if* $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$*, and* $a_n \neq 0$ *then* $f(x)$ *is said to have degree* n*.*

There is an interesting theorem from algebra that connects polynomials to the notion of degree and root. This theorem allows for a complete accounting of the roots of a polynomial.

Theorem 7 *The Fundamental Theorem of Algebra*
Let $f(x)$ *be an* n-degree *polynomial. Then the function* $f(x)$ *has* n *roots with multiplicity. That is, the total number of roots a polynomial has is* n*, where more than one of the same root might exist.*

Example 18 *Let* $f(x) = x^2 + 2x + 1$*. We first can observe the graph of this function in Figure 12. We notice that a root must exist at* $x = -1$ *since this is where the function equals 0. We notice something interesting however. We notice that the function does not cross the x-axis twice. Yet the fundamental theorem tells us we must have 2 roots. You will notice that this function can be factored into a more compact form:* $f(x) = (x+1)^2 = (x+1)(x+1) = x^2 + 2x + 1$*. Notice that the only way for* $f(x) = 0$ *we must have* $(x+1)(x+1) = 0$*. This can happen at either of these factors, even though they are the same. Hence, the root* $x = -1$ *has a* multiplicity *of 2. That is, the same root appears twice.*

Polynomials are flexible tools for practical applications. If you had taken a Calculus course, you had seen their ability to express any type of *differentiable* function as a polynomial. Management Scientists also use them for more complicated types of data analysis. Computer programmers also use them for their mathematical programs to perform complicated calculations in a shorter amount of time. Finally, polynomials allow us to determine types of algorithms (step-by-step procedures) that can be calculated in a reasonable amount of time, where *reasonable* is typically defined in terms of polynomials. To get a feeling for how the roots of polynomials work and how they are related to the degrees of polynomials, we have a few different polynomial functions graphed in Figure 13, each of a different degree.

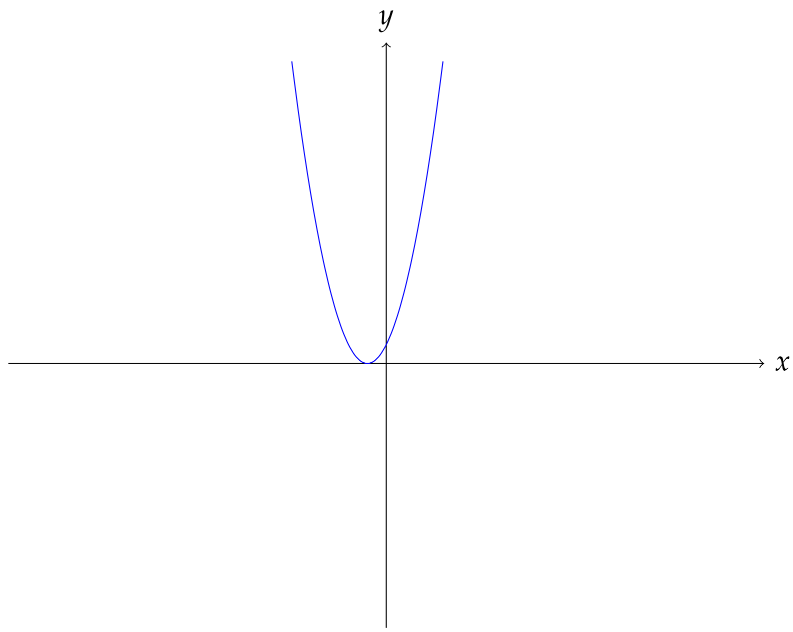


Figure 12: A Polynomial of Degree 2.

4.2.4 Power Functions

Another common function we will encounter in our studies is the *power function*. This type of function has a few different cases which we will briefly review below.

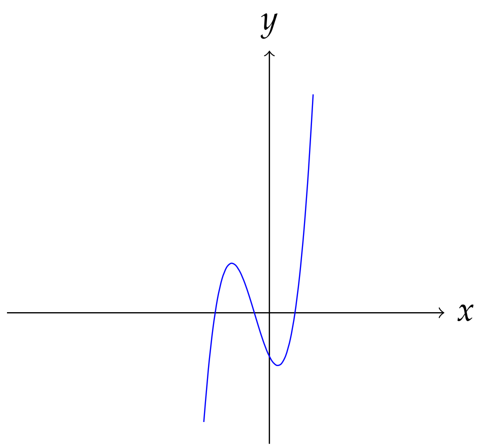
Definition 32 *The function $f(x) = x^a$ is called a* power function*, where $a \in \mathbb{R}$*

We have a few different cases for what the power could be. When the power is in one of these categories, the graph of the function tends to differ in appearance and behavior than the other types:

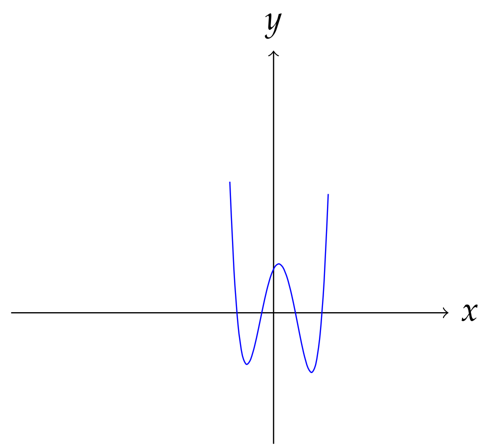
1. a is a positive integer
2. $a = \frac{1}{n}$ where n is a positive integer
3. $a = -1$

If a is a positive integer then the power function $f(x) = a$ is just a special case of a polynomial function. Figure 14 shows different graphs of functions for different positive integers of a. We can see an interesting pattern emerge. The blue lines represent functions when the power is an odd number. In the figure, the blue lines represent the power function when the power is 1,3,5 or 7. We notice that they all are very similar to each other in shape. We also notice the same pattern for the even integers. Looking at the red lines, which represent power of 2,4,6, we notice that they are all similar in shape. The power functions share this pattern for all even/odd powers.

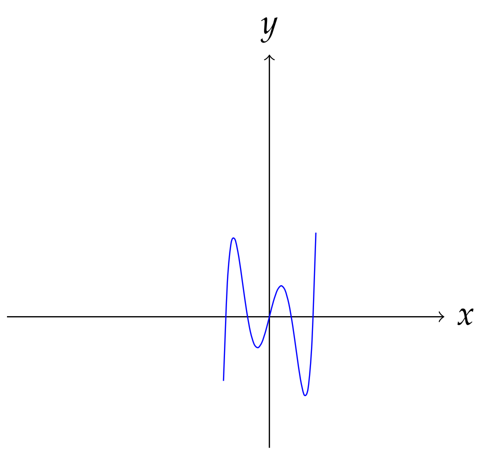
Recall that the other case for a is when $a = \frac{1}{n}$. These are commonly referred to as *root* functions, since $f(x) = x^{\frac{1}{n}} = \sqrt[n]{x}$. When n is an even number, then the domain for the root functions is always $[0, \infty)$. On the other hand, if n is odd, the domain is $\mathbb{R}$. This is so since it is possible to attain negative numbers when taking a negative number to an odd power. For



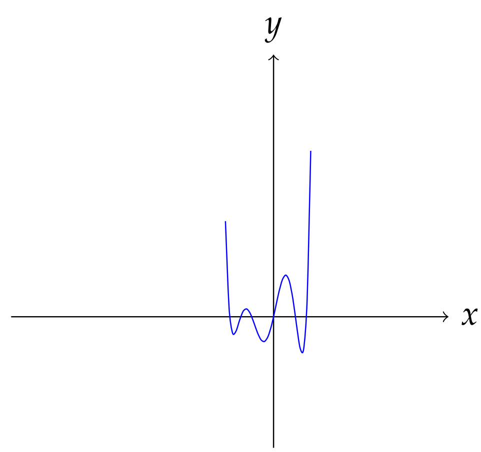
(a) Polynomial of degree 3



(b) Polynomial of degree 4



(c) Polynomial of degree 5



(d) Polynomial of degree 6

Figure 13: Different Polynomial Functions of Varying Degrees

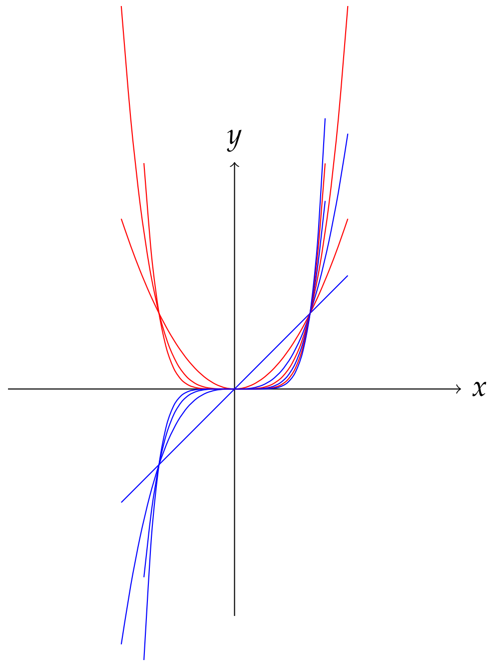


Figure 14: Different Power Functions When a is a Positive Integer. Red Graphs are When Powers are Even Integers, Blue Graphs are When Powers are Odd Integers.

example, $(-3)^3 = (-3)(-3)(-3) = (9)(-3) = -27$. Hence, $\sqrt[3]{-27} = -3$. It should be easy to see that odd numbered roots of negative numbers should also be negative. This is due to the fact that the number is first multiplied by itself an even number of times, and there is one last number (itself) to multiply the total factor to (the even powered). That is, if $b < 0$, and c is an odd number, by definition $c = 2 \cdot d + 1$ and so $b^c = b^{2 \cdot d + 1} = b^{2 \cdot d} \cdot b$. We know that any number taken to an even number is positive, and since $b < 0$, it therefore must be the case that $b^c < 0$. Figure 15 has a few different *root* functions that are graphed with different powers. Red plots are the odd powers and blue are the even.

