Calculus Portion of Math Camp: Topics to be Covered, and Some Definitions and Theorems

Ahmer Tarar
Department of Political Science
Texas A&M University
email: ahmertarar@tamu.edu

August 29, 2017

1 Readings

There is no required text for this section of math camp, which will mainly cover calculus. Most of the definitions and theorems below come from *Calculus with Analytic Geometry* (by Larson, Hostetler, and Edwards, 4th ed., Heath, 1993), which was my undergraduate calculus text. (The pages numbers and theorem numbers are from that text and are for my own reference, and shouldn't mean anything to you.) If you would like to use a text, any undergraduate calculus text would suffice, and I believe that a few are available for free online. An advanced text that I recommend you acquire if you plan to pursue game theory or statistical methods at an advanced level is *Mathematics for Economists* (by Simon and Blume, Norton, 1994).

2 Some Uses of Math in Political Science

Statistics: Testing theories using large data sets, or just looking for patterns in data (POLS 602 and 603)

Theory Generation: For example, game theory, social choice theory (POLS 601 and 604)

3 Mathematical Notation

In concisely presenting mathematical definitions and results (**theorems, propositions, lemmas,** etc.), the following is some standard notation.

The symbol $\forall$ means "for all" or "for every".

The symbol $\exists$ means "there exists".

The initials **s.t.** stand for "such that".

The symbol $\Rightarrow$ means "implies". Thus, $A \Rightarrow B$ should be read "A implies B", or "If A, then B", or "If A is true then B is true". [Sometimes interpreted as A is sufficient for B.]

Similarly, $A \Leftarrow B$ should be read "B implies A". [B is sufficient for A.]

Finally, $A \Leftrightarrow B$ means "A implies B and B implies A". Sometimes this is instead read as "A if and only if B". The "if" part refers to $A \Leftarrow B$, and the "only if" part refers to $A \Rightarrow B$.

The abbreviation **"iff"** stands for "if and only if". Thus, "A iff B" should be read "A if and only if B", which could also be written as $A \Leftrightarrow B$. That is, "A iff B" and $A \Leftrightarrow B$ mean exactly the same thing. Journal copy-editors in political science will often assume that "iff" is a typo and change it to "if" (in which case they've chopped off half of what you have proven), so beware.

The abbreviation **"WLOG"** stands for "without loss of generality". This term is sometimes used in proofs, when going through all of the cases is not necessary because the exact same arguments used in a subset of the universe of cases also covers the remaining cases. This term should be used with caution, as sometimes there *is* a loss of generality in restricting yourself to a subset of cases, in which case you have not really proved the result.

The symbol $\neg$ means "not" or "not true". Thus, $\neg A$ should be read "not A" or "A is not true". Similarly, $A \Rightarrow \neg B$ should be read "A implies not B" or "If A is true, then B is not true".

[Note the following simple but important logical result, called the **contrapositive**. If $A \Rightarrow B$, then $\neg B \Rightarrow \neg A$ (and vice-versa). That is, if A implies B, then if B is not true, then A is not true (since if A was true, then B would also be true). If B are **empirical implications** that necessarily follow from a **theory** A, then this is the basis of seeing whether the evidence **falsifies** the theory. The philosopher of science **Karl Popper** famously argued that good theories are those which haven't yet been falsified by the evidence, i.e., it is not the case that $\neg B$. In empirically evaluating theory A, we try to show that $\neg B$, and if we can't, the theory survives. Supportive evidence can't "prove" a theory to be true, because for a number of reasons it is almost never the case that $B \Rightarrow A$. For instance, there may be another theory A' for which it is also the case that $A' \Rightarrow B$.]

The symbol $\sum$ stands for addition. Let $s_1, s_2, \ldots, s_n$ be a sequence of numbers. Then $\sum_{i=1}^{n} s_i = s_1 + s_2 + \ldots + s_n$. The **index variable** i begins at 1 and then continues all the way to n, in increments of 1. If the sequence is infinitely long, then we use the notation $\sum_{i=1}^{\infty} s_i$ for the sum $s_1 + s_2 + \ldots$ Whenever you see a mathematical statement ending in $\ldots$, that means continue the preceding pattern forever.

The symbol $\prod$ is the analogous symbol for multiplication. That is, $\prod_{i=1}^{n} s_i = s_1 \times s_2 \times \ldots \times s_n$. If the sequence is infinite, then $\prod_{i=1}^{\infty} s_i = s_1 \times s_2 \times \ldots$

4 Set Theory

Many mathematical concepts are defined using set theory, so it is foundational.

Definition 1 *A **set** is a collection of distinct objects (of any kind).*

The members of a set are called **elements** of the set.

Common notation involved in expressing sets: capital letters and curly braces. For example, define the set $A = \{$ red, blue, yellow $\}$.

The objects must be distinct: $D = \{$ red, red, blue, yellow $\}$ is not a set.

The symbol $\in$ means "an element of". For instance, for the set we have just defined, it is the case that [red $\in A$], but [pink $\notin A$].

We can define a new set B as $B = \{x \in A \mid x \neq \text{red}\}$, read as "$B$ is the set consisting of all elements x of A such that x is not red" (we use lower-case letters to denote a generic element of a set, usually the letter x). In other words, $B = \{$ blue, yellow $\}$. This is a very convenient way of defining a new set, and we will use it repeatedly in POLS 601 and 604, so get very familiar with it. (Sometimes, it is instead $B = \{x \in A : x \neq \text{red}\}$.)

A **subset** of a set. Let A and B be sets. We say that B is a subset of A, denoted $B \subseteq A$, if $\forall x \in B, x \in A$ (alternatively, $x \in B \Rightarrow x \in A$).

We say that B is a **proper subset** of A, denoted $B \subset A$, if, in addition to the above condition, it is also the case that $\exists y \in A$ s.t. $y \notin B$.

Given A and B as defined above, it is true both that $B \subseteq A$ and $B \subset A$. On the other hand, if $C = \{$ red, brown $\}$, then $C \nsubseteq A$.

The **union** of two or more sets. Let A and B be sets. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.

The way we have defined A, B, and C above, $A \cup B = A$ and $A \cup C = \{$ red, blue, yellow, brown $\}$.

The **intersection** of two or more sets. Let A and B be sets. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

The way we have defined A, B, and C above, $A \cap B = B$ and $A \cap C = \{$ red $\}$.

The **null set**: $\emptyset = \{\}$. Also called the **empty set**. Note that for the above sets, $B \cap C = \emptyset$.

Set subtraction: the symbol is $\setminus$. Let A and B be sets. $A \setminus B = \{x \in A \mid x \notin B\}$.

The way we have defined A, B, and C above, $A \setminus B = \{$ red $\}$ and $A \setminus C = \{$ blue, yellow $\}$.

The **cardinality** of a set: number of elements. A set can be **finite** (e.g., A, B, and

C above) or **infinite**: either **uncountably infinite** (e.g., the set of all real numbers) or **countably infinite** (e.g., the set of positive integers).

