

Quantitative Methods (for Social Scientists)

Volume I: Calculus and Statistical Theory

Kevin Lingfeng Li

Table of contents

Preface	3
I Algebra and Pre-Calculus	4
1 Topics in Algebra	5
1.1 Set Theory	5
1.2 Solving Equations	6
1.3 Factoring and Expansion	7
1.4 Exponents and Logarithms	8
1.5 Summation and Product Notation	9
2 Pre-Calculus Topics	10
2.1 Functions and the Cartesian Plane	10
2.2 Slope and Zeroes of Functions	11
2.3 Special Types of Functions	12
2.4 Limits	14
2.5 Continuity	15
II Single Variable Calculus	16
3 Derivatives	17
3.1 Definition of a Derivative	17
3.2 Rules of Differentiation	18
3.3 Derivative of an Inverse Function	20
3.4 Partial Derivatives	21
3.5 Application: Multiple Regression	21
4 Optimisation	23
4.1 Local Extrema and Stationary Points	23
4.2 Concavity and Inflection Points	24
4.3 Global Extrema	24
4.4 Optimisation	25
4.5 Constrained Optimisation	26
5 Integrals	27
5.1 Indefinite Integrals	27
5.2 Definite Integrals and Riemann Sums	28
5.3 Fundamental Theorem of Calculus	30
5.4 Integration by Substitution	30
5.5 Integration by Parts	31

III	Probability and Statistical Theory	33
6	Probability Theory	34
6.1	Rules of Probability	34
6.2	Counting	34
6.3	Independence and Conditional Probability	34
6.4	Law of Total Probability	34
6.5	Bayes' Theorem	34
7	Random Variables	35
7.1	Discrete Random Variables	35
7.2	Continuous Random Variables	35
7.3	Expectations	35
7.4	Moments	35
8	Distributions	36
8.1	Discrete Uniform and Bernoulli Distribution	36
8.2	Binomial and Poisson Distribution	36
8.3	Continuous Uniform and Exponential Distribution	36
8.4	Normal (Gaussian) Distribution	36
8.5	Normal Approximation of the Binomial Distribution	36
9	Multivariate Random Variables	37
9.1	Joint Probability Functions	37
9.2	Marginal Distributions	37
9.3	Continuous Multivariate Distributions	37
9.4	Conditional Distributions	37
9.5	Covariance and Correlation	37
9.6	Independent Random Variables	37
9.7	Sums and Products of Random Variables	37

Preface

This book is one of the 5 companion-guides in the **Quantitative Methods** sequence, which provides the mathematics and statistics content useful for the study of Political Economics. The Quantitative Methods sequence consists of five volumes:

1. [Volume I: Calculus and Statistical Theory](#) (this volume) contains essential mathematics for Political Economics. Covers topics in single variable calculus, probability theory, and statistical theory.
2. [Volume II: Linear Algebra and Multivariate Calculus](#) covers topics in linear algebra and calculus. Highly recommended, but not formally required for Political Economics.
3. [Volume V: Proof and Analysis](#) covers introductory proofs in mathematics. Useful, but not essential for the rest of the Political Economics.

The 1st volume is a pre-requisite for the main series on Political Economics, and covers the core techniques in calculus, probability, and statistics needed. The 2nd volume is more advanced topics that are useful, but not formally required. Volume 3 contains advanced topics that are not required, but can be useful and interesting to those who enjoy Political Economics

This volume is a collection of mathematical and statistical topics that I consider to be essential to understand prior to starting any serious instruction in Political Economics. The book assumes a solid understanding of high-school level algebra, and discusses topics in algebra, single variable calculus, and probability and statistical theory. Note that this book is not a full mathematics course of those topics - topics have been selected based on what is commonly used in Political Economics and other social sciences.

Part I

Algebra and Pre-Calculus

Chapter 1

Topics in Algebra

1.1 Set Theory

Introduction to Sets

Key Definition: Sets

A **set** is a collection of objects. The objects within a set are the **elements** of that set. We can define a set by either listing every element out, or through describing the properties of the elements of the set.

- For example, $A = \{1, 2, 3, 4, 5\}$, or, $A = \{n \in \mathbb{Z} : 1 \leq n \leq 5\}$ - or in other words, A is the set of values of n is in the set of all *integers* \mathbb{Z} , such that n is between 1 and 5 inclusive.

Here are some common notation for sets:

- If two sets A and B have the same elements, they are equal: $A = B$.
- If a is an element of set A , we use the notation: $a \in A$.
- If a is not an element of set A , we use the notation: $a \notin A$.
- If all elements of set A are also within set B , then it is considered a **subset**, and notated $A \subseteq B$.
- If $A \subseteq B$, but $A \neq B$, then A is a **proper subset** of B , notated: $A \subset B$.

Set Operations

There are a few common operations of sets:

- The **union** of sets A and B , notated $A \cup B$, is the set of elements that belong to either A or B . For example, if $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, then $A \cup B = \{1, 2, 3, 4, 5\}$.
- The **intersection** of sets, notated $A \cap B$, is the set of elements that belong to A and B . For example, if $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, then $A \cap B = \{3\}$.
- The **cardinality** of a set A is the number of elements within set A , and is notated $|A|$. For example, if $A = \{1, 2, 3\}$, then $|A| = 3$. An **empty set** is a set with a cardinality of 0, or in other words, with no elements. It is notated with \emptyset .

Types of Sets and Notation

A **infinite set** is one with infinitely many elements. Common examples of infinite sets include:

- Set of all natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$.
- Set of all integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$.
- Set of all real numbers \mathbb{R} , which is any number on the real number line.

The set of all real numbers \mathbb{R} is the most common used set. A subset of \mathbb{R} is called an **interval**, and can be notated with either brackets or parentheses as follows:

- Closed interval: $[a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$
- Open interval: $(a, b) = \{x \in \mathbb{R} | a < x < b\}$
- We can also mix closed and open intervals: $[a, b) = \{x \in \mathbb{R} | a \leq x < b\}$

In all these intervals, a and b are called **endpoints** of the interval. Any point inside of an interval that is not an endpoint is an **interior point**.

1.2 Solving Equations

Solving Equations with One Variable

To solve an equation with one variable, we look to isolate the variable on one side. We can do this by either *adding/subtracting* to both sides of the equation, or *multiplying/dividing* the equation on both sides by the same value.

Example: Solve for x in the equation: $\frac{x+3}{4} = 6$

1. Multiply both sides by 4: $4\left(\frac{x+3}{4}\right) = 6 \times 4$
2. Simplify to get: $x + 3 = 24$
3. Now, subtract 3 from both sides: $x + 3 - 3 = 24 - 3$
4. Thus, we get $x = 21$

Solving Inequalities

Solving inequalities is the same as solving equations (for both one variable, and more, as we will see later). There is only one difference: if we multiply both sides by a negative number, we flip the direction of the inequality sign.

Solving System of Equations

To solve a system of equations (such as 2 different equations with 2 variables x and y), we do the following procedure:

1. With one of the equations, solve for one of the variables (either x or y generally)
2. Substitute that equation that we just solved, into the other equation. This will make that equation only have one variable. Solve that equation for the remaining variable.
3. Then, plug the variable we have found, into either of the original equations, and find the value of the other variable.

The solution to a system of equations is also the point where the two lines meet on a graph.

Example: Solve the system of equations: $3x + 2y = 13$, $x - 2y = 3$.

1. First, solve one of the equations for one of the variables.
 - Let us take $x - 2y = 3$. We can solve for x to get: $x = 2y + 3$
2. Now, plug the equation we just got into the other equation
 - The other equation is $3x + 2y = 13$. Plug in the x we found: $3(2y + 3) + 2y = 13$
3. Now, solve that equation for the remaining variable:
 - $3(2y + 3) + 2y = 13 \implies 6y + 9 + 2y = 13 \implies 8y = 4 \implies y = \frac{1}{2}$
4. Now, plug y back into either equation to find x :
 - $x = 2y + 3$ that we solved for earlier. Plug in y to get $x = 2(\frac{1}{2}) + 3 = 4$

Thus, the solution to the system of equations is $x = 4, y = \frac{1}{2}$

1.3 Factoring and Expansion

Factorisation is the process of changing an equation from the form $ax^2 + bx + c$ to the form $(x + i)(x + j)$.