The last case for power functions is the case where $a = -1$. These type of functions are commonly referred to as the *reciprocal function*. This is due to the fact that $x^{-1} = \frac{1}{x}$. The reciprocal function is defined on the domain $\mathbb{R} - \{0\}$. That is, all numbers except $x = 0$. While we can say that as $x \to 0$ the function either approaches $-\infty$ or ∞, depending on how x is approaching 0, we cannot say that the function itself is defined there. Important applications are relevant to the reciprocal function. For example, suppose K is a fixed cost of running a machine and x is the quantity of products to produce (this would be a decision in our example). Then the function $f(x) = \frac{K}{x}$ represents out unit fixed costs for $x > 0$. The graph of $f(x)$ is shown in Figure 16.

4.2.5 Rational Functions

Sometimes, although rare, we need to express functions for practical scenarios where the function structure is not a polynomial. Sometimes the structure of a function is the ratio between two polynomials. Such functions are commonly referred to as rational functions.

Definition 33 *Let $P(X)$ and $Q(x)$ be two polynomial functions. Then $f(x) = \frac{P(x)}{Q(x)}$ is called a* rational function. *It's domain is every number in the real numbers except those where $Q(x) = 0$. That is, the*

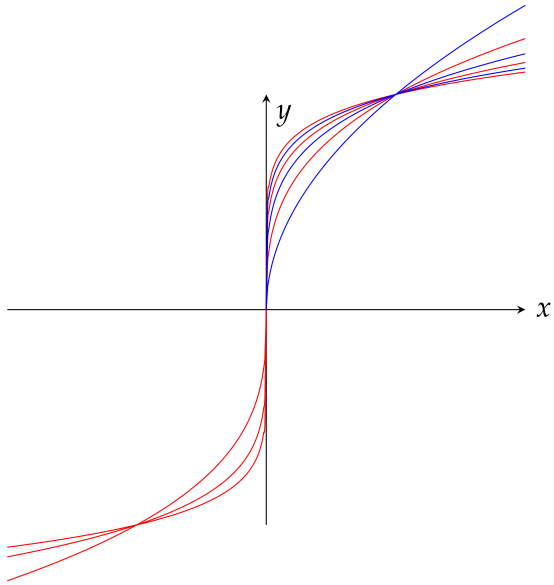


Figure 15: Different Root Functions When $a = \frac{1}{n}$. Blue Graphs are When n is an Even Integers, Red Graphs are When n is an Odd Integers.

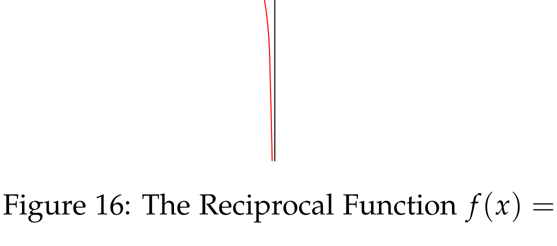


Figure 16: The Reciprocal Function $f(x) = \frac{1}{n}$

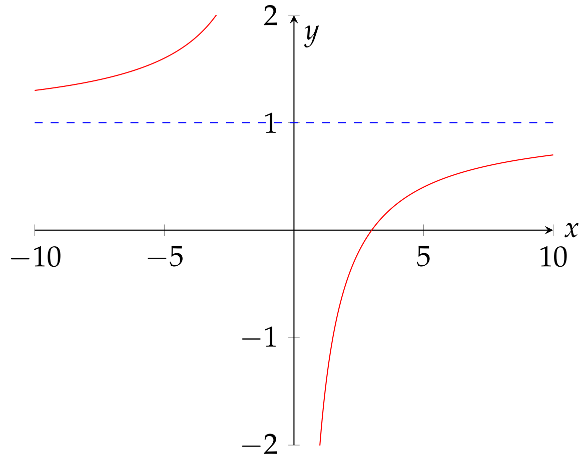


Figure 17: The Probability of Stocking Out when the Price is set to p

domain of $f(x)$ are all real numbers except the roots of $Q(x)$.

Example 19 *The popular newsvendor model is used to identify optimal ordering quantities when the demand is a random variable and follows a probability distribution. The basic model comprises of a demand for a particular product that is sold at p dollars and can be purchased at c dollars. Suppose that $c = 3$. In the model, the function $f(p) = \frac{p-c}{p} = \frac{p-3}{p}$ represents the probability of remaining in stock (that is, the probability of having enough inventory to satisfy all demand for a given time horizon) when the price is set to p dollars. This is an example of a rational function and is defined everywhere except when the price is 0 (although negative prices do not make much practical sense, it is possible to define this function over negative prices). The graph of this function is shown in Figure 17. As we can see, as we increase the price, the probability of being in stock increases and converges to 1 when we charge a price of ∞. We also notice that the probability of being in stock drops to 0 when we set the price exactly equal to the cost (in this example, 3).*

4.2.6 Algebraic Functions

We have now covered the basic functions which are all analogous to basic arithmetic operations such as addition, subtraction, multiplication and division. There are some functions, however, that are combinations of the previous functions we have discussed. If a function has an equation that is a combination solely of polynomials, rational functions and power functions then we typically call these functions *algebraic functions*.

Definition 34 *An* algebraic function *is a function that has as it's equation a combination of polynomials, rational functions and power functions.*

Example 20 *A few examples of algebraic functions are listed below:*

$f(x) = \sqrt[3]{\frac{x^3-1}{4x+3}}$

$f(x) = \frac{4 \cdot x^4 - (x+2)^3 + 1}{\sqrt{6x^2-2}}$

$f(x) = \sqrt[5]{5x-2} + \frac{1}{1-\sqrt{8x^5}}$

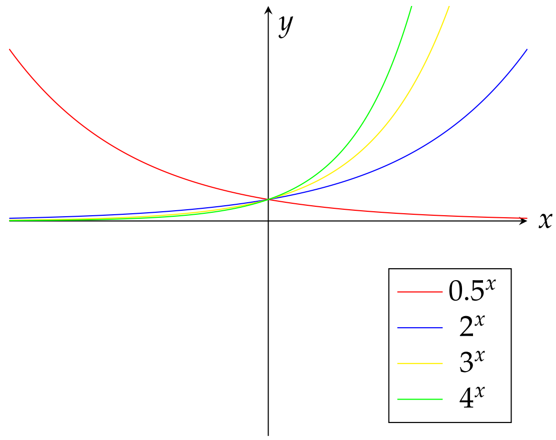


Figure 18: Different exponential functions

4.2.7 Exponential Functions

Yet another very common type of function which is employed in many different types of applications is the exponential function. Exponential functions describe rapid growths of whatever is being modeled. The notion of "rapid growth" is quite a common one. A few examples include computing compound interest, performing cash flow analysis, and determining the consequences of capacity expansion.

Definition 35 *An* exponential function *is a function that describes the rapid growth of a natural event. The general equation of an exponential function is*
$f(x) = a^x$
where $a > 0$.

The number a is commonly referred to as the base.

We can see that the larger the number a is, the faster the function will grow with small changes in x. Figure 18 illustrates different exponential functions for different bases. We notice an interesting association from the figure. First, when $a < 1$, the function converges to 0 as $x \to \infty$. Second,when $a > 1$, the function converges to ∞ as $x \to \infty$.