5 Some Important Pre-Calculus Topics

Real numbers: all numbers other than the **complex numbers** (also called **imaginary numbers**, involving $i = \sqrt{-1}$), so all numbers that we actually deal with and are familiar with

The *set* of real numbers is denoted by $\mathbb{R}$, i.e., $\mathbb{R} = \{x \mid x \text{ is a real number}\}$. This is graphically represented by the **real line**

We denote the set of **non-negative** real numbers as $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$, and the set of **positive** real numbers as $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$ (sometimes we use the term **strictly positive** to emphasize that we are not including 0). If we want to be fancy, note that we can alternatively define $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$

We define **non-positive** and **negative** analogously

The set of **integers** (also called whole numbers) is denoted $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$

The set of **rational numbers** is defined and denoted as $\mathbb{Q} = \{\frac{x}{y} \in \mathbb{R} \mid x, y \in \mathbb{Z}, y \neq 0\}$, i.e., the set of real numbers that can be expressed as the ratio of two integers

We can talk about $\mathbb{Z}_+$, $\mathbb{Z}_{--}$, $\mathbb{Q}_-$, $\mathbb{Q}_{++}$, etc.

Infinity (∞) and **negative infinity** ($-\infty$): are concepts, not numbers

For any $a, b \in \mathbb{R}$ such that $a < b$, we define the **closed interval** $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ and the **open interval** $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$

We can also talk about **half-closed** (alternatively, **half-open**) intervals: $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ and $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$

The following is a way of thinking about the real numbers: $\mathbb{R} = (-\infty, \infty)$

Division by 0 is **undefined**

The **absolute value** of a real number: for any $x \in \mathbb{R}$, $\mid x \mid= x$ if $x \geq 0$, $\mid x \mid= -x$ if $x < 0$ (i.e., is just the *magnitude* of the number)

The **distance** between two real numbers $x, y \in \mathbb{R}$: $d(x, y) =\mid x - y \mid$

The **Cartesian plane**, or **2-dimensional space**: $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$. The set of all **ordered pairs** of numbers: $(2, 1) \neq (1, 2)$. These are distinct **points** in the Cartesian plane.

x-axis and **y-axis**

Points in the Cartesian plane: $x = (x_1, x_2)$, $y = (y_1, y_2)$. **x-coordinate** and **y-coordinate**.

The **equation** of a **straight line** in the Cartesian plane: $y = mx + c$

m = **slope**, c = **y-intercept**, $-\frac{c}{m}$ = **x-intercept** (all fixed numbers, also called **constants**; x and y are called **variables**)

Slope $= \frac{\Delta y}{\Delta x}$

Plotting a straight line (just need two points)

The **distance** between two points $x, y \in \mathbb{R}^2$ (i.e., two points in the Cartesian plane) is defined as $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. This formula is a simple application of the **Pythagorean theorem**

A real-valued function of a real number (a **curve**; generalizing a straight line) [DEFINITION]

Definition 2 *A **function** f from a set X into a set Y (i.e., $f : X \to Y$) is a correspondence that assigns to each element x in X exactly one element y in Y. We call y the **image** of x*

under f and denote it by $f(x)$. The ***domain*** *of f is the set X, the* ***codomain*** *is the set Y, and the* ***range*** *consists of all images of elements in X.*

When $X, Y \subseteq \mathbf{R}$, then f is a **real-valued function of a real variable**, e.g., $y = f(x) = -2x + 5$ or $y = f(x) = x^2$

For such a function $y = f(x)$, we sometimes refer to y as the **dependent variable** and x as the **independent variable**

Linear function: $y = mx + c$, $m \neq 0$ (if $m = 0$, $y = c$ is a **constant** function)

Quadratic function: $y = ax^2 + bx + c$, $a \neq 0$. Also called a **parabola**. (Its curve is shaped like a hill if $a < 0$, and an upside-down hill if $a > 0$)

The **roots** of a function: the value(s) of x where the function (i.e., y) is 0 (for a linear function, the same as the x-intercept)

A quadratic function can have 0, 1, or 2 roots. If it has any roots, they are given by the **quadratic formula**: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. If $b^2 - 4ac < 0$, it has no roots (doesn't touch the x-axis). If $b^2 - 4ac = 0$, it has one root (touches the x-axis at one point), and if $b^2 - 4ac > 0$, it has two roots (touches the x-axis at two points)

More generally, **polynomial functions**: $y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x^1 + c$, where $n \geq 1$ is an integer (if we want to be concise, $n \in \mathbb{Z}_{++}$), c and the a_i's are constants, and at least one of the a_i's is not 0 ($i = 1, 2, \ldots, n$). If $a_n \neq 0$, then y is called an **n-th degree polynomial**. Linear functions are first-degree polynomials, quadratics are 2nd-degree polynomials, cubics are 3rd-degree polynomials, and so on.

Most of the functions you will encounter in your methods courses are polynomials. This is good, because we will see that polynomials have some nice properties (e.g., continuous and differentiable) that make them easy to deal with. If you have to find the roots of a polynomial, anything higher-degree than a quadratic can be quite tricky.

domain and range of a function

one-to-one function

onto function

Definition 3 *If to each value in its range there corresponds exactly one value in its domain, the function is said to be **one-to-one**. Moreover, if the range of f consists of all Y, then the function is said to be **onto**.*

Correspondence: generalization of a function, assigns a *subset* of Y to every $x \in X$.

Composite function

Definition 4 *Let $g : X \to Y$ and $f : Y \to Z$ be functions. Then the **composition** of g and f, or **composite function**, is a function $f \circ g : X \to Z$ defined by $(f \circ g)(x) = f(g(x))$ for all $x \in X$.*

As an example, suppose that $X = Y = Z = \mathbb{R}$, $g(x) = x^2$, and $f(x) = \frac{1}{x}$. Then $(f \circ g)(x) = \frac{1}{g(x)} = \frac{1}{x^2}$ for all $x \in \mathbb{R}$.

how to define and determine the slope of a function that is not a straight line: leads into the notion of the derivative of a function

6 Limits, Their Properties, Continuity

intuitive notion of limit: what does the function $y = f(x) = x^2$ approach as x approaches 1?

graphs showing that the limit of a function at a point need not correspond to the value of the function at that point i.e., limits cannot always be found by direct substitution, although they usually can

graphs showing that limit from below and limit from above can differ (also see p.92)

graphs showing that the limit of a function at a point can exist even if the function is not defined at that point

formal definition of the limit of a function (ϵ-δ definition) [DEFINITION]

Definition 5 *(definition of **limit**) Let f be a function defined on an open interval containing c (except possibly at c) and let L be a real number. The statement $\lim_{x \to c} f(x) = L$ means that for each $\epsilon > 0$ there exists a $\delta > 0$ such that $\mid f(x) - L \mid < \epsilon$ whenever $0 < \mid x - c \mid < \delta$.*

THEOREM 2.11 on p.93

Theorem 1 *(the existence of a limit) If f is a function and c and L are real numbers, then the limit of $f(x)$ as x approaches c is L if and only if $\lim_{x \to c^-} f(x) = L$ and $\lim_{x \to c^+} f(x) = L$.*

THEOREM 2.2 on p.77

Theorem 2 *(some basic limits) If b and c are real numbers and n is a positive integer, then the following properties are true.*

(1) $\lim_{x \to c} b = b$.

(2) $\lim_{x \to c} x = c$.

(3) $\lim_{x \to c} x^n = c^n$.