To do this, we need to:

1. Find 2 numbers i and j , that sum to b , and multiply to c
2. Once we have done that, simply put it into the form $(x + i)(x + j)$

Example: Factor $x^2 + 5x + 6$

1. We need to find two numbers i and j , that add to 5, and multiply to 6.
2. 2 and 3 are two numbers that add to 5, while multiplying to 6.
3. Now, put it into the factored form: $(x + 2)(x + 3)$

There is a special form of factoring: **difference of squares**.

- Essentially, any expression that takes the form of $a^2 - b^2$ can be rewritten as $(a + b)(a - b)$.
- For example: $x^2 - 9 = (x + 3)(x - 3)$.

Multiplication expansion undoes factoring. To do this, we must multiply each term of one section by each other term of the other section. This is more intuitive with an example.

Example: let us prove the difference of squares idea we saw earlier - multiply out $(x + 3)(x - 3)$:

$$\begin{aligned}(x + 3)(x - 3) &= (x \times x) + (x \times -3) + (3 \times x) + (3 \times -3) \\ &= x^2 - 3x + 3x - 9 \\ &= x^2 - 9\end{aligned}$$

So we can see, that indeed, difference of squares is true, and $x^2 - 9 = (x + 3)(x - 3)$.

1.4 Exponents and Logarithms

Exponents

Here are a few key properties of exponents that are important to note:

1. $a^x \times a^y = a^{x+y}$ and $a^x / a^y = a^{x-y}$
2. $a^{-x} = \frac{1}{a^x}$
3. $(a^x)^y = a^{xy}$
4. $a^0 = 1$
5. $(a \times b)^n = a^n \times b^n$ and $(a/b)^n = a^n / b^n$

Sometimes, you will have an equation that takes the form $x^n = a$, where n is some number.

How to get rid of n to isolate x ?

- We know that $n \times (1/n) = 1$.
- We also know that $(a^x)^y = a^{xy}$. Thus:

$$\begin{aligned}(x^n)^{(1/n)} &= a^{(1/n)} \\ x^{n \times (1/n)} &= x^1 = x = a^{(1/n)} \\ \text{thus, } x &= a^{1/n}\end{aligned}$$

Logarithms

Logarithms are the inverse for exponential functions. The relationship is as follows:

$$y = \log_a(x) \implies a^y = x$$

Or more intuitively, take the example $y = \log_2(8)$ is basically asking, what power y should I take base 2 to, in order to get an answer of 8.

- Of course, $2^3 = 8$, so the answer is $y = 3$.

Here are a few key properties to know about logarithms.

1. $\log(xy) = \log(x) + \log(y)$ and $\log(x/y) = \log(x) - \log(y)$
2. $\log(x^y) = y \times \log(x)$
3. $-\log(x) = \log(1/x)$
4. $\log(1) = \log(e^0) = 0$

Note that in statistics, when we say $\log(x)$, we are talking about natural logs (base e). We will specify any other base with a subscript in the logarithm. We can change the base of logarithms with the following formula:

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$$

1.5 Summation and Product Notation

Summation Notation

Key Definition: Summation Notation

Summation Notation is a way of expressing the sums of many values:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

- $i = 1$ indicates that we start summing from index $i = 1$, or basically, we start with x_1 .
- n is the value of the index at which we stop summing. This is often the sample size.
- x_i is the value of an element within set $X = \{x_1, x_2, \dots, x_n\}$ with index i . For example, is $i = 1$, $x_i = x_1$, the first element in set X .

Summation notation has a few properties:

1. $\sum cx_i = c \times \sum x_i$, where c is some constant.
2. $\sum (x_i + y_i) = \sum x_i + \sum y_i$
3. $\sum_{i=1}^n c = n \times c$, where c is some constant.

Product Notation

Key Definition: Product Notation

Product notation is a way of expressing the products of many values:

$$\prod_{i=1}^n x_i = x_1 \times x_2 \times \dots \times x_n$$

- $i = 1$ indicates that we start multiplying from index $i = 1$, or basically, we start with x_1 .
- n is the value of the index at which we stop multiplying. This is often the sample size.
- x_i is the value of an element within set $X = \{x_1, x_2, \dots, x_n\}$ with index i . For example, is $i = 1$, $x_i = x_1$, the first element in set X .

Product notation has the following properties:

1. $\prod cx_i = c^n \times \prod x_i$, where c is some constant.
2. $\prod_{i=1}^n c = c^n$, where c is some constant.

Converting Between Sum and Product Notation

We can convert between summation and product notation with the properties of logarithms::

$$\begin{aligned}\log\left(\prod_{i=1}^n cx_i\right) &= \log(cx_1 \times cx_2 \times \dots \times cx_n) \\ &= \log(c^n \times x_1 \times x_2 \times \dots \times x_n) \\ &= \log(c^n) + \log(x_1) + \log(x_2) + \dots + \log(x_n) \\ &= n \log(c) + \sum_{i=1}^n \log(x_i)\end{aligned}$$

Chapter 2

Pre-Calculus Topics

2.1 Functions and the Cartesian Plane

Functions

Key Definition: Functions

We have two sets, X and Y . A function $f : X \rightarrow Y$ is some rule, which assigns each element $x \in X$ to one, and only one element $y = f(x) \in Y$.

- For every input $x \in X$, we must only have one output $y \in Y$. If we have two outputs, that is not a function.

The **domain** of f is the set of possible values of set X . Basically, what values can we input into $f(x)$.

The set of possible outputs of function f is the **range** $R(f)$, and the range must be a subset of the set of Y : $R(f) \subseteq Y$.

Functions do not have to only have one input variable. For example, we can have a function $f(x, z)$, with two input values $x \in X$ and $z \in Z$, notated $f : X, Z \rightarrow Y$.

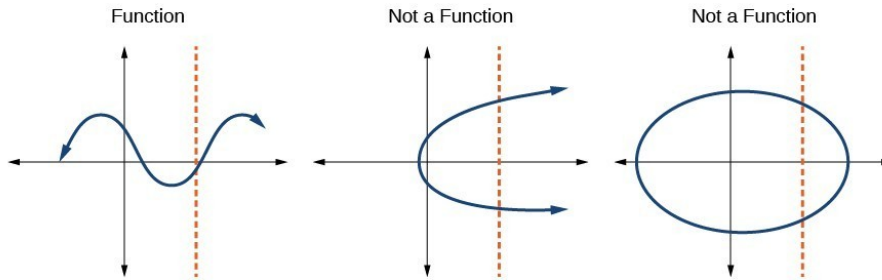
- However, we once again can only have one output $y \in Y$ for each different input (x, z) . If we have multiple outputs for one input, that is not a function.

Cartesian Plane

Remember that \mathbb{R} is the set of all real numbers $\mathbb{R} = (-\infty, \infty)$. That space is only one dimensional - a single number line. \mathbb{R}^2 is a 2 dimensional space - essentially two number lines, one perpendicular to the other. Essentially, a graph with x and y -axes. This 2-dimensional space is called a **cartesian plane**.

Every single point in the 2-dimensional cartesian plane can be specified with coordinates (a, b) , where both $a \in \mathbb{R}$ and $b \in \mathbb{R}$. a is the x coordinate, and b is the y coordinate - both which represent the distance from the **origin** with coordinates $(0, 0)$.