Example 21 *Suppose that an individual puts \$ 1,000 into a savings account that yields 5 % interest per year which compounds yearly and does not add any additional funds. What will be the balance of his account by the end of 10 years?*

Here we would like to compute the compounded interest. We are told that as each year passes, 5% of the balance is gained in interest. Hence, by the end of year 1, the individual would have the original \$1,000 plus the interest, which would be $1000 + (0.05) * 1000 = 1000(1 + 0.05) = (1000)(1.05)$. *At the end of the second year, the individual will have what they started with plus the interest gain in the first year. That is, the balance at the end of year one will be* $(1000)(1.05)$. *The end of year two will see a balance of the starting balance of* $1000(1.05)$ *plus the 5% interest of the balance. So the end of year two we see that the individual will have the original balance of* $(1000)(1.05)+$ *interest. In this instance the ending balance at the second year will be* $1000(1.05) + (1000(1.05))(0.05) = 1000(1.05 + 1.05(0.05)) = 1000(1.05)(1 + .05) = 1000(1.05)^2$. *We hence see a pattern emerge. Let x denote the number of years. Then the ending balance at year x of the account can be described by the exponential function* $f(x) = 1000(1.05)^x$. *The graph of this can be seen in Figure 19.*

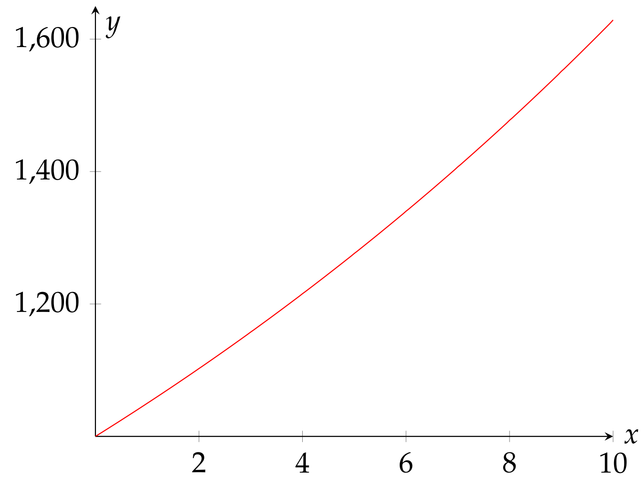


Figure 19: Compound Interest with a Starting Balance of \$1000

Example 22 *The previous example allows us to generalize the formula for compound interest. Let x be the number of years that passes by, r denote the interest rate, B denote the starting balance of an account. Then the balance after x years pass is described by the formula:*
$f(x) = B(1+r)^x$

4.2.8 Logarithmic Functions

We have seen the notion of exponential functions and their potential to explain natural events that exhibit relative quick growth patterns. The exponential functions provide a better mechanism than linear functions in trying to explain these types of events. Sometimes, however, we would like to explain events in the natural world whose values are very large. At times we need to have the values of different variables on the same scale. For example, it is difficult to explain the behavior of stock returns, which are often expressed in values that have a magnitude of 10^{-2} using the volume of those returns, which is often in a magnitude of 10^5. Explaining events that have as their data large values is often difficult not only to calculate but also to understand.

A common tool employed on data that has large values is that of the *logarithm*. Put simply, a *logarithm* is the opposite of the *exponential*. We had seen in the last section that when given a base and an input to the function, which is interpreted to be the exponent, that we can get large output values with small numbers being inputted as the exponent. The *logarithm* reverses this process. That is, when given a large value and a base, the *logarithm* of that large number is the exponent that the base is taken to in order to arrive at the large number in calculation.

Let us take an example to motivate the definition. We know from basic algebra that $2^3 = 8$. In this small example, we are given a base of 2 and an exponent of 3 to arrive at the number 8 when calculated. If we ask the question, "what value of y results in 8 when 2 is taken to this value", then mathematically this question can be phrased in the form of an equation, $2^y = 8$. Our minds will then attempt to see how many times we must multiply the value of 2 in order to arrive at 8. Let us try another small example. What is the value of y such that $10^y = 10000$? We clearly can see that multiplying 10 by itself 4 times will result in a value of 10000. These values of y are what is commonly referred to as the *logarithm* of a number when given a base.

Definition 36 *Let $a > 0$ and $x > 0$. Then the* logarithm *of x with respect to base a is defined to be the*

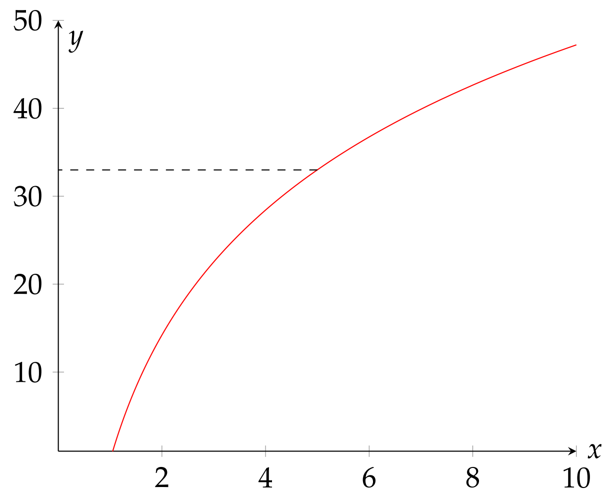


Figure 20: The Logarithm Function with base 1.05.

exponent $y = log_a(x)$ *such that*
$a^y = x$.

Logarithms help take very large numbers and express them in terms of their exponents with respect to a single base. They also help us solve some practical problems that involve exponential growth but our inquiries are about the opposite of what we had discussed in the previous section.

Example 23 *Suppose that an investor has a starting balance of $1. The investor would like to grow his account to $5 and decides to invest in an equity that yields 5% per year. It is assumed that interest is compounded yearly. How many years will it take for the investor to get to his goal?*

We can use an exponential function to describe the growth of the investors account. Let $f(x) = (1.05)^x$ *denote the balance amount over the course of x years. We would like to solve the equation* $5 = (1.05)^x$*. This is equivalent to finding* $log_{1.05}(5)$*. We can use a calculator to find this. Doing so gives us approximately* $y = log_{1.05}(5) = 32.99$ *Therefore, the investor would have to wait a minimum of 32.99 years to reach his goal. Rounding down, we can see that* $1.05^(32.99) = 5$*. Figure 20 shows the logarithm function with base 1.05.*

Similar to how exponents have algebraic laws that allow for simplification, logarithms have laws that allow us to easily manipulate them. Below is a list of these different laws.

Theorem 8 *Logarithm Laws*
Let a be a base, b and c be positive numbers and $r \in \mathbb{R}$ *. Then*
$log_a(b \cdot c) = log_a(b) + log_a(c)$
$log_a(\frac{b}{c}) = log_a(b) - log_a(c)$
$log_a(b^r) = rlog_a(b)$

We also have the ability to change bases when it is inconvenient or difficult to directly calculate logs with a given base:

Theorem 9 *Change of Base Formula*
Let $a > 0$ and let $b > 0$. Then we have
$log_a(x) = \frac{log_b(x)}{log_b(x)}$

Last, the exponential and logarithm functions are related by definition. He would hence have:

Theorem 10 *Let $a > 0$ and $x > 0$. Then the following is true: $log_a(a^x) = x$*
$a^{log_a(x)} = x$

These laws allow us to solve more complicated problems involving exponential growth and logarithms.

Example 24 *Suppose an investor is trying to identify the best asset in which to invest. The investor only has $1000 to start and would like to turn his investment to a balance of $2500 in 2 years. If interest is compound monthly, what interest rate should the investor look for?*

We are told that the investment is compounded monthly. Therefore the investor would like to earn 2500 in 24 months. Using our exponential function from earlier, we would have the following equation to solve: $2500 = (1000)(1+r)^{24}$. We can use the log laws to solve for the rate by rearranging and taking the log of both sides (any base will work, but we will use base 10):
$2500 = (1000)(1+r)^{24}$
$\frac{2500}{1000} = (1+r)^{24}$
$log_{10}(2.5) = log_{10}((1+r)^{24})$
$log_{10}(2.5) = 24log_{10}((1+r))$
$\frac{log_{10}(2.5)}{24} = log_{10}((1+r))$
$10^{(\frac{log_{10}(2.5)}{24})} = 10^{(log_{10}((1+r)))}$
$10^{(\frac{log_{10}(2.5)}{24})} = 1+r$
$10^{(\frac{log_{10}(2.5)}{24})} - 1 = r$

Using a calculator we find that $10^{(\frac{log_{10}(2.5)}{24})} - 1 = 1.038917 - 1 = 0.038917$ Reversing the process to double check our work, we see that $(1000)(1+(0.038917))^{24} = 1000(1.038917)^{24} = 2500.003$

One last important tool that is commonly employed is that of the *natural base e*. The number e is an irrational number that is roughly $e \approx 2.71828$. We will see later that e has special properties when it comes to calculus. When taking a log with respect to e, we use the notation $ln(x) = log_e(x)$. We hence have two very important functions that are often employed in practice:

Definition 37 *The exponential and logarithmic function with base e is defined as*
$f(x) = e^x$
$f(x) = ln(x) = log_e(x)$

Their graphs are displayed in Figure 21

4.2.9 Piece-wise Functions

Up to this point we have reviewed through most of the functions that we will encounter in this course. At times, however, we might require specially defined functions that are not necessarily

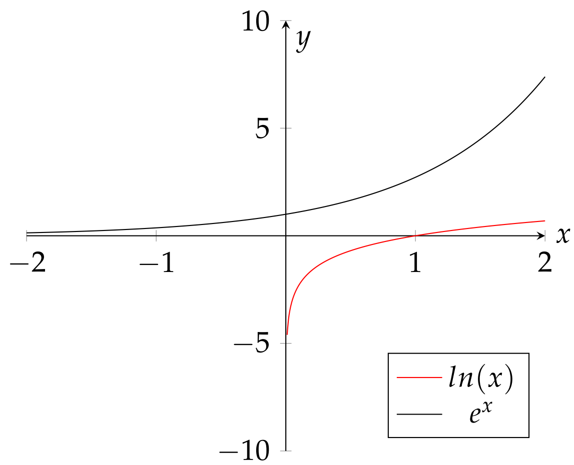


Figure 21: The functions $f(x) = e^x$ and $f(x) = ln(x)$.

continuous. You can think of a continuous function as one whose graph can be drawn over the entire domain without having to lift a pencil. With the possible exception of rational and algebraic functions, nearly all the functions we discussed were continuous. The need often arises in practice to define functions over different domains. We had seen that functions can be defined by equations. Do not loose sight, however, of the basic definition of a function: mapping points from one set to points of another set.