THEOREM 2.3 on p.78

Theorem 3 *(properties of limits) Let b and c be real numbers, and n a positive integer, and let f and g be functions whose limits exist as $x \to c$. Then the following properties are true.*

(1) Scalar multiple: $\lim_{x\to c}[b(f(x))] = b[\lim_{x\to c} f(x)]$.

(2) Sum: $\lim_{x\to c}[f(x) + g(x)] = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$.

(3) Difference: $\lim_{x\to c}[f(x) - g(x)] = \lim_{x\to c} f(x) - \lim_{x\to c} g(x)$.

(4) Product: $\lim_{x\to c}[f(x)g(x)] = [\lim_{x\to c} f(x)][\lim_{x\to c} g(x)]$.

(5) Quotient: $\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$ *(provided* $\lim_{x\to c} g(x) \neq 0$*).*

(6) Power: $\lim_{x\to c}[f(x)]^n = [\lim_{x\to c} f(x)]^n$.

THEOREMS 2.4 and 2.5 on p.79

Theorem 4 *(limit of a polynomial function) If p is a polynomial function and c is a real number, then* $\lim_{x\to c} p(x) = p(c)$.

Theorem 5 *(limit of a rational function) If r is a rational function given by* $r(x) = \frac{p(x)}{q(x)}$ *and c is a real number such that $q(c) \neq 0$, then* $\lim_{x\to c} r(x) = r(c) = \frac{p(c)}{q(c)}$.

THEOREMS 2.6 and 2.7 on p.80

Theorem 6 *(limit of a function involving a radical) If $c > 0$ and n is any positive integer, or if $c \leq 0$ and n is an odd positive integer, then* $\lim_{x\to c} \sqrt[n]{x} = \sqrt[n]{c}$.

Theorem 7 *(limit of a composite function) If f and g are functions such that* $\lim_{x\to c} g(x) = L$ *and* $\lim_{x\to L} f(x) = f(L)$*, then* $\lim_{x\to c} f(g(x)) = f(L)$.

we can usually evaluate limits by direct substitution, but when we cannot, other techniques exist

cancellation, example 1 on p.83

intuitive notion of what it means for a function to be continuous at a point continuous function [DEFINITION]

also definition on p.93

Definition 6 *A function f is called* ***continuous at*** *c if the following 3 conditions are all met.*

(1) $f(c)$ is defined.

(2) $\lim_{x\to c} f(x)$ exists.

(3) $\lim_{x\to c} f(x) = f(c)$.

Definition 7 *A function is called* ***continuous on an open interval (a,b)*** *if it is continuous at each point in the interval. A function that is continuous on the entire real line $(-\infty, \infty)$ is called* ***everywhere continuous****.*

Definition 8 *A function is called* ***continuous on a closed interval [a,b]*** *if it is continuous on the open interval (a, b) and $\lim_{x\to a^+} f(x) = f(a)$ and $\lim_{x\to b^-} f(x) = f(b)$. The function f is called* ***continuous from the right*** *at a and* ***continuous from the left*** *at b.*

Theorem 2.12

Theorem 8 *(properties of continuity) If b is a real number and f and g are continuous at $x = c$, then the following functions are also continuous at c.*

(1) scalar multiple: bf

(2) sum: $f + g$

(3) difference: $f - g$

(4) product: fg

(5) quotient: $\frac{f}{g}$ *(provided* $g(c) \neq 0$*)*

Theorem 2.13

Theorem 9 *(continuity of a composite function) If* g *is continuous at* c *and* f *is continuous at* $g(c)$*, then the composite function given by* $(f \circ g)(x) = f(g(x))$ *is continuous at* c*.*

Theorem 2.14

Theorem 10 *(Intermediate Value Theorem) If* f *is continuous on* $[a, b]$ *and* k *is any number between* $f(a)$ *and* $f(b)$*, then there is at least one number* c *in* $[a, b]$ *such that* $f(c) = k$*.*

Infinite limits (DEFINITION)

Definition 9 *(definition of infinite limits) The statement* $\lim_{x \to c} f(x) = \infty$ *means that for each* $M > 0$ *there exists a* $\delta > 0$ *such that* $f(x) > M$ *whenever* $0 <| x - c |< \delta$*. Similarly, the statement* $\lim_{x \to c} f(x) = -\infty$ *means that for each* $N < 0$ *there exists a* $\delta > 0$ *such that* $f(x) < N$ *whenever* $0 <| x - c |< \delta$*.*

Vertical asymptote (DEFINITION)

Definition 10 *If* $f(x)$ *approaches infinity (or negative infinity) as* x *approaches* c *from the right or left, then we call the line* $x = c$ *a* ***vertical asymptote*** *of the graph of* f*.*

Theorem 2.16

Theorem 11 *(properties of infinite limits) If* c *and* L *are real numbers and* f *and* g *are functions such that* $\lim_{x \to c} f(x) = \infty$ *and* $\lim_{x \to c} g(x) = L$*, then the following properties are true.*

(1) Sum: $\lim_{x \to c}[f(x) + g(x)] = \infty$.

(2) Difference: $\lim_{x \to c}[f(x) - g(x)] = \infty$.

(3) Product: $\lim_{x \to c}[f(x)g(x)] = \infty$ *if* $L > 0$, *and* $-\infty$ *if* $L < 0$.

(4) Quotient: $\lim_{x \to c} \frac{g(x)}{f(x)} = 0$.

(5) Similar properties hold for one-sided limits and for functions for which the limit of $f(x)$ *as* x *approaches* c *is* $-\infty$.

p.206

Definition 11 *(Limits at infinity) The statement* $\lim_{x \to \infty} f(x) = L$ *means that for each* $\epsilon > 0$ *there exists an* $M > 0$ *such that* $\mid f(x) - L \mid < \epsilon$ *whenever* $x > M$. *Similarly, the statement* $\lim_{x \to -\infty} f(x) = L$ *means that for each* $\epsilon > 0$ *there exists an* $N < 0$ *such that* $\mid f(x) - L \mid < \epsilon$ *whenever* $x < N$.

Definition 12 *If* $\lim_{x \to -\infty} f(x) = L$ *or* $\lim_{x \to \infty} f(x) = L$ *then the line* $y = L$ *is called a* ***horizontal asymptote*** *of the graph of* f.

THEOREM 4.10

Theorem 12 *(Some properties of limits at infinity) If* $\lim_{x \to \infty} f(x)$ *and* $\lim_{x \to \infty} g(x)$ *both exist, then:*

(1) $\lim_{x \to \infty}[f(x) + g(x)] = \lim_{x \to \infty} f(x) + \lim_{x \to \infty} g(x)$

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

Theorem 13 *If* r *is a positive rational number and* c *is any real number, then* $\lim_{x \to \infty} \frac{c}{x^r} = 0$.

7 Differentiation

motivation: the slope of a curve, or the instantaneous rate of change of y with respect to x

Definition 13 *The **derivative** of a function f at any point x (in its domain) is given by $\lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$, provided the limit exists.*

ALTERNATIVE DEFINITION:

Definition 14 *The **derivative** of a function f at any point x (in its domain) is given by $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$, provided the limit exists.*

Common notation for derivative: $f'(x)$, $\frac{dy}{dx}$, $\frac{d}{dx}[f(x)]$

terminology: the derivative of y with respect to x

Definition 15 *A function is called **differentiable** at x if its derivative exists at x and **differentiable on an open interval** (a, b) if it is differentiable at every point in the interval.*

THEOREM 3.2

Theorem 14 *(differentiability implies continuity) If f is differentiable at $x = c$, then f is continuous at $x = c$.*

can use the definition of the derivative to find it, but more convenient to use differentiation rules.