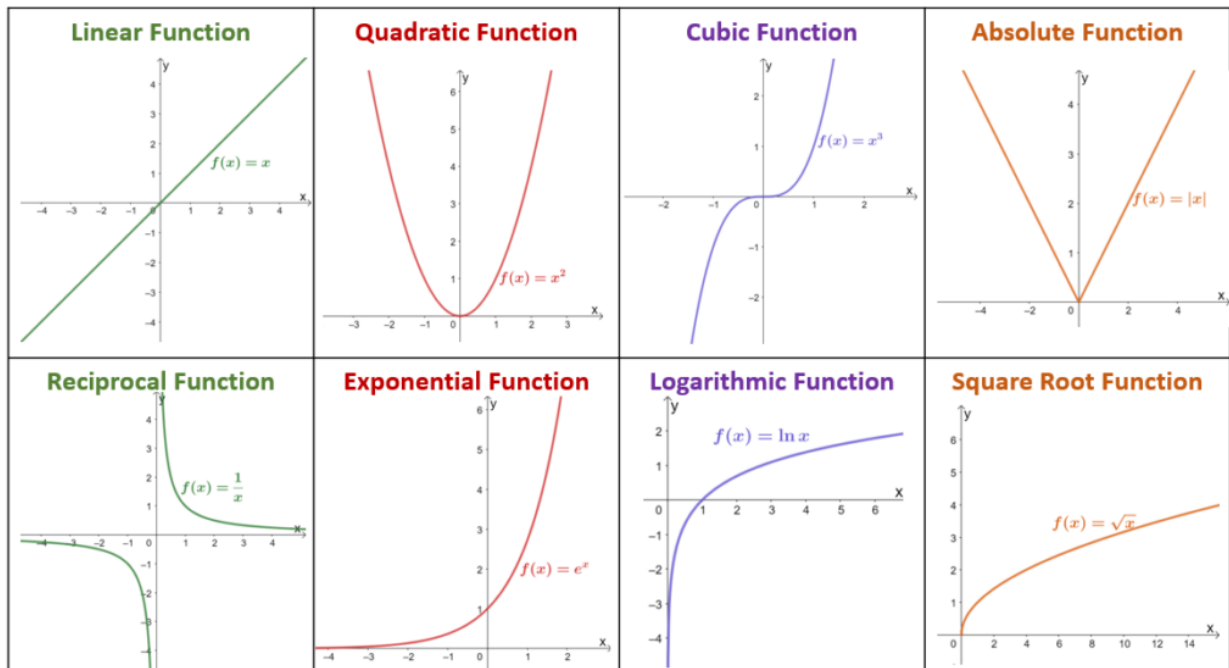
Any real-valued function of one variable can be represented in the \mathbb{R}^2 space. On the x -axis, we put the input number, and on the y -axis, we put the output of the function. For each input variable value, there must only be one output variable value. Graphically in \mathbb{R}^2 , we can show this:



Of course, functions can be in more than 2 dimensions. A function with two input variables $f(x, z)$, would be depicted in the 3-dimensional space \mathbb{R}^3 , as $x \in \mathbb{R}$, $z \in \mathbb{R}$, and output $f(x) \in \mathbb{R}$. The more input variables, the more dimensions the function is depicted in.

Common Types of Functions

There are a few very common types of functions that we will encounter:



2.2 Slope and Zeroes of Functions

Linear Equations and Slope

Linear equations take the form $f(x) = mx + b$, where m is the slope.

💡 Key Definition: Slope

More formally, slope is the change in y over the change in x , or the amount y changes, for every increase of one unit of x . We can calculate the slope between two points (x_1, y_1) and (x_2, y_2) as follows:

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

Zeroes and Roots of Functions

Key Definition: Zeroes and Roots

The zeroes/roots of a function are the input values of the function that make the output function equal to 0.

- Mathematically, for what x does $f(x) = 0$.
- Graphically. Since output $f(x)$ is graphed on the y -axis, we can see at what x -values the function cross the line $y = 0$ (the x -axis line).

To solve for zeroes/roots, simply set the function equal to 0, and solve for x .

- Some functions can have more than one zero, especially polynomials, and some can have none.

Example: Find the root of $f(x) = x - 5$

1. Set function output equal to 0: $0 = x - 5$
2. Solve for x and we get: $x = 5$.

For **quadratic functions**, there are two ways to find roots. We can factor, covered in section 1.2.

Example: Find the root of $f(x) = x^2 + 5x + 6$

1. Set function output equal to 0: $0 = x^2 + 5x + 6$
2. Factor: $0 = (x + 2)(x + 3)$
3. $x = -2$ and $x = -3$ make this equation true.

We can also use the **quadratic formula** to find the roots of unfactorable equations. Given a quadratic in the form of $ax^2 + bx + c$, the roots are:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Note: we will often have two roots in quadratics (although not always).

2.3 Special Types of Functions

Composite Functions

Composite functions are when the output of one function, is used as the input of another function. Composite functions are defined as the following:

$$f \circ g = f[g(x)]$$

Essentially, we calculate the output value of $g(x)$, and immediately input that value into $f(x)$.

For example, if $h(x) = x^2$, and $f(x) = x - 5$, then $f \circ h = (x^2) - 5$

- Note, $f \circ h$ and $h \circ f$ are different, so make sure you are doing the correct one!

Surjective, Injective, and Bijective Functions

A function $f : X \rightarrow Y$ is a **surjective** function if its range is equal to the set of Y . Mathematically, $R(f) = Y$.

- In other words, for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$.
- For example: $f(x) = x$ is surjective, as the range of $f(x) = x$ is $(-\infty, \infty)$, which covers the entire set of Y . Thus, $R(f) = Y$.

A function $f : X \rightarrow Y$ is an **injective** function if distinct inputs are mapped to distinct outputs. Basically, no two input values of $x \in X$ produce the same value output $f(x)$.

- Mathematically, for all $x_1 \in X, x_2 \in X$, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$
- For example: $f(x) = x$ is injective, because no two inputs of x produce the same output $f(x)$.

A function is **bijective** if it is both injective and surjective.

- $f(x) = x$, as we have shown previously is both injective and surjective, and thus bijective.

Inverse Function

Let $f : X \rightarrow Y$ be a bijective function. The sets X and Y are such that $X \subseteq \mathbb{R}$ and $Y \subseteq \mathbb{R}$. There is some unique function $f^{-1} : Y \rightarrow X$, called the **inverse function** of f , which is essentially the opposite of function f : when you input the outputs of $f(x)$ into f^{-1} , you get the respective input of $f(x)$.

- Or mathematically: $y = f(x)$ if and only if $x = f^{-1}(y)$

Because f^{-1} essentially reverses function $f(x)$, the following must be true:

$$\begin{aligned}f^{-1}(f(x)) &= x \text{ for all } x \in X \\f(f^{-1}(y)) &= y \text{ for all } y \in Y\end{aligned}$$

To find the inverse function, we simply solve for the input variable.

Example: Find the inverse of $f(x) = 2x + 3$

1. Let us define y as the output of $f(x)$, such that $f(x) = y = 2x + 3$
2. Now, solve for x : $y = 2x + 3 \implies 2x = y - 3 \implies x = \frac{y-3}{2}$

Let us prove that this is an inverse function with the properties we showed above.

$$\begin{aligned}\text{1st property: } f^{-1}(f(x)) &= x \text{ for all } x \in X \\f^{-1}(y) &= \frac{y-3}{2} \implies f^{-1}(f(x)) = \frac{(2x+3)-3}{2} = \frac{2x}{2} = x \\ \text{2nd property } f(f^{-1}(y)) &= y \text{ for all } y \in Y \\f(x) = 2x + 3 &\implies f(f^{-1}(y)) = 2\left(\frac{y-3}{2}\right) + 3 = y - 3 + 3 = y\end{aligned}$$

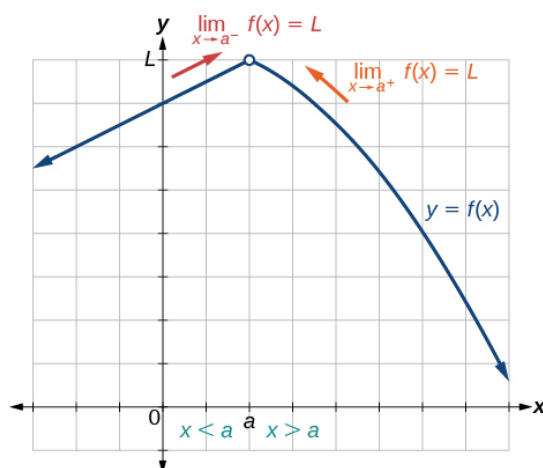
2.4 Limits

💡 Key Definition: Limits

A limit is when a function approaches some specific output value $y = f(x) = L$, as the input x moves closer and closer to some value c .

- Or more intuitively, as the x value becomes infinitely closer to $x = c$, the $y = f(x)$ value becomes infinitely close to $f(x) = L$.
- Limits are notated with the form: $\lim_{x \rightarrow c} f(x) = L$

For example, the figure below shows a function with a limit:



For a limit to exist, it must approach the same y value L , from both sides of $x = c$.