It is possible to take the set under consideration and "break it up" into different subsets. After that is done, we could define a mapping from one set to another set by passing through numbers belonging to different subsets to different equations. Such functions are often referred to as *Piece-wise Functions*. We will take a small example to motivate this definition:

Example 25 *A production facility charges a fixed cost of \$1000 for every time the machines are switched on for production. This facility, however, offers a discount to customer that order in excess of 500 units. They will charge a fixed cost of only \$100. The per-unit fixed cost for order sizes less than or equal to 500 units would be* $f_1(x) = \frac{1000}{x}$ *and the per-unit fixed cost for order sizes greater than 500 units would be* $f_2(x) = \frac{100}{x}$. *Therefore, the per-unit fixed cost function when ordering x units would be*

$$f(x) = \begin{cases} \frac{1000}{x} & x \leq 500 \\ \frac{100}{x} & x > 500 \end{cases}$$

The graph of the function is shown in Figure 22

Piece-wise functions allows us to define different equations for different areas in the domain. This is a powerful tool, as illustrated in the above example. It can become, however, quite cumbersome when we perform calculus on these functions. We will omit the complexities and problems these types of functions cause for us. It is important to understand, however, as we will see later, that these types of functions will not always allow for easy analysis in calculus.

We have the general definition of a piece-wise function on the real number line below:

Definition 38 *Let* $A \subseteq \mathbb{R}$ *and let* $A_1, A_2, \ldots, A_n$ *be subsets of* A *such that* $A_1 \cup A_2 \cup \cdots \cup A_n = A$ *and* $\forall i, j \in \{1, \ldots, n\}$ $A_i \cap A_j = \varnothing$. *Also let* $\forall i \in \{1, \ldots, n\}, f_i \colon A_i \to \mathbb{R}$ *be functions on each of these*

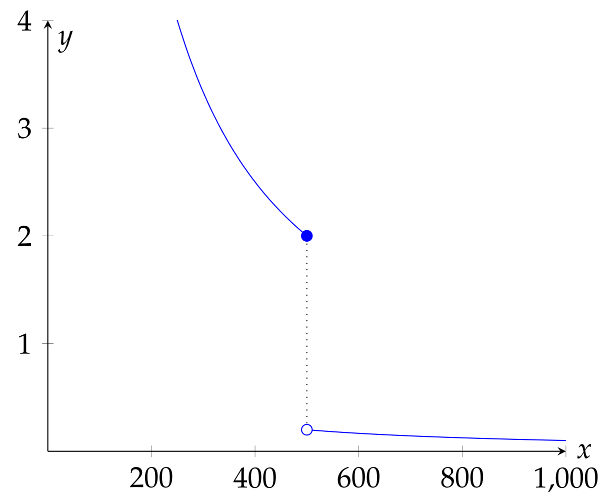


Figure 22: The Piece-wise function $f(x)$.

sets. Then the piece-wise function $f\colon A \to \mathbb{R}$ is defined as

$$f(x) = \begin{cases} f_1(x) & x \in A_1 \\ \vdots & \vdots \\ f_n(x) & x \in A_n \end{cases}$$

The definition states that if we wish to define a function on a domain A differently, then we must ensure that we break up the domain into disjoint sets. That is, no number can be in two different subsets (that was the empty intersection term in the definition). Furthermore, once we break the domain into different sets, we must define a function for each subset. The overall function value for f is then determined based on the set that the number of interest belongs to.

In the example above we saw that if a number such as 400 was used as input, then we would use the first formula, since 400 belongs to set of numbers less than or equal to 500. So in that case our function would be $f(400) = f_1(400) = \frac{1000}{400} = 2.5$. On the otherhand, if the value of the input is greater than 500, such as 600, we would use the second function: $f(600) = f_2(600) = \frac{100}{600} = \frac{1}{6}$.

One more point to add to piece-wise functions. We notice that in the example graph that there is a solid and an open circle on the graph. The solid point denotes the fact that the point belongs to the graph just before that point. The open point denotes the fact that the point does *not* belong to the graph the point is on. So in our example, the open circle on $x = 500$ indicates that the point $(500, 0.2)$ does not belong to the graph of that function. That is, it is not a point which can be found in the function definition. The point $(500, 2)$ on the other hand is solid, which indicates that the point $(500, 2)$ belongs to the function of the graph that it is located on. The idea is to emphasize the notion of membership.

4.3 Increasing and Decreasing Functions

Now that we have discussed different types of functions as defined by equations, we have a few properties that can discuss about general functions. The first is the notion of increasing and decreasing functions.

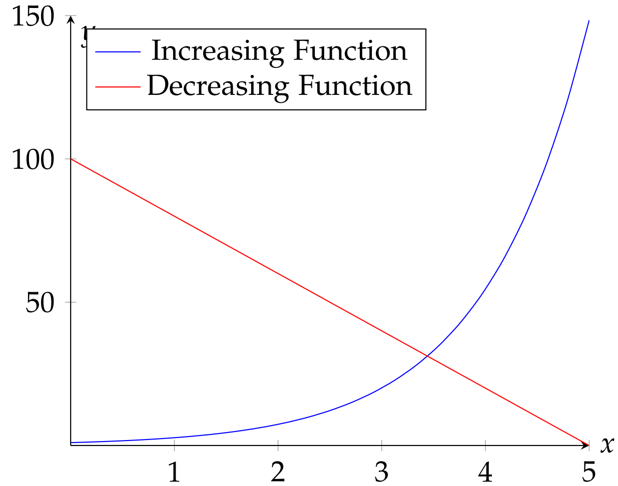


Figure 23: Increasing and Decreasing Functions.

Definition 39 *A function $f\colon A \to \mathbb{R}$ is said to be* increasing *if $\forall x_1, x_2 \in A$ such that $x_1 \leq x_2$ then $f(x_1) \leq f(x_2)$.*
A function $f\colon A \to \mathbb{R}$ is said to be decreasing *if $\forall x_1, x_2 \in A$ such that $x_1 \leq x_2$ then $f(x_1) \geq f(x_2)$.*

In other words, a function is increasing if when we increase values of x starting from the smallest number in the domain (sometimes $-\infty$), then the function $f(x)$ increases in value and never decreases. The opposite is true of decreasing functions. That is, if we start at the smallest value of x in the domain, then as we increase the value of x, the function $f(x)$ decreases. Figure 23 illustrates these two cases.

4.4 Algebra and Composition of Functions

When given a few definitions of functions there are multiple different types of operations that one can perform on functions. The laws described below breakdown the different types of operations that one can perform on functions:

Definition 40 *Let $f\colon A \to B$, $g\colon C \to D$, then the following is true:*
$(f+g)(x) = f(x) + g(x)$
$(f-g)(x) = f(x) - g(x)$
$(f \cdot g)(x) = f(x) \cdot g(x)$
$(\frac{f}{g})(x) = \frac{f(x)}{g(x)}$ *where* $g(x) \neq 0$

Another interesting operation that can be performed on functions is that of compositions. Compositions are very useful when it comes to calculus. The basic idea of compositions is that we first map an element from the domain of a set to the codomain. We then have a second function which maps that element to the codomain of the second function. We can see this visually in Figure 24

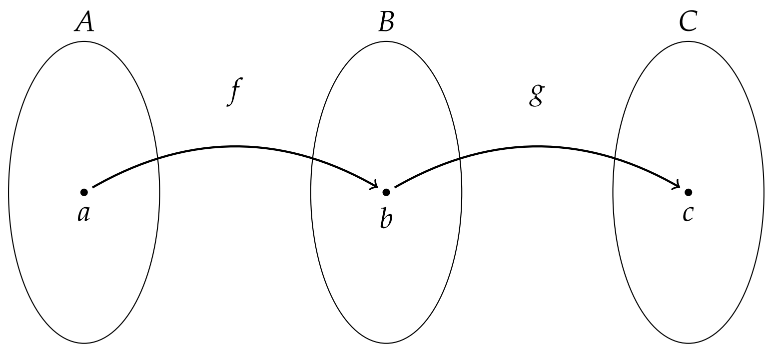


Figure 24: Function Composition Diagram.