THEOREM 3.3

Theorem 15 *(constant rule) The derivative of a constant is 0. That is, if c is a real number, then* $\frac{d}{dx}[c] = 0$.

THEOREM 3.4

Theorem 16 *(power rule) If* n *is a rational number, then* $\frac{d}{dx}[x^n] = nx^{n-1}$.

THEOREM 3.5

Theorem 17 *(constant multiple rule) If* f *is a differentiable function and c is a real number, then* $\frac{d}{dx}[cf(x)] = cf'(x)$.

THEOREM 3.6

Theorem 18 *(sum and difference rules) Let* f *and* g *be differentiable functions. Then* $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$ *and* $\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$.

THEOREM 3.8

Theorem 19 *(product rule) Let* f *and* g *be differentiable functions. Then* $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$.

THEOREM 3.9

Theorem 20 *(quotient rule) The quotient of two differentiable functions,* f *and* g*, is itself differentiable at all values of* x *for which* $g(x) \neq 0$. *Moreover,* $\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$ *(provided* $g(x) \neq 0$*).*

The Chain Rule for composite functions

THEOREM 3.11

Theorem 21 *(Chain Rule) If $y = f(u)$ is a differentiable function of u and $u = g(x)$ is a differentiable function of x, then $y = f(g(x))$ is a differentiable function of x and $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$ or, equivalently, $\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$.*

THEOREM 3.12

Theorem 22 *(General Power Rule) If $y = [u(x)]^n$, where u is a differentiable function of x and n is a rational number, then $\frac{dy}{dx} = n[u(x)]^{n-1}\frac{du}{dx}$ or, equivalently, $\frac{d}{dx}[u^n] = nu^{n-1}u'$.*

IMPLICIT DIFFERENTIATION: useful method when we have an equation involving y and x, but we cannot write y explicitly in terms of x.

EXAMPLE: Suppose $y^3 + y^2 - 5y - x^2 = -4$. This equation implicitly defines y as a function of x, but we cannot "isolate" y on one side of the equation and then calculate $\frac{dy}{dx}$.

8 Applications of Differentiation

Application of derivatives: finding the instantaneous rate of change of y with respect to x (slope of a curve at any given point)

Application of derivatives: finding the relative maxima and minima of a function (first and second order conditions)

graphical representation of a quadratic function and relative maxima/minima

Definition 16 *Let f be defined on an interval I containing c. We say that $f(c)$ is the **minimum of** f **on** I if $f(c) \leq f(x)$ for all x in I. We say that c is the **minimizer of** f **on** I.*

Definition 17 *Let f be defined on an interval I containing c. We say that $f(c)$ is the*

maximum of f ***on*** I *if* $f(c) \geq f(x)$ *for all* x *in* I*. We say that* c *is the* ***maximizer of*** f ***on*** I*.*

THEOREM 4.1

Theorem 23 *(Extreme Value Theorem) If* f *is continuous on a closed interval* $[a, b]$*, then* f *has both a minimum and a maximum on the interval.*

Definition 18 *If there is an open interval on which* $f(c)$ *is a maximum, then* $f(c)$ *is called a* ***relative maximum*** *of* f*.*

Definition 19 *If there is an open interval on which* $f(c)$ *is a minimum, then* $f(c)$ *is called a* ***relative minimum*** *of* f*.*

Definition 20 *If* f *is defined at* c*, then* c *is called a* ***critical number*** *of* f *if* $f'(c) = 0$ *or if* f' *is undefined at* c*.*

THEOREM 4.2

Theorem 24 *If* f *has a relative minimum or relative maximum at* $x = c$*, then* c *is a critical number of* f*.*

In practice, with the functions we will work with, relative maxima or minima occur at critical numbers where $f'(c) = 0$ rather than where f' is undefined.

THEOREM 4.3

Theorem 25 *(Rolle's Theorem) Let* f *be continuous on the closed interval* $[a, b]$ *and differentiable on the open interval* (a, b)*. If* $f(a) = f(b)$ *then there is at least one number* c *in* (a, b) *such that* $f'(c) = 0$*.*

THEOREM 4.4

Theorem 26 *(Mean Value Theorem) If f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists a number c in (a, b) such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

Definition 21 *A function f is said to be* ***strictly increasing*** *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) < f(x_2)$.*

A function f is said to be ***weakly increasing*** *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) \leq f(x_2)$.*

Definition 22 *A function f is said to be* ***strictly decreasing*** *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) > f(x_2)$.*

A function f is said to be ***weakly decreasing*** *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) \geq f(x_2)$.*

Definition 23 *A function is called* ***strictly monotonic*** *on an interval if it is either strictly increasing on the entire interval (in which case we say the function is* ***monotone increasing*** *on the interval), or strictly decreasing on the entire interval (in which case we say the function is* ***monotone decreasing*** *on the interval).*

THEOREM 4.5

Theorem 27 *Let f be a function that is differentiable on the interval (a, b).*

(1) If $f'(x) > 0 \ \forall x \in (a, b)$, then f is strictly increasing on (a, b).

(2) If $f'(x) < 0 \ \forall x \in (a, b)$, then f is strictly decreasing on (a, b).

(3) If $f'(x) = 0 \ \forall x \in (a, b)$, then f is constant on (a, b).

THEOREM 4.6

Theorem 28 *(First Derivative Test) Let c be a critical number of a function f that is continuous on an open interval I containing c. If f is differentiable on the interval, except possibly at c, then $f(c)$ can be classified as follows.*

(1) If f' changes from negative to positive at c, then $f(c)$ is a relative minimum of f.

(2) If f' changes from positive to negative at c, then $f(c)$ is a relative maximum of f.

(3) If f' does not change signs at c, then $f(c)$ is neither a relative minimum nor a relative maximum.

Definition 24 *Let f be differentiable on an open interval. We say that the graph of f is* ***concave upward*** *if f' is strictly increasing on the interval and* ***concave downward*** *if f' is strictly decreasing on the interval.*

THEOREM 4.7

Theorem 29 *(Test for concavity) Let f be a function whose second derivative exists on an open interval I.*

(1) If $f''(x) > 0$ for all x in I, then the graph of f is concave upward.

(2) If $f''(x) < 0$ for all x in I, then the graph of f is concave downward.

Definition 25 *Let f be a function whose graph has a tangent line at $(c, f(c))$. The point $(c, f(c))$ is called a* ***point of inflection*** *if the concavity of f changes from upward to downward (or vice versa) at that point.*

THEOREM 4.9

Theorem 30 *(Second derivative test) Let f be a function such that $f'(c) = 0$ and the second derivative of f exists on an open interval containing c.*

(1) If $f''(c) > 0$, then $f(c)$ is a relative minimum.

(2) If $f''(c) < 0$, then $f(c)$ is a relative maximum.

(1) If $f''(c) = 0$, then the test fails.