- For example, in the figure above, from both the left and right of $x = c$, the function $f(x)$ approaches L . Thus, a limit exists here.

However, if the function were to approach different y values on both sides of $x = c$, then there would be no limit.

- We can notate the limit from the left side of $x = c$ as $\lim_{x \rightarrow c^-} f(x) = L$, and the limit from the right side of $x = c$ as $\lim_{x \rightarrow c^+} f(x) = L$.

Importantly, $f(x)$ does not need to be defined at point $x = c$ for the limit to exist.

- For example, the figure above has a function $y = f(x)$ that is not defined at $f(c)$, however, the limit still exists.

Properties of Limits

Let us have functions f and g , where $\lim_{x \rightarrow c} f(x) = A$, and $\lim_{x \rightarrow c} g(x) = B$. The following properties are thus true:

1. $\lim_{x \rightarrow c} [f(x) + g(x)] = \lim_{x \rightarrow c} f(x) + \lim_{x \rightarrow c} g(x)$
2. $\lim_{x \rightarrow c} [\alpha \times f(x)] = \alpha \times \lim_{x \rightarrow c} f(x)$, where α is some constant.
3. $\lim_{x \rightarrow c} [f(x) \times g(x)] = \lim_{x \rightarrow c} f(x) \times \lim_{x \rightarrow c} g(x)$
4. $\lim_{x \rightarrow c} [f(x)/g(x)] = \lim_{x \rightarrow c} f(x) / \lim_{x \rightarrow c} g(x)$

2.5 Continuity

Continuity as a concept is quite intuitive: if a function has no breaks or jumps within an interval, then it is continuous within that interval.

We can express this intuition in more mathematical terms. Suppose that function f has some input interval containing $x = c$. f is continuous at point c , given these 2 criteria are both met:

1. $\lim_{x \rightarrow c} f(x)$ exists, i.e. $\lim_{x^- \rightarrow c} f(x) = \lim_{x^+ \rightarrow c} f(x)$.
2. $\lim_{x \rightarrow c} f(x) = f(c)$.

These rules make sense intuitively:

1. The limit must exist for continuity, since if a function approaches a different Y value on each side at a certain $x = c$, then it cannot be continuous. Thus, the first condition must be true.
2. If the function is not defined at c , or is defined by not equal to the limit, that means there will be a break or jump in the function. Thus, the second condition must be true for continuity.

Let us assume functions f and g are both continuous at $x = c$. Then, continuity has a few properties:

1. $f(x) + g(x)$ is continuous at $x = c$.
2. $f(x) - g(x)$ is continuous at $x = c$.
3. $f(x) \times g(x)$ is continuous at $x = c$.
4. $|f(x)|$ is continuous at $x = c$.
5. $\alpha \times f(x)$ is continuous at $x = c$, where α is some constant.
6. $f(x)/g(x)$ is continuous at $x = c$ given $g(c) \neq 0$.

Finally, if f is continuous in interval $[a, b]$, then f must have a maximum and minimum value within $[a, b]$.

Part II

Single Variable Calculus

Chapter 3

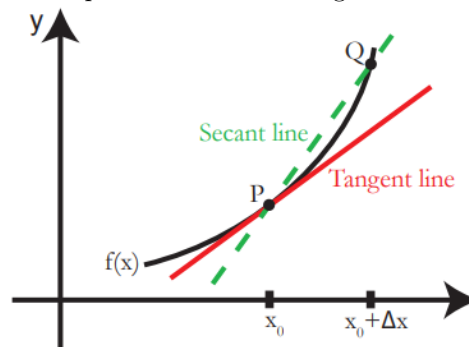
Derivatives

3.1 Definition of a Derivative

💡 Key Definition: Derivative Function

A derivative function f' outputs the slope (rate of change) of the function of f at any input. So, if we want to know the slope of f at $f(a)$, we would plug a into f' .

The figure shows a function f and a point P . The red tangent line shows the slope at point P :



We know the definition of slope between two points (from section 1.2) is:

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

First, since we are dealing with functions where output $y = f(x)$, let us replace the y 's in our equation with $f(x)$'s:

$$\Delta y / \Delta x = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

We can rewrite x_2 and x_1 relative to a single point of x as $x_1 = x$ and $x_2 = x + h$. Why? this is because point x_2 is just some x -distance away from point x_1 . Now, let us replace x_1 and x_2 with x and $x + h$:

$$\Delta y / \Delta x = \frac{f(x + h) - f(x)}{x + h - x} = \frac{f(x + h) - f(x)}{h}$$

So that is essentially another way to rewrite the definition of a slope between two points.

- However, the issue is, that a derivative is the slope at **one point**, not between two points. What we have calculated so far is the slope between two points (the secant green line in the figure above).
- We need the red-tangent line - the slope at one point. How do we calculate the slope at one point?

To calculate the slope at one point, we can slowly move the two points x and $x + h$ closer and closer to each other.

- How do we do that? With **limits** (that we covered in section 1.6). By making h smaller and smaller until it approaches 0, the two points will get closer and closer until they are on top of each other, and then, we will have the slope at one point.

💡 Key Definition: Formal Definition of a Derivative

Thus, below is the **formal definition of a derivative**. We can calculate any derivative with this rule (although, this can become difficult sometimes).

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

A note on notation: so far, I have used f' (apostrophes) to notate a derivative. We can also use the form $\frac{dy}{dx}$ to indicate a derivative. We can change y and x in this notation if we are using different variables.

Finally, it is also possible to take the derivative of a derivative. This is a second-order derivative, and we notate it f'' (with two apostrophes).

3.2 Rules of Differentiation

You can calculate any derivative by plugging it into the formal definition from above. However, this can be tedious and involve a lot of algebra.

However, there are a few generalised forms of functions, whose derivatives are consistent. These rules allow us to quickly calculate derivatives.

- Note: these rules are all obtained from the definition of the derivative - they do not magically appear from thin air.

Standard Derivatives

Linear Rule: $[\alpha x]' = \alpha$, where α is a constant.

Derivative of a constant: $[\alpha]' = 0$, where α is a constant.

Power Rule: $[x^k]' = kx^{k-1}$, where k is some constant and x^k is a monomial.

Sum Rule: $[f(x) \pm g(x)]' = f'(x) \pm g'(x)$.

Constant Rule: $[\alpha \times f(x)]' = \alpha \times f'(x)$, where α is some constant.

Product Rule: $[f(x) \times g(x)]' = f'(x) \times g(x) + f(x) \times g'(x)$.

Quotient Rule: $[f(x)/g(x)]' = \frac{f'(x) \times g(x) - f(x) \times g'(x)}{[g(x)]^2}$.

Chain Rule: $[f(g(x))]' = f'[g(x)] \times g'(x)$ (used for composite/nested functions).

Exponential and Log Derivatives

Exponential: $[e^x]' = e^x$.

Exponential Nested: $[e^{u(x)}]' = e^{u(x)} \times u'(x)$.

Other Based Exponentials: $[b^x]' = \ln(b) \times b^x$.

Natural Log Derivative: $[\ln(x)]' = \frac{1}{x}$.

Log derivative with a power: $[\ln(x^k)]' = \frac{k}{x}$.

Nested Log Derivative: $[\ln(u(x))]' = \frac{u'(x)}{u(x)}$

Trigonometric Derivatives

Derivative of sine: $[\sin(x)]' = \cos(x)$.

Derivative of cosine: $[\cos(x)]' = -\sin(x)$.

Example 1: Find the derivative of $f(x) = 5x^4 - 6x^3 + x^2 - 5x + 6$.

1. Sum rule says that $f'(x) = [5x^4]' - [6x^3]' + [x^2]' - [5x]' + [6]'$
2. Constant rule says that $f'(x) = 5[x^4]' - 6[x^3]' + [x^2]' - 5[x]' + [6]'$
3. Power rule says that: $f'(x) = 5(4x^3) - 6(3x^2) + (2x) - 5 + 0$
4. Simplify: $f'(x) = 20x^3 - 18x^2 + 2x - 5 + 0$

Example 2: Find the derivative of $f(x) = (x^3)(2x^4)$.