Definition 41 *Let $f \colon A \to B$, $g \colon B \to C$. Then the composition function $h \colon A \to C$ is defined as:*
$h(x) = (f \circ g)(x) = f(g(x))$

Example 26 *When given functions, we can find the formulas for the composition of functions by replacing every occurrence of the input in one equation with the equation of the input function. Here are a few examples:*
$f(x) = x^2 + 2 \cdot x - 1$, $g(x) = e^x$.
$(f \circ g)(x) = f(g(x)) = f(e^x) = (e^x)^2 + 2(e^x) - 1 = e^{2x} + 2e^x - 1$
$(g \circ f)(x) = g(f(x)) = g(x^2 + 2x - 1) = e^{x^2+2x-1}$

$f(x) = \sqrt{x-1}$, $g(x) = ln(x) + x$.
$(f \circ g)(x) = f(g(x)) = f(ln(x) + x) = \sqrt{(ln(x) + x) - 1}$
$(g \circ f)(x) = g(f(x)) = g(\sqrt{x-1}) = ln((\sqrt{x-1})) + (\sqrt{x-1})$

In addition to combining functions, we can also simplify functions by observing that some functions are just compositions of others. That is, we can take a complicated equation and break it into different functions.

Example 27 *Let $f(x) = 4e^{2x+1} + 1$. We can break this function into three distinct functions:*
$f_1(x) = 4x$
$f_2(x) = e^x$
$f_3(x) = 2x + 1$

Hence, we can write $f(x)$ as:
$f(x) = 4e^{2x+1} + 1 = f_1(e^{2x+1}) + 1 = f_1(f_2(f_3(x))) + 1$

Being able to break these functions into smaller ones will help when we perform calculus calculations.

4.5 One-To-One and Onto Functions

Another set of interesting and important properties of functions are two concepts known as *surjectiveness* and *injectivness*. These two properties help us better understand certain behaviors of the functions themselves.

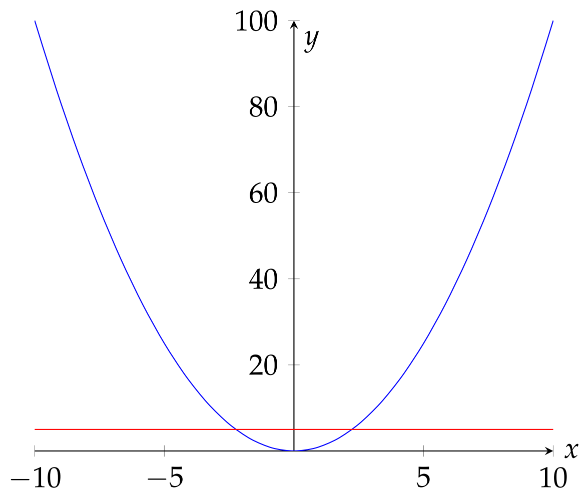


Figure 25: Horizontal Line Test for $f(x) = x^2$.

Definition 42 *A function* $f\colon A \to B$ *is said to be* injective *(or* one-to-one*) if* $x_1, x_2 \in A$ *and* $x_1 \neq x_2$ *then* $f(x_1) \neq f(x_2)$*. That is, given two numbers in the domain, the function of those two numbers will produce two different numbers.*

Definition 43 *A function* $f\colon A \to B$ *is said to be* surjective *(or* onto*) if* $\forall y \in B, \exists x \in A$ *such that* $f(x) = y$*. That is, for every element in the codomain, there must exist an element in the domain such that after passing that element through the function you get the original element in the codomain under observation.*

Example 28 *A function is one-to-one if the graph of the function passes the horizontal line test. This test is conducted by drawing a horizontal line and "moving" it downward and upward. If there is a point on the graph that intersects with the horizontal line twice, then the function is not one-to-one. For example, the function* $f(x) = x^2$ *has two points that intersect with a horizontal line. This means that two different x values will produce the same y-value, which means this is not a one-to-one function.*

4.6 Inverse Functions

If a function is one-to-one and onto, then we can typically find something known as an *inverse function* when given a function.

Definition 44 *Let* $f\colon A \to B$ *be a function that is one-to-one and onto. Then the* inverse function *of* f *is the function* $f^{-1}(x)$ *and has the property that* $f(f^{-1}(x)) = f^{-1}(f(x)) = x$*.*

Example 29 *Let* $f\colon [0, \infty) \to [0, \infty)$ *be defined as* $f(x) = x^2$*. This function is one to one and onto (the domain is only positive real numbers, not the whole number line, which is why it is one to one, differing from the previous example where the domain was all real numbers). We have as it's inverse function* $f^{-1}(x) = \sqrt{(x)}$*. We see that number in this domain are uniquely mapped to numbers in the range. So,* $\sqrt{x^2} = x = (\sqrt{x})^2$

5 Motivating Calculus

If you have spent a great deal of time attempting to understand the fundamentals of Set Theory and Pre-Calculus, then you are ready to understand Calculus. As we mentioned in the introduction, calculus is **not hard**! Don't let your friends scare you. Yes, Calculus does have some scary symbols sometimes. However, do not let this dissuade you. Calculus, all of calculus, breaks down very simplistically into one of three problems, the latter two of which stem from the first: limits, rates of change, and areas. We will motivate our need for Calculus in Business by looking at a few examples.

5.1 Motivating Limits

We will motivate the concept of a *limit* by observing a common problem in Business, namely the concept of *net present value*. Generally in our economy, the value of our dollar is always changing. Nominally, what \$1000 would buy us today will not buy us the same amount tomorrow. Day to day, these differences in value are not too noticeable. However, over a long period of time, the changes in value of our dollar affect every aspect of our lives.

Let us consider for the moment that a business would like to keep it's \$1,000 in cash in an account that does not earn any interest. At first glance, we may think this is a smart idea. If we are not spending the cash, then how can we possibly "loose it"? This is equivalent to stuffing the \$1,000 under a mattress. Suppose the business keeps it "under the mattress" for 10 years. The problem with doing this, of course, is that the value of the \$1,000 will most certainly change over the course of the 10 year time period. That is, while we technically still have the full \$1,000, we would be able to purchase much less with it in 10 years from now than we are today.

Therefore, what the business was able to purchase today is no longer able to purchase in 10 years from now with the same \$1,000. Ideally, we want to understand just how much "less" we can buy in 10 years from now by computing what the worth of that amount would be in terms of *today's money*. That is, while we will still have the \$1,000, can we relate it's worth in 10 years to an actual dollar amount in today's current environment? Also, why is it that we can't purchase as much with the same amount of money in 10 years than today? It's not like anyone took money away from us. Well, technically, you did "loose" money, not so much because someone "took away" any of your \$1,000, but rather, by changing how much that \$1,000 can purchase.

As a result of many factors, the cost of goods and services has a natural tendency to rise year to year. For example, rent may cost us only \$1,000 this year, but our landlord may raise this to \$1,100 for next year. The landlord needs to raise this rent not because he is greedy, but because his energy, food, and other costs in his life have changed as well. To make up for the change, he needs to pass this down to you. This cycle of course continues to spiral throughout the year for every year, which is what in turn raises the cost of goods and services, among other factors. Typically, we have observed in the past that the cost of goods and services rise on average by about 2% every year.

This means that your cash on hand is now worth 2% less than the year before. What cost you \$1,000 this year will cost you \$1,020 next year. This means that your \$1,000 is now worth \$20 less. In other words, your \$1,000 next year will only be worth the equivalent of \$980 this year. This can easily be computed by noticing the following: $\$980 = \$1000 - \$20 = \$1000 - (0.02)(\$1000) = \$1000(1 - (0.02))$. Now suppose you let the money sit another year. Given that another year has passed, this means that your original \$1000 devalues once again by another 2%. Starting in the end of the first year that passed, you had the equivalent of \$980 in terms of today's money.

If another year passes, then it is this amount that devalues by 2%. Following a similar logic as before, the \$1,000 in two years would be worth in today's money the equivalent of \$980(1-0.02). Notice that we have an expression above that we can substitute in for \$980: $980(1-.02) = (1000(1-0.02))(1-0.02) = 1000(1-0.02)^2$. Hence, we see a pattern emerge. If we would like to determine what our original money would be worth in t years but in terms of *today's money*, then we can easily find this by using the equation $1000(1-0.02)^t$.

In finance and economics, the percentage change in value per time frame (in our example, the time frame is a single year) is commonly referred to as a *rate*, and the starting amount of money is commonly referred to as a *principal*. When the rate is applied to the principal, then the amount is said to be *compounded*. Notice here that we only applied the rate once per time period, since we multiplied by it once per year. We can abstract out this problem by P_0 denote the initial amount of money, $P(t)$ denote the worth of P_0 in t years in terms of today's money, r denote the percentage rate of decline (which is equivalent to the percentage change in the cost of goods and services year to year), and t represent the number of time frames we are considering (in our example, the number of years that passed). As a result, we can figure out what our \$1,000 can buy us, value wise, in 10 years from now in terms of today's money. Our general equation for this is:

$$P(t) = P_0(1-r)^t$$

Answering our question from before, we can see that $P(10) = (1000)(1-0.02)^{10} = 1000(0.98)^{10} = 1000(0.81707) = 817.07$. To put this into perspective, this would mean that over 10 years, by letting out \$1,000 sit around not doing anything has effectively left us with \$817.07 in terms of today's money. In other words, our \$1,000 in 10 years will "feel like" we have only \$817.07 today. That's a $\frac{817.07-1000}{1000} = \frac{-182.93}{1000} = -0.18293$ percentage change. That is, we lost 18.293% of our initial money over 10 years by just letting it "sit safely under the mattress".