9 L'Hopital's Rule

Indeterminate forms: $\frac{0}{0}$ and $\frac{\infty}{\infty}$

what if we want to determine a limit, where direct substitution gives one of these forms

Theorem 31 *Let f and g be functions that are differentiable on an open interval (a, b) containing c, except possibly at c itself. If the limit of $\frac{f(x)}{g(x)}$ as x approaches c produces the indeterminate form $\frac{0}{0}$ or $\frac{\infty}{\infty}$, then $\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}$ provided the limit on the right exists (or is infinite). This rule also applies if c is ∞ or $-\infty$.*

10 Functions of Several Variables

Definition 26 *Let D be a set of ordered pairs of real numbers, i.e., $D \subseteq \mathbb{R}^2$. If to each ordered pair (x, y) in D there corresponds a real number $f(x, y)$, then f is called a **function of x and y**. The set D is the **domain** of f, and the corresponding set of values for $f(x, y)$ is the **range** of f.*

If $z = f(x, y)$, x and y are called independent variables, and z is the dependent variable

2-dimensional analog of an open interval: an open disc

Definition 27 *The **open disc** of **radius** $\delta > 0$ centered at the point (x_0, y_0) is the set*

$\{(x, y) \mid \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta\}$.

Intuitively, the set of all points within distance δ of (x_0, y_0).

Definition 28 *The* ***closed disc*** *of* ***radius*** $\delta > 0$ *centered at the point* (x_0, y_0) *is the set* $\{(x, y) \mid \sqrt{(x - x_0)^2 + (y - y_0)^2} \leq \delta\}$.

Sometimes the term "ball" is used instead of "disc"

Sometimes the disc is called the **δ-neighborhood** about (x_0, y_0), and we can talk about a closed neighborhood and an open one.

Definition 29 *A point* (x_0, y_0) *in a plane region* R *is an* ***interior point*** *of* R *if there exists a* δ*-neighborhood around* (x_0, y_0) *that lies entirely in* R.

Definition 30 *If every point in a plane region* R *is an interior point, then we call* R *an* ***open region***.

Definition 31 *A point* (x_0, y_0) *in a plane region* R *is a* ***boundary point*** *of* R *if every open disc centered at* (x_0, y_0) *contains points inside* R *as well as points outside* R.

Definition 32 *If a region contains all its boundary points, then we say that the region is* ***closed***.

A plane region that contains some but not all of its boundary points is neither open nor closed, and is analogous to the half-open interval $[a, b)$ or $(a, b]$.

Definition 33 *A region in the plane is called* ***bounded*** *if it is a subregion of a closed disc in the plane.*

Definition 34 *(limit of a function of two variables) Let f be a function of two variables defined, except possibly at (x_0, y_0), on an open disc centered at (x_0, y_0), and let L be a real number. Then $\lim_{(x,y)\to(x_0,y_0)} f(x, y) = L$ if for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mid f(x, y) - L \mid < \epsilon$ whenever $0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$.*

Definition 35 *A function f of two variables is* ***continuous at a point*** *(x_0, y_0) in an open region R if $f(x_0, y_0)$ is defined and is equal to the limit of $f(x, y)$ as (x, y) approaches (x_0, y_0). That is, $\lim_{(x,y)\to(x_0,y_0)} f(x, y) = f(x_0, y_0)$. The function f is* ***continuous in the open region*** *R if it is continuous at every point in R.*

Polynomial functions are continuous everywhere in their domains, and these are the most common types of functions encountered in political science applications.

11 Partial Derivatives

Definition 36 *If $z = f(x, y)$, then the* ***first partial derivatives*** *of f with respect to x and y are the functions f_x and f_y defined by $f_x(x, y) = \lim_{\Delta x\to 0} \frac{f(x+\Delta x,y)-f(x,y)}{\Delta x}$ and $f_y(x, y) = \lim_{\Delta y\to 0} \frac{f(x,y+\Delta y)-f(x,y)}{\Delta y}$, provided the limits exist.*

To find f_x, we treat y as a constant and differentiate with respect to x, using the same differentiation rules as before (for single-variable functions).

To find f_y, we treat x as a constant and differentiate with respect to y.

Common notation for partial derivatives, with $z = f(x, y)$: $f_x(x, y)$, $\frac{\partial z}{\partial x}$, $\frac{\partial}{\partial x} f(x, y)$

If we plot the function $z = f(x, y)$, it is a surface in 3-dimensional Euclidean space as opposed to a curve in 2-dimensional Euclidean space, and we interpret $\frac{\partial z}{\partial x}$ as a function that gives the slope of the surface **in the x direction** (i.e., along the direction of the x-axis) at any given point (x_0, y_0, z_0) on the surface

It is also interpreted as the instantaneous amount by which z increases (or decreases, if the partial is negative) as x increases by 1 unit, **holding the value of y constant**

also typically called "marginal effects"

Higher-order partial derivatives: $\frac{\partial}{\partial x}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial x^2} = f_{xx}$

$\frac{\partial}{\partial y}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial y^2} = f_{yy}$

$\frac{\partial}{\partial y}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial y \partial x} = f_{xy}$ (mixed partial derivative)

$\frac{\partial}{\partial x}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial x \partial y} = f_{yx}$ (mixed partial derivative)

THEOREM 14.3

Theorem 32 *(equality of mixed partial derivatives) If f is a function of x and y and f, f_x, f_y, f_{xy}, and f_{yx} are continuous on an open region R, then for every (x, y) in R, $f_{xy}(x, y) = f_{yx}(x, y)$.*

The continuity condition is satisfied for the polynomial functions that we deal with in political science applications.

THEOREM 14.15

Theorem 33 *(Extreme Value Theorem) Let f be a continuous function of two variables x and y defined on a closed bounded region R in the xy-plane.*

(1) There is at least one point in R where f takes on a minimum value.

(2) There is at least one point in R where f takes on a maximum value.

Definition 37 *Let f be a function defined on a region R containing (x_0, y_0).*

*(1) $f(x_0, y_0)$ is a **relative minimum** of f if $f(x, y) \geq f(x_0, y_0)$ for all (x, y) in an open disc containing (x_0, y_0).*

*(2) $f(x_0, y_0)$ is a **relative maximum** of f if $f(x, y) \leq f(x_0, y_0)$ for all (x, y) in an open disc containing (x_0, y_0).*

Definition 38 *Let f be a function defined on an open region R containing (x_0, y_0). We call (x_0, y_0) a **critical point** of f if one of the following is true.*

(1) $f_x(x_0, y_0) = 0$ and $f_y(x_0, y_0) = 0$.

(2) $f_x(x_0, y_0)$ or $f_y(x_0, y_0)$ does not exist.

THEOREM 14.16

Theorem 34 *If $f(x_0, y_0)$ is a relative extremum of f on an open region R, then (x_0, y_0) is a critical point of f.*

THEOREM 14.17

Theorem 35 *(Second-partials test) Let f have continuous first and second partial derivatives on an open region containing a point (a, b) for which $f_x(a, b) = 0$ and $f_y(a, b) = 0$. To test for relative extrema of f, we define the quantity $d = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$.*

(1) If $d > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a relative minimum.

(2) If $d > 0$ and $f_{xx}(a, b) < 0$, then $f(a, b)$ is a relative maximum.