Two methods to do this problem. Method 1 (product rule):

1. Product rule says that $[f(x) \times g(x)]' = f'(x) \times g(x) + f(x) \times g'(x)$. Let us define $f(x) = x^3$ and $g(x) = 2x^4$.
2. Let us find $f'(x)$ and $g'(x)$. Using power rule: $f'(x) = 3x^2$, and $g'(x) = 2[x^4]' = 2(4x^3) = 8x^3$
3. Put it into the product rule form: $f'(x) = (3x^2)(2x^4) + (x^3)(8x^3)$.
4. Simplify: $f'(x) = 6x^6 + 8x^6 \Rightarrow f'(x) = 14x^6$.

Method 2 (power rule):

1. Simplify $f(x)$ to get $f(x) = 2x^7$.
2. Use constant rule and power rule: $f'(x) = 2[x^7]' = 2(7x^6) \Rightarrow f'(x) = 14x^6$

Example 3: Find the derivative of $f(x) = (3x^2 + 5x - 7)^6$

1. This is a composite function, so we use chain rule $[f(g(x))]' = f'[g(x)] \times g'(x)$. Let us define $f(x) = x^6$ and $g(x) = 3x^2 + 5x - 7$.
2. Let us find $f'(x)$ and $g'(x)$. Power rule says $f'(x) = 6x^5$. Sum, constant, and power rule says $g'(x) = 6x + 5$.
3. Put it into the form of chain rule: $f'(x) = 6(3x^2 + 5x - 7)^5 \times 6x + 5$.
4. Distribute out $6x + 5$ to get $f'(x) = 36x(3x^2 + 5x - 7)^5 + 30(3x^2 + 5x - 7)^5$.

3.3 Derivative of an Inverse Function

Let $f : X \rightarrow Y$ be a bijective function $y = f(x)$ (see sections 1.4-1.5). We want to find the derivative of the inverse function $f^{-1} : Y \rightarrow X$.

The simplest way to do this is to find the inverse function (as shown in section 1.5), and calculate the derivative of that.

- However, the function $x = f^{-1}(y)$ is not always easily differentiable with the rules we have introduced.
- We can utilise an alternative approach to find inverse derivatives.

Remember the identity of inverse functions shown in section 1.5:

$$f^{-1}(f(x)) = x \text{ for all } x \in X$$

We know that chain rule says:

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

We can define our inverse original f' function as f for the chain rule, and the original $f(x)$ as $g(x)$ for the chain rule.

Thus, using chain rule, we can get the derivative of the identity of inverse functions:

$$f^{-1}(f(x)) = x, \text{ now differentiate both sides:} \\ \frac{df^{-1}(f(x))}{dy} \times \frac{df(x)}{dx} = 1, \text{ since } y = f(x) \implies \frac{df^{-1}(y)}{dy} \times \frac{df(x)}{dx} = 1$$

Now, isolate the derivative of the inverse function to get the **derivative of the inverse**:

$$\frac{df^{-1}(y)}{dy} = \frac{1}{\frac{df(x)}{dx}}, \text{ and since } y = f(x), x = f(y), \implies \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

We can use this property to solve for the derivative of the inverse of a function that we do not know how to differentiate with our above rules.

Example: Find the derivative of the inverse function for $f(x) = \ln(x)$.

1. First, find the inverse $f^{-1}(y)$. We do this by solving for x in $y = \ln(x)$. Through the definition of logarithms, we know that $x = e^y$. Thus, $f^{-1}(y) = e^y$.
2. Using derivative rules, we know that $f'(x) = \frac{1}{x}$.
3. Using the inverse derivative relationship, we know $f'(y) = \frac{1}{f'(x)}$.
4. Thus, $f'(y) = \frac{1}{1/x} = x$.
5. Since we solved $x = e^y$, we can plug in to get: $f'(y) = e^y$.

3.4 Partial Derivatives

Suppose that we have a function $f : X_1, X_2, \dots, X_n \rightarrow \mathbb{R}$. Or in other words, a function with more than one variable input.

- However, we might only be interested in the rate of change (slope) relative to only one of the variables.
- A partial derivative is what we will need to find - the rate of change only in relation to one variable.

Key Definition: Partial Derivative

A partial derivative is defined as:

$$\frac{\partial}{\partial x_i} f(x_1, \dots, x_i, \dots, x_n) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

- Where $x_1, \dots, x_i, \dots, x_n$ are all the inputs of function f .
- Where we are only interested in the rate of change in respect to x_i (hence why only x_i has a $+h$ in the numerator).
- As a result, we treat x_1, \dots, x_n that are not x_i as simply constants.

To solve a partial derivative, we take the derivative in respect to our specific variable x_i , while treating all other input variables x_1, \dots, x_n like a constant.

- So essentially, we treat other variables like we would the number 5 - i.e. the derivative of a constant is 0, so the derivative of other variables is also 0.

Example: Find the partial derivative in respect to x for $f(x, z) = 5x^2z^3 + 2x + 3z + 5$

1. Use sum rule to split up derivative: $\frac{\partial f(x, z)}{\partial x} = \frac{\partial}{\partial x} [5x^2z^3] + \frac{\partial}{\partial x} [2x] + \frac{\partial}{\partial x} [3z] + \frac{\partial}{\partial x} [5]$.
2. Take out constants from first term (remember z is constant here) $\frac{\partial f(x, z)}{\partial x} = 5z^3 \frac{\partial}{\partial x} [x^2] + \frac{\partial}{\partial x} [2x] + \frac{\partial}{\partial x} [3z] + \frac{\partial}{\partial x} [5]$.
3. Find derivatives, remember z is constant here: $\frac{\partial f(x, z)}{\partial x} = 10z^3x + 2 + 0 + 0$.
4. Thus, $\frac{\partial f(x, z)}{\partial x} = 10z^3x + 2$.

3.5 Application: Multiple Regression

Partial derivatives are very useful in economic modelling and econometrics/statistics. We will introduce an application of partial derivatives with linear regression model.

The multiple linear regression model takes the following form (do not worry too much about this, this will be covered in the Econometrics sequence):

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$$

You probably have heard that multiple linear regression is powerful because it allows us to see the independent effects of explanatory variables on output variables (if you have not, do not worry).

Let us say we are interested in only the effect of X_1 on Y . We can take the partial derivative in respect to X_1 as follows:

$$\begin{aligned}\frac{\partial \hat{Y}}{\partial X_1} &= \frac{\partial \hat{Y}}{\partial X_1}[\hat{\alpha}] + \frac{\partial \hat{Y}}{\partial X_1}[\hat{\beta}_1 X_1] + \frac{\partial \hat{Y}}{\partial X_1}[\hat{\beta}_2 X_2] + \dots + \frac{\partial \hat{Y}}{\partial X_1}[\hat{\beta}_k X_k] \\ \frac{\partial \hat{y}}{\partial X_1} &= 0 + \hat{\beta}_1 + 0 + \dots + 0 \Rightarrow \frac{\partial \hat{y}}{\partial X_1} = \hat{\beta}_1\end{aligned}$$

So we can see, the effect of X_1 on \hat{Y} is $\hat{\beta}_1$, and this is not affected by any other explanatory variables X_2, \dots, X_n . This shows how multiple linear regression allows us to “hold other variables constant” and isolate the effect of one variable.

Alternatively, we can also use partial derivatives to find the effect of one variable in a moderating effect regression:

$$\begin{aligned}\hat{Y} &= \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 \\ \frac{\partial \hat{Y}}{\partial X_1} &= 0 + \hat{\beta}_1 + 0 + \hat{\beta}_3 X_2 \\ \frac{\partial \hat{Y}}{\partial X_1} &= \hat{\beta}_1 + \hat{\beta}_3 X_2\end{aligned}$$

As you can see, the relationship between X_1 and Y here depends on the value of X_2 . In more intuitive words, given a one unit increase in X_1 , there is an expected $\hat{\beta}_1 + \hat{\beta}_3 X_2$ increase in Y .