However, we only applied the full interest a single time every year. The practical interpretation of this would be that the cost of goods and services changes only once per year. Obviously, this is not realistic. The reason is that the cost of goods and services, in actuality, changes every day. Assume that the cost of goods and services changes a total of m times throughout the year by a rate of $\frac{r}{m}$. We divide by m since the full rate r is not actually applied m times a year. For example, we are not going to see daily fluctuations in the price of milk or rent by 2%. To simplify the rate that the cost of goods and services change by, we will just simply assume that the yearly percentage rate is applied in a *uniform* manner every time it is applied throughout the year. If this is the case, then on average, if our cost of goods and services change m times throughout the year, then the average percentage rate it changes by would be $\frac{r}{m}$.

Using our formula above, we can see that in a single year, our principle has been compounded a total of m times at a rate of $\frac{r}{m}$. Hence, after one year, we would have:

$$P(t) = P_0\left(1-\frac{r}{m}\right)^m$$

What if, however, we wanted to compound this over a period of t years? Since we know that interest will be applied a total of m times in a single year, and we are doing this compounding for t years, then the total number of times it is compounded would be mt. Therefore, if we compound m times per year at a rate of $\frac{r}{m}$, for t years, the value of our principle in t years in terms of today's money will be:

$$P(t) = P_0 \left(1 - \frac{r}{m}\right)^{mt}$$

Going back to our example from before, our initial \$1000, compounded m times per year at a rate of $\frac{0.02}{m}$ for 10 years will be worth in today's money $P(10) = (1000)(1 - \frac{0.02}{m})^{10m}$. Now practically, the cost of goods and services could be argued to change by even the second of our day to day lives. This can be argued to be the case due to second by second trading of commodities such as crude oil, coffee, barley, etc. These prices change by the second, and so, an argument could be made that the interest of $\frac{r}{m}$ is applied many many times throughout the year.

Let us select a few values of m and see what the worth would be under the assumption that we apply the interest these different number of times. Suppose we apply it only once per year, once per month, once per week, once per day, once per hour, and once per minute. Under these different number of times we compound, we compound more and more within a single year. For example, if we are compounding once per month for the entire year, then we compound it 12 times. If we compound it once per day, then we are compounding a total of 365 times in a year, etc. So, under the assumption that we compound many times within the year for different number of times, we observe the following:

m	$(1000)(1 - \frac{0.02}{m})^{10m}$
1.00	817.07
12.00	818.59
52.00	818.70
365.00	818.73
8760.00	818.73
525600.00	818.73

What if we compound an *infinite* number of times within a year? As we can see from the table, that as we increase the number of times we compound within a single year, we notice that the worth of our 1000 is *converging* to a single number. What do we mean by "converging"? It means that as we move the value of m closer to ∞, the values of the corresponding function $(1000)(1 - \frac{0.02}{m})^{10m}$ appear to get really close to a value of 818.73. The value of the function *at infinity* is what we call the *limit* of this function as m approaches infinity. When we move values of the independent variable x in a function (in our case, we can consider the variable m to be the independent variable in our example function) towards another value a (in our case, ∞), then if the values of $f(x)$ get closer, but not necessarily touch, another value L, then we write

$$\lim_{x \to a} f(x) = L$$

In our case, our function was $(1000)(1 - \frac{0.02}{m})^{10m}$ and our independent variable m was approaching infinity. We can see the resulting values of the function get very close to the value of 818.73. So, we would write:

$$\lim_{m \to \infty} (1000)(1 - \frac{0.02}{m})^{10m} = 818.73$$

Understanding what happens to a mathematical function as we "move" the value of an independent variable closer to value (where the value is any real number, infinity, or negative infinity) is the fundamental problem of *limits*. Limits serve as the bedrock for all of Calculus. We will see in the next lecture how to compute limits algebraically without having to list a few numbers like we have done above.

5.2 Motivating Rates of Change

The second fundamental problem in calculus that we concern ourselves over is the problem of *rate of change*. Rate of change is an incredibly important topic to understand for business applications, especially in the sub-branch of mathematics known as *optimization*. Real-world business decisions rest on the idea of minimizing or maximizing some objective. For example, a business may seek as part of it's strategy to minimize cost as much as possible. Others may seeks to minimize risk as a primary objective. Alternatively, others may wish seek to maximize their profit, revenue, or even *service levels*. In business applications, regardless of what the business seeks to accomplish, it is trying to do so by making decisions that will result in the lowest or highest possible values for their objectives.

At times, however, minimizing and maximizing may be difficult to accomplish. The reason is that our mathematical models of economics and business generally need to account for many complicated factors in the real world. If we model our real world business problem too simplistically, then we are ignoring important factors in the real world that directly impact our objectives. If we model our real world business problem too complex, then solving for the "best" decision becomes too complicated, computationally budrensome, and in some cases, infeasible. Therefore, modeling business problems is not only a science, but an art, that rests upon highly subjective observations from analyst.

Suppose we would like to figure out how many units to manufacture. If we manufacture too many units, we may not be able to sell all of our units. If we manufacture too little units, then we miss the potential to sell more, and we are left with a "cost" known as *opportunity cost*, which is the cost of not selecting a different decision. If a firm produces Q units for which it sells at a price of p and it costs them c per unit to manufacture, then we know that the firm's total profit is $\pi(Q) = pQ - cQ$ (in business math, the symbol π represents "profit", and it is **not** to be understood to be the typical interpretation of being the ratio of the circumference to the diameter of a circle).

In the real world, firms cannot just set the price p to anything it wants to. The laws of Economics typically state that the total demand for a product rests upon the price of that product. Every consumer has a certain *willingness to pay*. That is, if the firm sets the price p too high, then it can expect to see less demand for the product, since there will be a certain number of consumers that can no longer afford, or be willing to spend, money on the product for that price. Therefore, the total market demand for a product and the price p of the product are related to each other. This relation between Q (total market demand) and price p is often referred to as a *demand curve*.

Every product and market has a different demand curve that can be modelled mathematically with different equations. These equations rest on many factors like whether or not there is only one firm that produces the product (this is called a *non-competitive market*), many firms that make the produce (*competitive markets*), the level of brand loyalty, income of the consumer, cost of goods and services, product attributes, and many more.

For our purposes, we will make a *simplifying assumption* regarding the relation between the

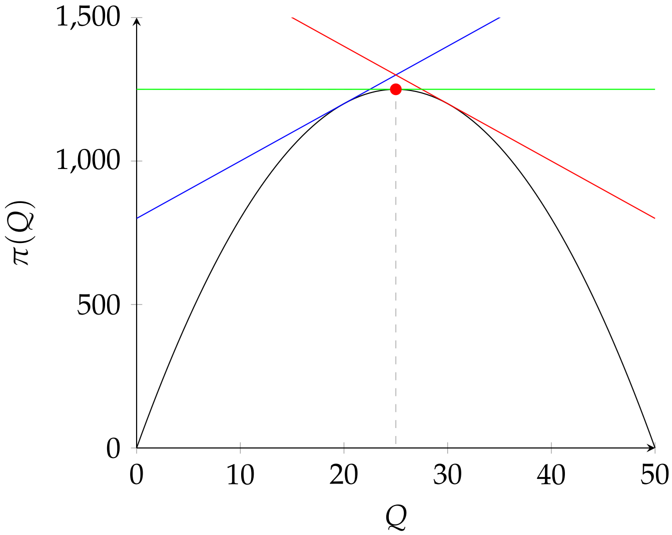


Figure 26: The profit function.

price and total market demand of our product in focus. A simplifying assumption is a statement about our model that makes the real world "more simple". Here, we will assume that there is only one firm (a monopoly) that makes this product, brand loyalty and product attributes are negligible, and that price and total quantity that the market demands are related by the following equation: $p = 1000 - 2Q$. Furthermore, suppose that the firm will produce only what the market demands. That is, we are assuming that the market is operating at *equilibrium*, namely, that the supply is equal to the demand. If we assume that the firm spends $c = \$900$ to manufacture each unit, then the firm's profit function for manufacturing Q units, selling at a price of $p = 1000 - 2Q$ would be:

$$\begin{aligned} \pi(Q) &= pQ - cQ \\ &= (1000 - 2Q)Q - 900Q \\ &= -2Q^2 + 100Q \end{aligned}$$

We can see this function plotted in Figure 26. Notice that if the firm produces too little, they will not earn as much profit. This is due to the fact that more consumers are willing to pay more for the product. However, if the firm produces too much, they will necessarily need to lower the price so as to match the supply and the demand. Therefore, there is a maximum level of profit that a firm can make. Now obviously, we can plot the function and see what the best number of units to produce is. Clearly, we see that this is 25, at a profit level of \$1250. This is the absolute most profit the firm can earn in the market (unless they manage to lower their variable costs).