*(3) If $d < 0$, then $(a, b, f(a, b))$ is a **saddle point**.*

(4) The test gives no information if $d = 0$.

All of the above can be extended in a straightforward way to functions of 3 or more variables, in which case we have to use some matrix notation.

Application: the **method of least squares**

We believe some variable y is a function of some variable x, and that the function is a linear one

We have data on x and y, i.e. a set of n "observations" $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

We want to plot the straight line $y = f(x) = ax + b$ that is the "best fit" to this data, i.e., we want to determine the "best fit" values of a and b

The "least squares" method of determining the best fit is to minimize the **sum of the squared errors** $S = \sum_{i=1}^{n}[f(x_i) - y_i]^2 = \sum_{i=1}^{n}[ax_i + b - y_i]^2$

The x_i's and the y_i's are fixed numbers, and hence S is a function of two variables, a and b, i.e., $S = S(a, b)$

We want to find the point (a_0, b_0) where $S(a, b)$ achieves its minimum value

Partial derivative of S with respect to a: $S_a(a, b) = \sum_{i=1}^{n} 2x_i[ax_i + b - y_i] = 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i$

Partial derivative of S with respect to b: $S_b(a, b) = \sum_{i=1}^{n} 2[ax_i + b - y_i] = 2a\sum_{i=1}^{n} x_i + 2nb - 2\sum_{i=1}^{n} y_i$

Setting these equal to 0 gives 2 equations in 2 unknowns (a_0 and b_0), and solving these 2 **simultaneous equations** gives $a_0 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$ and $b_0 = \frac{1}{n}[\sum_{i=1}^{n} y_i - a_0\sum_{i=1}^{n} x_i]$

We can apply the second-partials test to verify that these values of a_0 and b_0 yield a minimum of S

12 Constrained Optimization

This is an advanced topic, and we will only discuss it briefly.

Consider the function $z = f(x, y) = -x^2 - y^2$. A moment's reflection should convince you that the plot of this function is a 3-dimensional symmetric "hill" whose peak (maximum) is at the point $(x, y) = (0, 0)$, at which $z = f(0, 0) = 0$. That is, this is the global maximum, and the function has no minimum. You can use Theorem 35 to verify this formally.

Suppose we want to find the maximum of this function under the **equality constraint** $2x + y = 1$. That is, we want to solve the following problem.

Maximize $f(x, y) = -x^2 - y^2$ subject to $2x + y = 1$.

Why might we impose the constraint $2x + y = 1$? Perhaps for some reason, this relationship between x and y has to hold. Alternatively, suppose we define the domain of f to be $X = \{(x, y) \in \mathbb{R}^2 \mid 2x + y = 1\}$.

Graphically, we visualize this problem as follows. $f(x, y)$ is defined on all of $\mathbb{R}^2$. Suppose, however, we consider the restriction of this function to the straight line (in the x-y plane) $2x + y = 1$, or $y = 1 - 2x$. The resulting plot, instead of being a 3-dimensional surface, will be a 2-dimensional curve (where the "x-axis" is the straight line $y = 1 - 2x$, and the "y-axis" is the z-axis). We want to find the maximum of this curve.

One way to solve this problem is by direct substitution. That is, plug $y = 1 - 2x$ into $f(x, y) = -x^2 - y^2$ to get the new function $f^*(x) = -x^2 - (1 - 2x)^2$. Then we just engage in unconstrained maximization of this function with respect to x; we can engage in unconstrained optimization because the constraint is already embodied in the substitution, and hence we can now ignore it. We can simplify $f^*(x) = -x^2 - (1 + 4x^2 - 4x) = -5x^2 + 4x - 1$. $f^{*'}(x) = -10x + 4$, and setting this equal to 0 (the FOC) gives $x = \frac{2}{5}$. The SOC is $f^{*''}(x) = -10 < 0$, and hence this is a maximum. Finally, we substitute $y = 1 - 2x = 1 - 2(\frac{2}{5}) = \frac{1}{5}$. Thus, $(x, y) = (\frac{2}{5}, \frac{1}{5})$ is the solution to the constrained maximization problem, and at this point, $f(\frac{2}{5}, \frac{1}{5}) = -(\frac{2}{5})^2 - (\frac{1}{5})^2 = -\frac{1}{5}$.

An alternative method, which is especially useful for more complicated constrained maximization problems, is the **Method of Lagrange Multipliers**. Recall that we want to solve the following problem.

Maximize $f(x, y) = -x^2 - y^2$ subject to $2x + y = 1$.

More generally, suppose we want to maximize some function $f(x, y)$ subject to the equality constraint $g(x, y) = c$, where c is some real number (possibly 0). In the above example, $f(x, y) = -x^2 - y^2$, $g(x, y) = 2x + y$, and $c = 1$.

Then we define a function $L(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$, where L is called the **Lagrangian function** or simply the **Lagrangian**, and λ is called the **Lagrange multiplier**.[1] Note that L is a function of 3 variables instead of 2. Then we find the critical points of L, i.e., the points (x, y, λ) at which $\frac{\partial L}{\partial x} = 0$, $\frac{\partial L}{\partial y} = 0$, and $\frac{\partial L}{\partial \lambda} = 0$ all three hold.

Note that for our example, $L(x, y, \lambda) = -x^2 - y^2 - \lambda(2x + y - 1)$, $\frac{\partial L}{\partial x} = -2x - 2\lambda$, $\frac{\partial L}{\partial y} = -2y - \lambda$, and $\frac{\partial L}{\partial \lambda} = -2x - y + 1$. Setting each of these partials equal to 0 and solving the resulting system of 3 equations in 3 unknowns gives us $x = \frac{2}{5}$, $y = \frac{1}{5}$, and $\lambda = -\frac{2}{5}$. Note that these are the same values of x and y that the substitution method (which we know is a sound method) gave us. This should strike you as being pretty remarkable.

Theorem 18.1 on p.416 of Simon and Blume gives the justification for this method, namely that if (x^*, y^*) is a solution to the constrained maximization problem, then there exists a real number λ^* such that (x^*, y^*, λ^*) is a critical point of L.

(Simon and Blume, p.317)

Definition 39 *If f is a function of x and y and f_x and f_y exist at, and are continuous at, every (x, y) in an open region R, then f is* ***continuously differentiable*** *(or C^1) on R.*

[1] Alternatively, we can add instead of subtract, as in $L(x, y, \lambda) = f(x, y) + \lambda[g(x, y) - c]$, as the solution to the constrained maximization problem will be the same either way. Thus, some authors add.

(Simon and Blume, p.416)

Theorem 36 *Let f and g be C^1 functions of x and y. Suppose that (x^*, y^*) is a solution of the problem: maximize $f(x, y)$ subject to $g(x, y) = c$. Suppose further than (x^*, y^*) is not a critical point of g. Then there is a real number λ^* such that (x^*, y^*, λ^*) is a critical point of the Lagrangian function $L(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$. That is, at (x^*, y^*, λ^*), $\frac{\partial L}{\partial x} = 0$, $\frac{\partial L}{\partial y} = 0$, and $\frac{\partial L}{\partial \lambda} = 0$ all three hold.*

We won't pursue this topic further, but the method can be expanded to handle functions of more than 2 variables, multiple equality constraints, and multiple inequality constraints (e.g., that $x \geq 0$ and $y \geq 0$) as well.