Chapter 4

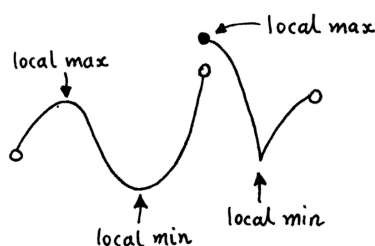
Optimisation

4.1 Local Extrema and Stationary Points

💡 Key Definition: Maximum and Minimum

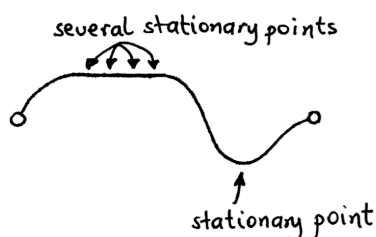
Maximums and minimums are quite intuitive concepts - for function f , the maximum is input that gives the highest output value, and the minimum is the input that gives the lowest output value. More formally:

- A **local maximum** of function $f : I \rightarrow \mathbb{R}$ is a point $a \in I$ with the property $f(a) \geq f(x)$ for all x sufficiently near $x = a$. A **strict local maximum** has the condition $f(a) > f(x)$ for all x sufficiently near $x = a$
- A **local minimum** of function $f : I \rightarrow \mathbb{R}$ is a point $a \in I$ with the property $f(a) \leq f(x)$ for all x sufficiently near $x = a$. A **strict local minimum** has the condition $f(a) < f(x)$ for all x sufficiently near $x = a$



Take an open interval set I , and function $f : I \rightarrow \mathbb{R}$. A **stationary point** at $a \in I$ exists if $f'(a) = 0$.

More intuitively, a stationary point is when the function is neither increasing or decreasing, and since the derivative function is a measure of the rate of change, a function neither increases or decreases when $f'(a) = 0$. Visually:



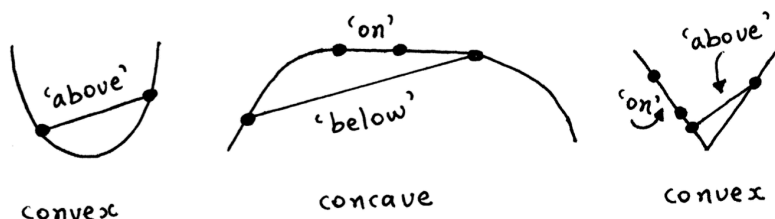
4.2 Concavity and Inflection Points

A function $f : I \rightarrow \mathbb{R}$ is called **convex**, if a line segment joining any two points on the graph of f lies above or on the graph of f .

A function $f : I \rightarrow \mathbb{R}$ is called **concave**, if a line segment joining any two points on the graph of f lies below or on the graph of f .

A function $f : I \rightarrow \mathbb{R}$ is **strictly convex/concave**, if a line segment joining any two points on the graph of f lies always above, or always below, the graph of f .

Below are three figures of convex and concave functions. The first function is strictly convex:

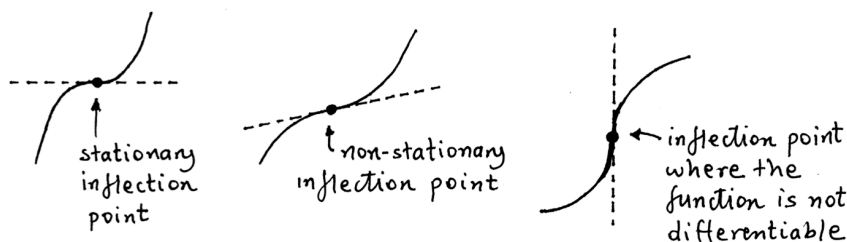


💡 Key Definition: Concavity with Second Derivatives

If function $f : I \rightarrow \mathbb{R}$ is differentiable on an interval I , we also have an alternative definition of concavity.

- If for function f in interval I , $f''(x) > 0$, then the function is convex.
- If for function f in interval I , $f''(x) < 0$, then the function is concave.

If a function f changes from convex to concave at point a , or from concave to convex (or in other words, $f''(a) = 0$ and it changes sign before and after a), that point is an **inflection point**. Inflection points do not always have to be differentiable, just the concavity has to change. Inflection points can also be stationary points, but not always, if $f'(x) = f''(x) = 0$.



4.3 Global Extrema

Take the function $f : A \rightarrow \mathbb{R}$, in which A is domain of f such that $A \subseteq \mathbb{R}$.

A **global maximum** of f is a point $a \in A$ such that $f(a) \geq f(x)$ for all $x \in A$. A **strict global maximum** is when the condition $f(a) > f(x)$ for all $x \in A$.

A **global minimum** of f is a point $a \in A$ such that $f(a) \leq f(x)$ for all $x \in A$. A **strict global minimum** is when the condition $f(a) < f(x)$ for all $x \in A$.

What is the difference between a local and global extrema?

Well, a local extrema only has to be the maximum/minimum in its “neighbourhood”, meaning there can be multiple local extrema. There is only one global minimum and maximum in a function.

Global extrema are a subset of local extrema. That means we can find global extrema by first finding local extrema, then eliminating some local extrema that are not global extrema. There can only be one global maximum or minimum in a function.

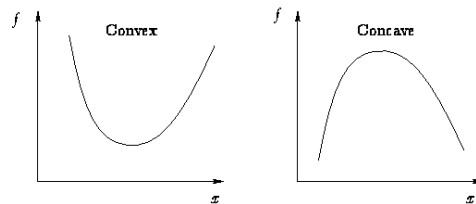
4.4 Optimisation

💡 Key Definition: Optimisation

Optimisation is about finding the maximum or minimum point of a function. This is useful because we often want to minimise error, or maximise utility, or so on.

If a function is differentiable on an open interval, then a local extrema can only occur at a stationary point of this function. Since global extrema are a subset of local extrema, that means that global extrema must also occur at a stationary point of the function.

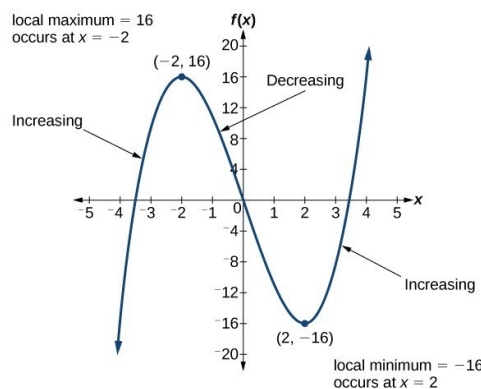
For a stationary point to be a local maximum, the function should be concave. For a stationary point to be a local minimum, the function should be convex. Why? - consider the following diagram:



Thus, we can identify local extrema of a function $f(x)$ in an interval with these steps:

1. Find $f'(x)$, then find for what x makes $f'(x) = 0$. These are the stationary points.
2. Find $f''(x)$. Plug in the x we solved for above. If $f''(x) > 0$, the function is convex, thus there is a local minimum. If $f''(x) < 0$, the function is concave, thus there is a local maximum.
3. Find which extrema is most extreme - that is likely (but not guaranteed) to be the extrema in the interval.

However, we are not done - we have to calculate the edge cases. These are the end-points of our interval. Why? - look at the following figure below - on the interval $[-3, 7]$, clearly $x = 7$ will have a larger maximum value than any local maximum, since the function tends towards infinity as x approaches infinity:



Thus, we have to check the edges of our intervals and see if they are the maximum or minimum in the interval, and compare them to the local extrema we found. Once we have found the x value that makes f output its maximum/minimum value within an interval, we have completed optimisation.

Example

Find the maximum value of $f(x) = 2x^3 - 9x^2 + 12x$ in the interval $x \in [0, 3]$.

1. Find the first derivative: $f'(x) = 6x^2 - 18x + 12$.
 2. Find when $f'(x) = 0$. Thus, $0 = 6x^2 - 18x + 12$, factor so $\implies 0 = 6(x - 2)(x - 1)$. Thus, the two x values that make $f'(x) = 0$ are $x = 2, x = 1$.
 3. Find the 2nd derivative to distinguish maximums and minimums: $f''(x) = 12x - 18$.
 4. Plug in our stationary points $x = 2, x = 1$ we found earlier to find minimums and maximums: $f''(2) = 6$, $f''(1) = -6$. Since maximums are when $f''(x) < 0$, thus $x = 1$ is the only local maximum. We can plug in $x = 1$ into the original function to find the local maximum value of $f(1) = 5$.
 5. Test our edge cases of our interval $[0, 3]$. $f(0) = 0$, $f(3) = 9$. We can see that $f(3) = 9 > f(1) = 5$. Thus, $x = 3$ is the maximum point in this interval.
-

4.5 Constrained Optimisation

Sometimes, we have constraints when doing optimisation. This is best demonstrated with an example.