However, in many applications in business, the equation to work with is not always obvious, and many times, we cannot even produce a plot of the profit function when given more than one decision. Therefore, we need to develop a set of analytical tools to be able to find the "best output" so as to achieve the highest level of possible profit. We typically do this by observing the *rate of change* of the function. Notice in the plot that if we were to pick a point to the left of 25 and draw a line that is tangent to the curve, we get a straight line with a *positive slope* (the blue line).

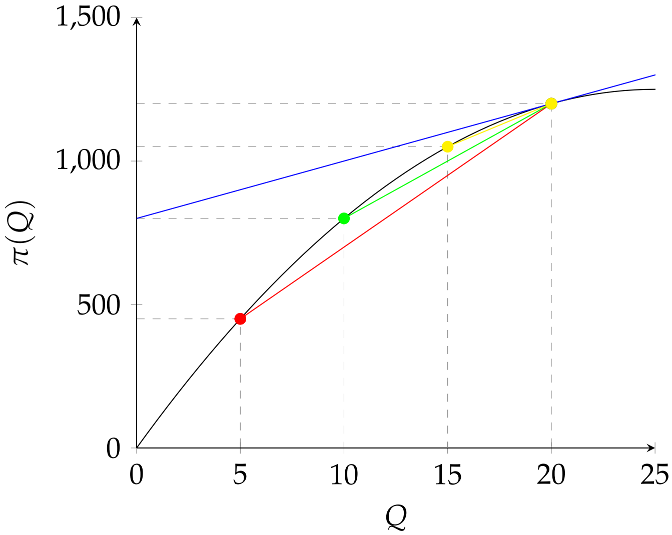


Figure 27: The profit function.

Recall that a straight line with a positive slope is a line that goes from down to up as we move horizontally from left to right. Also notice that if we were to pick a point on the right side of 25 and draw a tangent line to the curve, we would notice that we get a line with a *negative slope* (the red line), that is, the line is moving from up to down as we move horizontally from left to right. Intuitively, we should then expect that the slope of line that is tangent to the curve at the point 25 should be 0,a flat horizontal line (the green line).

Therefore, if we have a mechanism to be able to find the slope of the tangent line to a given curve $f(x)$ at a point x, then we would be able to use this mechanism to find optimal values of points. How can we find the slope of such a tangent line at a point to a curve $f(x)$? We can do so by *approximating* the slope of the line that intersects two points $f(x_1)$ and $f(x_2)$. This line is often referred to as the secant line. Suppose we would like to find the slope of the tangent line at $Q = 20$ (the slope of the blue line in Figure 27). Let us pick another point to the left of this, say $Q = 5$, and incrementally "move" Q up closer to the value of $Q = 20$ by every 5. So we will consider the points $Q = 5, Q = 10, Q = 15$. For each value of $Q_i, Q_i \in \{5, 10, 15\}$, we will compute the slope of the line that passes through the points $(Q_i, \pi(Q_i))$ and $(20, \pi(20)) = (20, 1200)$. In other words, we want to find the slope of the red line, green line, and yellow line in Figure 27.

Recall from algebra that you can find the slope of a line between two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ by finding the ratio of difference between the y-values and x-values (recall "rise over run"). That is, in order to find the slope of the line between points $(Q_i, \pi(Q_i))$ and $(20, 1200)$, we find the y-difference: $\Delta y = y_2 - y_1 = 1200 - \pi(Q_i)$ and the x-difference: $\Delta x = x_2 - x_1 = 20 - Q_i$. So the slope between these two points would be $m_i = \frac{\Delta y}{\Delta x} = \frac{1200 - \pi(Q_i)}{20 - Q_i}$. We can see in the table below that as we "move" the value of Q closer to the value of $Q = 20$, the resulting slopes m_i appear to be *converging* to a number, namely 20:

Q_i	$m_i = \frac{1200-\pi(Q_i)}{20-Q_i}$
5.00	50.00
10.00	40.00
15.00	30.00
16.00	28.00
17.00	26.00
18.00	24.00
19.00	22.00
19.90	20.20
19.99	20.02

Does this methodology look familiar? Recall in the previous section we discussed the idea of *limits*. That is, the concept of taking the value of an independent variable as close as possible to another value and seeing where the value of the corresponding $f(x)$ "converges" (if it converges). In this instance, our "function" is actually the expression $\frac{1200-\pi(Q_i)}{20-Q_i}$ and the independent variable is Q_i. We are taking Q_i and moving it "towards" the value of 20. Hence, we notice that the slope of the line tangent to a curve $f(x)$ can be found using the process describe above by finding:

$$\lim_{x \to a} \frac{f(a) - f(x)}{a - x}$$

Recall that $\Delta x = a - x$. We can hence rewrite this as: $a = \Delta x + x$. We also notice that as $x \to a$, this is the same as write $x \to \Delta x + x$, and the only way this happens is if $\Delta x \to 0$. So an equivalent way of writing the limit above would be:

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

This number, as we have motivated in our derivations, represents the slope of the line tangent to the function $f(x)$ at the point x. This is called the *derivative* of the function $f(x)$ at the point x, and we can represent this as:

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

Going back to our original problem, we can find the derivative of our profit function at $Q = 25$ by successively approximating the slopes, and "seeing" where the limit converges to. This is illustrated in the table below:

Q_i	$m_i = \frac{120=50-\pi(Q_i)}{25-Q_i}$
5.00	40.00
10.00	30.00
15.00	20.00
20.00	10.00
21.00	8.00
22.00	6.00
23.00	4.00
24.00	2.00
24.90	0.20
24.99	0.02

As we can see, as we bring the value of Q_i closer and closer to the value of $Q = 25$, we can see that the slope of the secant line gets closer and closer to the value of 0. This would imply that $f'(25) = 0$. In other words, the slope of the line tangent to our profit function at $Q = 25$ is equal to 0. This concept is the bedrock of all optimization theory. Generalizing this idea, we state without proof, that if we have a function $f(x)$ that represents the outcome of making a decision x, and if x^* represents the "best" decision that either maximizes or minimizes the objective $f(x)$, then it must be true that $f'(x^*) = 0$. This gives us a way to analytically *optimize* functions that involve decision making. The way we optimize decisions is by finding an equation for the *derivative*, or, the slope of a line tangent to a curve at a given value of x, set it equal to 0, and solve for x. Doing so will (possibly) give us values for the "best" decision. We will explore this idea further in a later lecture when we discuss derivatives in more detail.

5.3 Motivating Areas

Recall our discussion on taking summations over *uncountable sets*. Recall that the example we took was to "add" the number 1 $\frac{1}{n}$ of the time for a total of n times. We did this so that we can "add" the number 1 an uncountably infinite number of times. There are many practical examples in business that exploit this idea. Take for example inventory levels over time, where time can be modeled using an uncountable set such as an interval, like $[0, \infty)$, or even the entire real number line $\mathbb{R}$. In these instances, our basic notions of addition and multiplication go "out the window", as we illustrated earlier. The reason is simple. If we cannot "count", then we cannot add, and if we cannot add, then we cannot multiply. This is problematic. However, there is an answer to this problem, and it lies at the heart of the reason for why Calculus was constructed (i.e. limits).

Suppose we would like to figure out how much total inventory cost we spent over a period of time, where time runs in the uncountable set $T = [0, 250]$ (suppose for example sake that our unit of time is days). Assuming we have *continuous production*, that is, we continuously have inventory adding up in our facility, we can model the inventory level at time $x \in T$ as $f(x) = .005x^2$ That is, $f(x)$ represents the inventory level at time x, and we want to know how much total inventory we held from time 0 to time 250. The interval is $T = [0, 250]$ is uncountable since we can find a bijection between this and $\mathbb{R}$ (I'll leave that as an exercise for you! Hint, find a bijection between any interval [a,b] and (0,1). The function $f : (0, 1) \to \mathbb{R}$, $f(x) = ln\left(\frac{1}{x} - 1\right)$ is a bijection between (0,1) and $\mathbb{R}$, and you'll have your proof). Since time is not countable, we cannot take "all" the numbers, pass them through the function, and "add" like we were able to in the finite and countable cases. Instead, we need to resort to *approximation*. That is, the function for inventory level is bijective on the domain $[0, 250]$, which means that given any two points in time, the inventory level is different. This is problematic, since we cannot easily find the total inventory held over a time frame. If I told you that you held 5 units over a day, then you know that at half of the day, you still have 5 units of inventory. And so the total inventory held over 1 days would be $(5)(1)$. However, for our case, the inventory is not 5 over the entire day. In fact, it is *always changing*. This means that we cannot simply multiply (5)(1) to find the total inventory held over the day. In fact, the total inventory held over the day would actually be more (or less, depending on how you approximate it) than 5. This, of course, is problematic, and all intuition goes out the window!