13 Logarithmic Functions

We know that $10^0 = 1$, $10^1 = 10$, $10^2 = 100$, and so on.

Also, $10^{-1} = \frac{1}{10} = 0.1$, $10^{-2} = \frac{1}{10^2} = 0.01$, and so on.

So the number 10 can be raised to any real number (not just integers) power, and the result is always a positive real number.

So for any positive real number, we can ask the power to which 10 needs to be raised in order to get that number.

This leads to the notion of the logarithm of base 10. For any positive real number x, define the function $f(x) = \log_{10}(x) =$ the number such that 10 raised to that number equals x

In other words, $\log_{10}(x) = a$ means that $10^a = x$.

Note that the base of the log function doesn't have to be 10, but 10 is a commonly-used based, in fact so common that sometimes authors in math just write $\log x$ to mean $\log_{10} x$

In empirical work in political science, we more often use the **natural logarithm**, whose base is the number $e \approx 2.718$. The reason for this is one of the great mysteries of the world.

For the natural logarithm, the standard notation is $\ln x$. That is, $\ln x = \log_e x$, and $\ln x = a$ means that $e^a = x$

For a *strictly positive* independent variable x that take on a very wide range of values (e.g., 1 to 1,000,000), the empirical analyst often includes $\ln x$ in a regression equation rather than x itself, in order to suppress this variation. This can be seen from the graph of $y = \ln x$

Sometimes for a highly-dispersed variable x that takes on values of 0 or even negative values, the empirical analyst will transform it to make it always positive (e.g., by adding a fixed positive constant to every observation of x), and then ln it

THEOREM 6.2

Theorem 37 *(Logarithmic properties) If a and b are positive numbers and n is a rational number, then the following properties are true.*

(1) $\ln(1) = 0$

(2) $\ln(ab) = \ln(a) + \ln(b)$

(3) $\ln(a^n) = n \ln a$

(4) $\ln(\frac{a}{b}) = \ln a - \ln b$

In **maximum likelihood estimation**, if L is the **likelihood function**, it is easier to find the maximizer of $\ln(L)$ (the **log-likelihood**) than of L, because L is the product of many terms, whereas by property (2) $\ln(L)$ is the sum of many terms, and it is easier to differentiate (and thereby find the maximizer of) a sum than a product. And because $y = \ln(x)$ is a strictly increasing function, L and $\ln(L)$ have the same maximizer

THEOREM 6.3

Theorem 38 *Let u be a differentiable function of x.*

(1) $\frac{d}{dx}[\ln x] = \frac{1}{x}$ *(provided $x > 0$)*

(2) $\frac{d}{dx}[\ln u] = \frac{1}{u}\frac{du}{dx}$ *(provided $u > 0$), from the Chain Rule*

14 Integration

14.1 Antiderivatives and Indefinite Integration

So far, given a function $f(x)$, find its derivative

Suppose instead that, given a function $f(x)$, we are asked to find a function whose derivative is $f(x)$, i.e., find a function $F(x)$ such that $F'(x) = f(x)$

For example, if $f(x) = 2$, we know that $F(x) = 2x$, since $F'(x) = 2 = f(x)$

Similarly, if $f(x) = 2x$, we know that $F(x) = x^2$

Definition 40 *A function F is called an **antiderivative** of the function f if for every x in the domain of f, $F'(x) = f(x)$.*

No function has a *unique* antiderivative, since if $F(x)$ is an antiderivative of $f(x)$, then so is $F(x) + C$ for any $C \in \mathbb{R}$.

THEOREM 5.1

Theorem 39 *If F is an antiderivative of f on an interval I, then G is an antiderivative of f on the interval I if and only if G is of the form $G(x) = F(x) + C$ for all x in I, where C is a constant.*

If we have a function $f(x)$, the symbol used to denote the **antidifferentiation** "process", i.e., the antidifferentiation "operator", is $\int f(x)\ dx$, which is read as "integrate the function $f(x)$ with respect to the variable x (this is what the dx denotes)"

x is called the **variable of integration** (similar to the notion of taking the derivative with respect to x)

To **integrate** a function means to find its antiderivative

Thus, $\int 2x\ dx = x^2 + C$

We add the constant to indicate that the equation is true for any value of C, and hence this is really a family of solutions (antiderivatives)

The term **indefinite integral** is a synonym for antiderivative

Definition 41 *(definition of integral notation for antiderivatives) The notation $\int f(x)\ dx = F(x) + C$, where C is an arbitrary constant, means that F is an antiderivative of f. That is, $F'(x) = f(x)$ for all x in the domain of f.*

Note that integration and differentiation are "inverse" operations, as seen from $\int F'(x)\ dx = F(x) + C$

That is, integrating the derivative of a function gives us the original function (plus an arbitrary constant)

Moreover, if $\int f(x)\ dx = F(x) + C$, i.e., $F(x)$ is an antiderivative of $f(x)$, then differentiating both sides of the equation gives $\frac{d}{dx}[\int f(x) dx] = f(x)$

That is, differentiating the integral of a function gives us back the function

the two processes "cancel out" each other, and hence they are "inverse" processes

THEOREM 5.2

Theorem 40 *The following integration rules hold, where C and k are any constants.*

(1) $\int 0 \, dx = C$

(2) $\int k \, dx = kx + C$

(3) $\int kf(x) \, dx = k \int f(x) \, dx$

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

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

(6) $\int x^n \, dx = \frac{x^{n+1}}{n+1} + C$ *(provided $n \neq -1$)*

These rules make it easy to integrate any polynomial function, and the power rule (6) is the key

14.2 Area, Riemann Sums, and the Definite Integral

Definition 42 *(definition of sigma notation) The sum of n terms $a_1, a_2, a_3, \ldots, a_n$ is written as $\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$, where i is called the* ***index of summation****, a_i is called the* ***i-th term*** *of the sum, and the* ***upper and lower bounds of summation*** *are n and 1, respectively.*

THEOREMS 5.3 and 5.4

Theorem 41 *The following rules hold.*

(1) $\sum_{i=1}^{n} ka_i = k \sum_{i=1}^{n} a_i$*, where k is a constant*

(2) $\sum_{i=1}^{n} [a_i + b_i] = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$

(3) $\sum_{i=1}^{n} [a_i - b_i] = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i$

(4) $\sum_{i=1}^{n} c = cn$*, where c is a constant*

For a common shape such as a rectangle, we say that the formula for the **area** of the rectangle is base times height, where we have some intuitive notion of what we mean by the area of a shape

It is more accurate to say that we *define* the area of a rectangle to be its base times its height

We can use that formula to derive the area of other simple shapes such as a triangle, by embedding them within rectangles and then being ingenious

What about the area under a never-negative curve $y = f(x)$ on the interval $[a, b]$, where by "under" we mean down to the x-axis?