Take this scenario: a company wants to spend money on advertising. They can choose to spend x amount of money on TV advertising, and y amount of money on YouTube advertising. $f(x, y) = xy$ is a function where the output is the increase in sales as a result of advertising. However, the company only has a total budget of \$10 to spend on both, thus $x + y = 10$.

Simply, we want to maximise $f(x, y) = xy$, given the constraint $x + y = 10$.

1. Start by solving for one variable within the constraint. Let us solve for y , so $x + y = 10 \implies y = 10 - x$.
2. Now, substitute the constraint into the function we want to maximise: $f(xy) = xy \implies f(x) = x(10 - x) \implies f(x) = -x^2 + 10x$.
3. Find the first derivative of our constraint function: $f'(x) = -2x + 10$.
4. Then, find when $f'(x) = 0$ for stationary points: $0 = -2x + 10 \implies x = 5$.
5. Now, find second derivative to see if our stationary point $x = 5$ is a maximum. $f''(x) = -2$, thus $f''(5) = -2$, thus $f''(5) < 0$ so it is a maximum.
6. Now that we know $x = 5$ is a local maximum, plug into constraint to find value of y : $y = 10 - x \implies y = 5$.
7. Thus, $x = 5, y = 5$ is the maximum of function $f(x, y) = xy$. Plugging in $x = 5, y = 5$, then the maximum output value if $f(5, 5) = 25$.

Thus, the general workflow of constrained optimisation is:

1. Solve for one variable in the constraint.
2. Plug into our function we want to optimise, and complete the optimisation process as normal.
3. Once we find the maximum value of one variable, plug into the constraint to find the other.

Chapter 5

Integrals

5.1 Indefinite Integrals

Integrals, also called anti-derivatives, are the opposite of derivatives - i.e. applying an integral to a derivative results in the original function as they cancel each other out.

Key Definition: Indefinite Integral

Let $X, Y \subseteq \mathbb{R}$. Let function $f : X \rightarrow Y$ be a function.

An **indefinite integral** of $f : X \rightarrow Y$ is a function F of form:

$$F(x) = \int f(x)dx$$

Such that its derivative is function f :

$$\frac{d}{dx} \left(\int f(x)dx \right) = f(x)$$

So essentially, integrals reverse derivatives, and derivatives reverse integrals.

Note: there are many indefinite integrals F , for any given function f . This is because when we take a derivative, we lose any constants, since $\frac{d}{dx}[c] = 0$.

- Many different functions F with different c , when taking the derivative, will result in function f .
- This is why you will see a $+C$ at the end of integrals - which indicates that any constant C in integral F will result in the same derivative f .

Integration Rules

Since integrals are the opposite of derivatives, we can reverse many derivative rules to find rules of integration:

Constant Rule: $\int \alpha f(x)dx = \alpha \int f(x)$, where α is a constant.

Sum Rule: $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$.

Reverse power rule: $\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C$, when constant $\alpha \neq -1$.

Nested power rule: $\int [f(x)]^\alpha f'(x)dx = \frac{1}{\alpha+1} [f(x)]^{\alpha+1} + C$.

Positive natural log integral: $\int x^{-1}dx = \ln(x) + C$, when $x > 0$.

Negative natural log integral: $\int x^{-1}dx = \ln(-x) + C$, when $x < 0$.

Reverse nested log integral: $\int \frac{f'(x)}{f(x)} = \ln(f(x)) + C$.

Exponential integral: $\int e^x dx = e^x + C$.

Nested exponential integral: $\int e^{f(x)} f'(x) dx = e^{f(x)} + C$.

Cosine integral: $\int \cos(x) dx = \sin(x) + C$.

Sine integral: $\int \sin(x) dx = -\cos(x) + C$.

Arctan integral: $\int \frac{1}{x^2+1} dx = \arctan(x) + C$.

Arcsine integral: $\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin(x) + C$, when $x \in (-1, 1)$.

Example 1:

Find $\int f(x)$ where $f(x) = 3x^2 + 2x + 1$:

1. Using sum rule, we know: $\int f(x) = \int 3x^2 dx + \int 2x dx + \int 1 dx$.
2. Using constant rule, we know: $\int f(x) = 3 \int x^2 dx + 2 \int x dx + \int 1 dx$.
3. Using reverse power rule, we know: $\int f(x) = 3 \left(\frac{1}{2+1} x^{2+1} \right) + 2 \left(\frac{1}{1+1} x^{1+1} \right) + x + C$.
4. Simplifying: $\int f(x) = 3 \left(\frac{1}{3} x^3 \right) + 2 \left(\frac{1}{2} x^2 \right) + x + C$.
5. Multiply out: $\int f(x) = x^2 + x^2 + x + C$.

Example 2:

Find $\int(fx)$ where $f(x) = e^x e^{e^x}$:

1. We know that $\frac{d}{dx}[e^x] = e^x$.
2. Using nested exponential integral rule: $\int e^{f(x)} f'(x) dx = e^{f(x)} + C$, we can define $f(x) = e^x$ and $f'(x) = e^x$.
3. Putting it into this form, we get $\int f(x) = e^{e^x} + C$.

5.2 Definite Integrals and Riemann Sums

Integrals are not only the opposite of derivatives, but are also a function for the area under a curve.

For example, say we have some function $f(x)$. The area between $f(x)$ and $y = 0$ can be calculated with the integral $F(x) = \int f(x) dx$.

Key Definition: Definite Integral

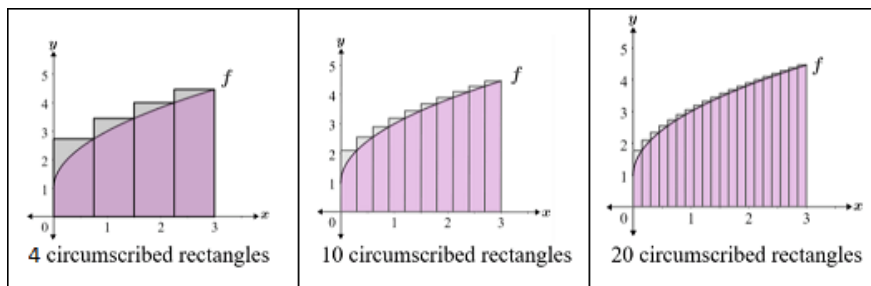
A definite integral is the function F which measures the area under the curve of function f between $x = a$ and $x = b$. We notate the area under the curve of f between a and b as:

$$\int_a^b f(x) dx$$

Now knowing that a definite integral is the area under a curve between a and b , how do we actually evaluate this function and calculate this area?

With a straight linear $f(x)$, this is quite easy. However, when $f(x)$ is non-linear and curved, it might seem quite hard to find the area under the curve.

However, there is a strategy we can use - we can approximate the area under the curve using rectangles. The figure below shows this:



Notice how with more and more rectangles (with each rectangle having a smaller width), the more and more accurate our approximation of the area under a curve becomes.

How do we calculate the area of a rectangle? Of course: width \times height.

- Let us call the width of each rectangle Δx (change in x).
- What is the height of each rectangle? Well, it is the value of $y = f(x)$, the output of the function.
- Thus, the area of each rectangle is $f(x) \times \Delta x$.

Thus, to approximate the area under the curve, we can sum up all the rectangles as follows:

$$\sum_{i=1}^n f(x_i) \Delta x$$

However, as we noted, the smaller the width of each rectangle is, the more accurate our estimation of the area. How do we make the width of each rectangle infinitely smaller?

We can use a limit! With $\lim_{\Delta x \rightarrow 0}$, we can infinitely reduce the width of each rectangle, until we have a perfect measurement of the area under the curve.

💡 Key Definition: Riemann Integral

The **Riemann integral** is a way to calculate the definite integral - the area under the function f between points a and b . Mathematically:

$$\int_a^b f(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{i=1}^n f(x_i) \Delta x$$

A Riemann Integral can, in theory, calculate any definite integral.