Suppose we split the time interval into 10 equal sized intervals. We would then have $[0, 250] = [0, 25] \cup [25, 50] \cup [50, 75] \cup [75, 100] \cup [100, 125] \cup [125, 150] \cup [150, 175] \cup [175, 200] \cup [200, 225] \cup [225, 250]$. Now, select points from each interval, let's go with the points of the right of the intervals: 25,50,75,100,125,150,175,200,225,250. The goal is to compute the total inventory held during that time. Obviously, we cannot do that with the uncounable set, with the exception of the inventory level function being a constant. For example, suppose I told you that the inventory held over the time $[0, 25]$ was the same, say 1.25 held for time 0 to 25. Since the length of the interval is 25 (25 - 0), then the total inventory held for all time points between $[0, 25]$ would just simply be the multiplication of the two: (25)(1.25) = 31.25. In other words, suppose I wanted to know how much inventory in total was held in this time frame. 25 minutes passed, and we held 1.25 units per minute. Then we obtain our multiplication to obtain the total inventory held. This is a very practical type of situation, since we often need to compute inventory holding costs, which may be charged by the minute! And, this approximation easy to compute! Notice that this is also the area of a rectangle with width 25 and height 1.25. So "area" in this instance represents total inventory held within the 25 minutes.

The areas of rectangles represent "total inventory held" over each interval. So, if we were to add the areas of the rectangles, we would get an approximate value of the "total inventory" over all possible time values. So let's do that:

$$
\begin{aligned}
\text{Total Inventory} &\approx \sum_{i=1}^{10} 25 f(x_i) \\
&= 25 \sum_{i=1}^{10} f(x_i) \\
&= 25[f(x_1) + f(x_2) + f(x_3) + f(x_4) + f(x_5) + f(x_6) + f(x_7) + f(x_8) + f(x_9) + f(x_{10})] \\
&= 25[f(25) + f(50) + f(75) + f(100) + f(125) + f(150) + f(175) + f(200) + f(225) + f(250)] \\
&= 25[.005(25)^2 + .005 + (50)^2 + .005(75)^2 + .005(100)^2 + .005(125)^2 \\
&+ .005(150)^2 + .005(175)^2 + .005(200)^2 + .005(225)^2 + .005(250)^2] \\
&= 25(3.125 + 12.5 + 28.125 + 50 + 78.125 + 112.5 + 153.125 + 200 + 253.125 + 312.5) \\
&= 25(1203.125)
\end{aligned}
$$

However, this is only an approximation to the actual total, since as time immediately changes, so too does the inventory level. If we increase the number of rectangles (which subsequently decreases the width of each interval/rectangle), we may be able to get a better approximation. The table below shows the summations over different numbers of intervals:

Number of Intervals	Total Inventory
10	30078.12
100	26433.59
1000	26080.74
10000	26045.57
100000	26042.06
1000000	26041.71

As we can see, the more intervals we use, the summation eventually converges to a number. This also is shown visually in Figure 28, where the more intervals we use, the better the approximation of the area under the curve. If we notice, the difference between approximations is very small. If we were to take the number of intervals to infinity, then we obtain something that is called the *integral*, which would represent the area under our inventory level function (which in our case, the area under our curve from time 0 to time 250 represents the total inventory held over that time frame). We denote this area by:

$$\int_0^{250} f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x_n$$

This looks complicated, but do not fear it. Just keep in mind the intuitive interpretation of the symbol. The elongated "S" just simply means "take a summation using all values in an uncountable set". The bottom number indicates the starting value of the interval, the top number indicates the ending value of the interval. $f(x)$, in our case, represented the inventory level at time x. The symbol dx is just notation to specify the specific variable that we are "summing" over, which in our case is x, that is, the time. On the right hand side, the $\lim\limits_{n\to\infty}$ notation just says "set a value equal to n, plug it in the equation to the right, take a bigger value of n, do the same thing, keep doing this until we get to ∞. As a result of this "replace n with a value, compute, etc" procedure, we get a sequence of numbers that we hope will converge to a single number, where "converge" just means that the numbers at the end of the sequence will start to be very close to each other.

The Δx_n is specifying the length of the interval (rectangle) when we are approximating with n rectangles (intervals). If we assume the size of the interval is the same for all n intervals from 0 to 250, then the width of each interval would be $\Delta x_n = \frac{250-0}{n}$. The x_i, in our derivation, just means to set the value of x_i to the end of the interval when there are n intervals. Earlier we had $x_1 = 25, x_2 = 50, etc$. If we had 100 intervals, we would have $x_1 = 2.5, x_2 = 5, x_3 = 7.5, etc$, since these are the end points of the 100 intervals $[0, 2.5], [2.5, 5.0], [5.0, 7.5], \ldots, [245, 247.5], [247.5, 250]$.

And so, we have motivated our three fundamental problems in Calculus. In practice, when we can interpret a quantity as an area over an uncountable set, you better bet that we will be using integrals! Again, don't be afraid of the scary looking symbols. Just try to grasp the underlying meaning here. Everything in calculus, everything, is about a limit. In addition (no pun intended), we typically have very practical interpretations of rates of change and areas under functions. All of Calculus is about finding limits, finding rates of change (i.e. derivatives or slopes at a single point), and finding areas under curves. It really is THAT SIMPLE!

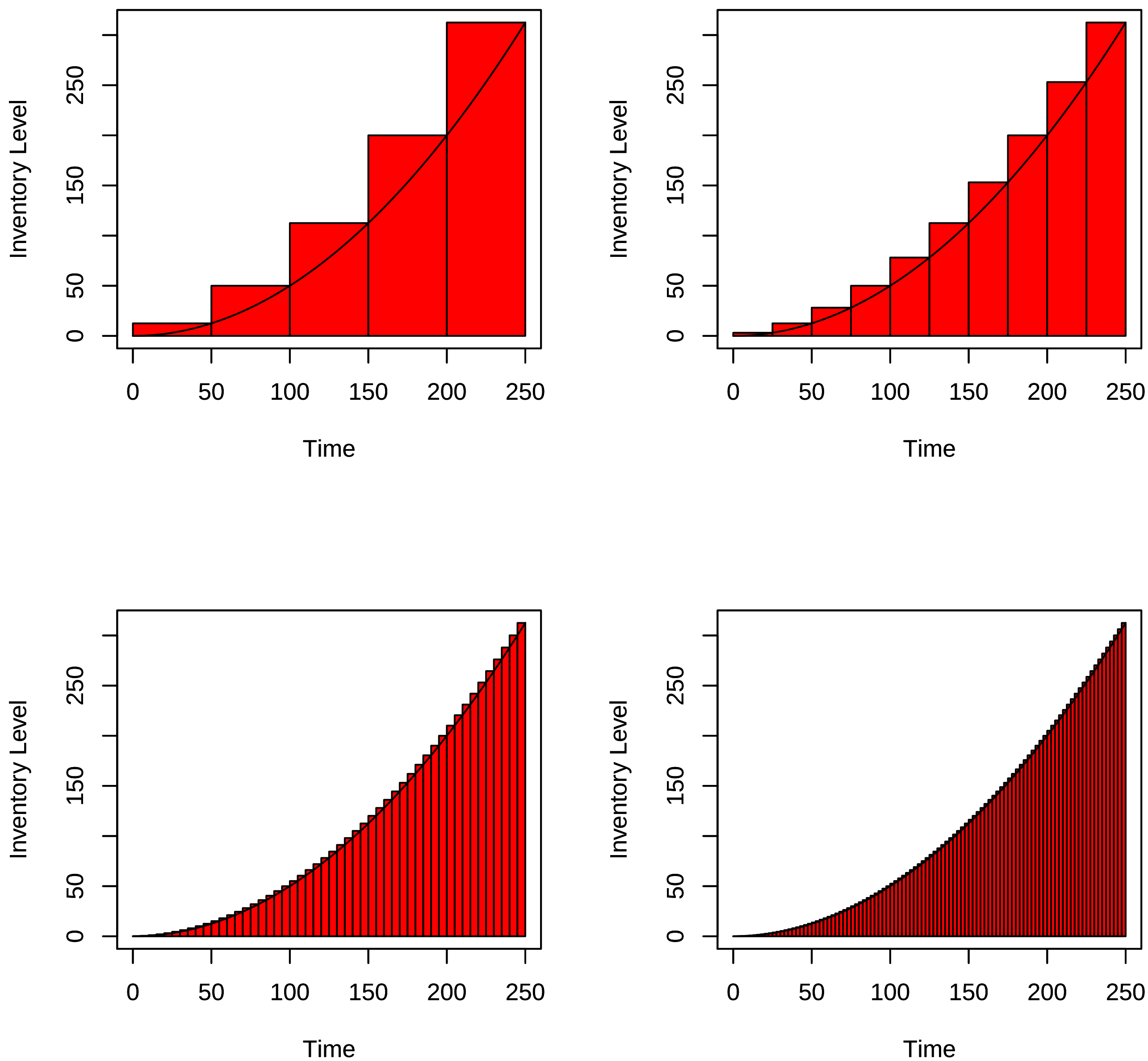


Figure 28: Four plots. The top right is an approximation of the total inventory using only 5 intervals. The top right is an approximation using 10 intervals. The bottom left is an approximation using 100 intervals. The bottom right is an approximation using 100 intervals