A Riemann integral can calculate any integral. However, just like the formal definition of derivatives, sometimes, this can be very difficult to compute.

- Luckily for us, there is an alternative way to calculate definite integrals, that we will discuss in the next section.

5.3 Fundamental Theorem of Calculus

 Key Definition: The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus links definite and indefinite integrals. The theorem states:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Where F is the indefinite integral of function f .

More intuitively, it says that if we evaluate the indefinite integral F , we can then plug in $F(b)$ and $F(a)$ and find the difference, in order to evaluate the definite integral. This allows us to not use Riemann Sums.

Example:

Find the area under the curve between $x = 1$ and $x = 5$ for function $f(x) = 3x^2$.

1. Let us find indefinite integral. Using reverse power rule: $\int f(x)dx = F(x) = x^3 + C$.
 2. Now, fundamental theorem of calculus says $\int_a^b f(x)dx = F(b) - F(a)$.
 3. Let us plug in b and a . $F(b) = F(5) = 5^3 = 125$, and $F(a) = F(1) = 1^3 = 1$.
 4. Find the difference: $F(b) - F(a) = F(5) - F(1) = 125 - 1 = 124$.
 5. Thus, $\int_1^5 3x^2 dx = 124$.
-

5.4 Integration by Substitution

Integration by substitution is a technique that allows us to solve more complex integrals.

 Key Definition: Integration by Substitution

Integration by substitution, derived from the chain rule, is a rule that states:

$$\int f[u(x)]u'(x)dx = F[u(x)] + c$$

Or in other words, if we can rearrange a function into the form $f[u(x)]u'(x)$, we know its derivative will be in the form $F[u(x)] + c$.

So, to solve an integral, we need to:

1. Look at our original function in the integral, and see what $f(x)$ and $u(x)$ could make our original function take the form $f[u(x)]u'(x)$. We need to make sure our function in this form is equal to the original function (so we are solving the same problem). This can involve some tinkering, including dividing our new form by a constant (see example 2).
2. Once we have our integral in that form, we know the integral will be equal to $F[u(x)] + c$. So we must find $F(x) = \int f(x)$, and we will have our integral (since $u(x)$ is already defined in the first step).

If we can re-represent our integral in the form $f(x)$'s and $u(x)$'s in the form of integration by substitution, and ensure that the $f(x)$ we chose in that process is able to be integrated, we can find the integral.

Example 1: find $\int (x+4)^5 dx$.

1. Let us think how we can rewrite this in the form $f[u(x)]u'(x)$. We could define $f(x) = u^5$ and $u(x) = x+4$ so that $f[u(x)]$ section looks like $(x+4)^5$.
2. However, we cannot forget about the extra $u'(x)$ at the end of the form $f[u(x)]u'(x)$. What is $u'(x)$. Using simple derivation on $u(x) = x+4$, we know $u'(x) = 1$. Thus, in the form $f[u(x)]u'(x)$, we have $(x+4)^5 \times 1$, which is equivalent to our original equation.
3. Using integration by substitution, we know $\int f[u(x)]u'(x)dx = F[u(x)] + c$. So knowing $f(x)$ and $u(x)$, how do we get the integration answer in the form $F[u(x)] + c$?
4. We need to first find what $F(x) = \int f(x)dx$ is. The integral of u^5 , using reverse power rule, is $\frac{1}{5+1}u^{5+1} + c = \frac{1}{6}u^6 + c$.
5. Now that we know $F(x)$, and $u(x)$ from the very first step, we can put it in the form $F[u(x)] + c$. Thus, our answer is $\frac{1}{6}(x+4)^6 + c$.

Example 2: find $\int e^{4x^2} x dx$.

1. Let us think how we can rewrite this in the form $f[u(x)]u'(x)$. We could define $f(x) = e^u$ and $u(x) = 4x^2$ so the $f[u(x)]$ section looks like e^{4x^2} .
2. However, we cannot forget about the extra $u'(x)$ at the end of the form $f[u(x)]u'(x)$. What is $u'(x)$? Using power rule of $u(x) = 4x^2$, $u'(x) = 8x$. Thus, in the form $f[u(x)]u'(x)$, we know have $e^{4x^2} 8x$.
3. But notice, our new form $e^{4x^2} 8x$ still does not equal our original equation $f(x) = e^{4x^2} x$. We can adjust this by multiplying our new form by $\frac{1}{8}$, getting $\frac{1}{8}e^{4x^2} 8x$.
4. Using integration by substitution, we know $\int f[u(x)]u'(x)dx = F[u(x)] + c$, thus $\frac{1}{8} \int e^{4x^2} 8x dx = F[u(x)]$. But what is $F = \int f(x)$? since $f(x) = e^u$, using integration rules, $F(x) = e^u + c$.
5. Since the integral by substitution is $F[u(x)]$, we input our $u(x) = 4x^2$ into this to get $F[u(x)] = e^{4x^2} + C$.
6. Let us not forget the $\frac{1}{8}$ outside of our interval! Thus, our integral is $\frac{1}{8}e^{4x^2} + c$.

5.5 Integration by Parts

Integration by parts is another technique for solving more difficult integrals.

Key Definition: Integration by Parts

Integration by parts, derived from the product rule, is defined as:

$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$$

So, if we can find some way to represent our initial integral equation in the form $f(x)g'(x)$, and $f'(x)g(x)$ is able to be integrated, we can find the integral.

So, to solve an integral, we need to:

1. Look at our original function in the integral, and see what $f(x)$ and $g'(x)$ could make our original function take the form $f(x)g'(x)$. We need to make sure our function in this form is equal to the original function (so we are solving the same problem).
2. Then we can solve for $\int f'(x)g(x)dx$ and $\int g'(x)dx = g(x)$.
3. With that, we can find the final integral which is in the form $f(x)g(x) - \int f'(x)g(x)dx$.

So, if we can find some way to represent our initial integral equation in the form $f(x)g'(x)$, and $f'(x)g(x)$ is able to be integrated, we can find the integral.

Example 1: Find $\int xe^x dx$.

1. Let us define $f(x) = x$ and $g'(x) = e^x$. This means the form $f(x)g'(x)$ will equal our original equation.
2. Thus, we know $f'(x) = 1$ and $\int g'(x)dx = \int e^x dx = e^x$.
3. That means $\int f'(x)g(x)dx = \int 1e^x dx = \int e^x = e^x + c$.
4. Now with all the parts, we can put into the final form $f(x)g(x) - \int f'(x)g(x)dx$. Thus, our answer is $xe^x - e^x + c$.

Example 2: find $\int \ln(x)dx$.

1. Let us define $f(x) = \ln(x)$ and $g'(x) = 1$. That means the form $f(x)g'(x)$ will equal our original equation.
2. Thus, we know $f'(x) = \frac{1}{x}$ and $g(x) = \int g'(x) = \int 1dx = x$.
3. That means $\int f'(x)g(x)dx = \int \frac{1}{x}x dx = \int 1dx = x + c$.
4. Now with all the parts, we can put into the final form $f(x)g(x) - \int f'(x)g(x)dx$. Thus, our answer is $\ln(x) - x + c$.

Part III

Probability and Statistical Theory

Chapter 6

Probability Theory

6.1 Rules of Probability

6.2 Counting

6.3 Independence and Conditional Probability

6.4 Law of Total Probability

6.5 Bayes' Theorem

Chapter 7

Random Variables

7.1 Discrete Random Variables

7.2 Continuous Random Variables

7.3 Expectations

7.4 Moments

Chapter 8

Distributions

- 8.1 Discrete Uniform and Bernoulli Distribution
- 8.2 Binomial and Poisson Distribution
- 8.3 Continuous Uniform and Exponential Distribution
- 8.4 Normal (Gaussian) Distribution
- 8.5 Normal Approximation of the Binomial Distribution

Chapter 9

Multivariate Random Variables

9.1 Joint Probability Functions

9.2 Marginal Distributions

9.3 Continuous Multivariate Distributions

9.4 Conditional Distributions

9.5 Covariance and Correlation

9.6 Independent Random Variables

9.7 Sums and Products of Random Variables