We can approximate this area by drawing narrow rectangles approximating this region under the curve, and then add up the areas of those rectangles

As the width of these rectangles decreases, the approximation becomes more accurate

In the limit, as the width approaches 0, the approximation approaches perfection, and we define this limit to be the area under the curve

Definition 43 *(area of a region in the plane) Let f be continuous and non-negative on the interval $[a, b]$. The **area** of the region bounded by the graph of f, the x-axis, and the vertical lines $x = a$ and $x = b$ is:*

area $= \lim_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x$, *where n is the number of rectangles used in the approximation,* $\Delta x = \frac{b-a}{n}$ *is the width of each rectangle, and c_i is any value of x in the i'th rectangle.*

The approximation $\sum_{i=1}^{n} f(c_i)\Delta x$ is called a **Riemann sum**, although the latter in general does not require the width of each rectangle to be the same. It also does not require the function f to be non-negative, and we only need to impose this restriction when defining the notion of an *area* under a curve

More generally, we allow the function to be negative as well, and we allow rectangle i to have width Δx_i, and we then define a Riemann sum as $\sum_{i=1}^{n} f(c_i)\Delta x_i$

When we allow the rectangles to have different widths, we call the largest width the **norm** of the partition (the rectangle structure), and denote it by $\|\Delta\|$

Note that $\|\Delta\| \to 0$ implies that $n \to \infty$, but the converse is not true

Now that we no longer require f to be non-negative and the approximation rectangles to have equal widths, we define the notion of a definite integral, which is related to the idea of an area but is more general

Definition 44 *(the definite integral) If f is defined on the closed interval $[a, b]$ and the limit of a Riemann sum of f exists, then we say f is **integrable** on $[a, b]$ and we denote the limit by:*

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

*This limit is called the **definite integral** of f from a to b. The number a is the **lower limit** of integration, and the number b is the **upper limit** of integration.*

Although we use the same $\int$ notation for indefinite integrals and definite integrals, they are different entities. A definite integral is a number (because it is a limit of a certain function), whereas an indefinite integral is a family of functions. But the Fundamental Theorem of Calculus will tell us the link between these two and why we use the same notation

THEOREM 5.6

Theorem 42 *If a function f is continuous on the closed interval $[a, b]$, then f is integrable* *on* $[a, b]$.

THEOREM 5.7

Theorem 43 *(the definite integral as the area of a region) If* f *is continuous and non-negative on the closed interval* $[a, b]$*, then the area of the region bounded by the graph of* f*, the* x*-axis, and the vertical lines* $x = a$ *and* $x = b$ *is given by:*

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

Definition 45 *We make the following 2 notational definitions.*

(1) If f *is defined at* $x = a$*, then* $\int_a^a f(x)\ dx = 0$

(2) If f *is integrable on* $[a, b]$*, then* $\int_b^a f(x)\ dx = -\int_a^b f(x)\ dx$

THEOREM 5.8

Theorem 44 *If* f *is integrable on the three closed intervals determined by* a*,* b*, and* c*, then:*

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

Theorem 45 *(properties of definite integrals) If* f *and* g *are integrable on* $[a, b]$ *and* k *is a constant, then the following properties are true.*

(1) $\int_a^b kf(x)dx = k\int_a^b f(x)dx$

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

(3) $\int_a^b [f(x) - g(x)] = \int_a^b f(x)dx - \int_a^b g(x)dx$

14.3 The Fundamental Theorem of Calculus

The definite integral, which we generally think of as giving the area under a curve (but it is more general than that), requires us to compute a limit of a Riemann sum, which is tedious.

The FTC gives us a much more convenient way of computing a definite integral, while also telling us why we use the same $\int$ notation for definite and indefinite integrals, despite them being fundamentally different things

Theorem 46 *(Fundamental Theorem of Calculus) If a function f is continuous on the closed interval $[a, b]$, then $\int_a^b f(x)dx = F(b) - F(a)$, where F is any function such that $F'(x) = f(x)$ for all $x \in [a, b]$.*

In other words, to compute $\int_a^b f(x)dx$, we simply determine an anti-derivative (i.e., an indefinite integral) F of f, and then compute $F(b) - F(a)$. No need to calculate a limit of a Riemann sum

As an example, suppose we want to compute $\int_0^5 3x^2dx$. We know that $\int 3x^2dx = x^3 + C$ (where C is a constant). That is, let $F(x) = x^3 + C$. The FTC tells us that $\int_0^5 3x^2dx = F(5) - F(0) = (5^3 + C) - (0^3 + C) = 125$. (Note that the constant C always drops out, and so we could have just let $F(x) = x^3$. The constant is only important for indefinite integration, not definite integration, and hence we normally leave it out. Notice how the FTC is worded near the end.)

We often use the notation $\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a)$

The definite integral $\int_a^b f(x)dx$ is a number. Sometimes we are interested in treating the upper limit of integration (b) as a variable x instead of a constant b, and taking the derivative of the definite integral (which is now a function of that variable) with respect to that variable. This leads to the second FTC.

THEOREM 5.13

Theorem 47 *(Second Fundamental Theorem of Calculus) If f is continuous on an open* *interval I containing a, then for every x in the interval,* $\frac{d}{dx}[\int_a^x f(t)dt] = f(x)$.

Note that in the second FTC we use the notation $f(t)$ in the definite integral, rather than $f(x)$. That's because x is now being used to denote the upper limit of integration, and hence we want to use a different label for the variable of integration. And remember that the definite integral is *not* a function of its variable of integration; in fact, any variable can be used (e.g., t instead of x), and the definite integral is exactly the same (it just depends on f, b, and a). All of this may seem a bit confusing and arcane, and the second FTC is used less often than the more straightforward first FTC, but the second is sometimes used in applied work, so it's useful to know that it exists. You can study it in detail if you ever need to take the derivative of a definite integral (where the upper limit of integration is a variable).

15 Calculus and Probability Theory

In the section of math camp on probability theory, you will learn about **continuous random variables**, which are more complicated than **discrete random variables**. We just briefly discuss this topic here, in order to give you a sense of how calculus plays a role in probability theory.

In a **discrete probability distribution**, we assign an exact probability to each of the (finite or countably infinite) possible values that the discrete random variable (that is drawn from the discrete probability distribution) can take.

The simplest discrete probability distribution is the single Bernoulli trial (also called a binomial trial): the random variable takes the value of 1 ("success") with probability $q \in (0, 1)$, and the value of 0 ("failure") with probability $1 - q$

If Y is a discrete random variable with the **probability function** $p(y)$, then the **expected value** of Y, $E(Y)$, is defined to be $E(Y) = \sum_y y \cdot p(y)$

So for the above Bernoulli trial, $E(Y) = q$

Continuous random variables, whose value is determined by a draw from a **continuous probability distribution**, can take on an (uncountably) infinite number of possible values

Two simple continuous probability distributions: the uniform distribution, and the normal distribution

We characterize continuous probability distributions not by the probability they assign to each possible value (as for discrete distributions), i.e., by a probability function $p(y)$, but by a **cumulative distribution function** $F(y)$, defined by $F(y) = \Pr(Y \leq y)$ for any $y \in (-\infty, \infty)$ (the CDF also applies to discrete probability distributions, but is an optional tool there, but crucial to continuous probability distributions)

If $F(y)$ is differentiable on an interval I, then we define the **probability density function** $f(y)$ to be the derivative of the cumulative distribution function, i.e., $f(y) = F'(y)$ for all y in I

We associate common continuous probability distributions with the graph of their density functions, e.g., draw the graph of $f(y)$ for the normal and uniform distributions

$F(y) = \Pr(Y \leq y) = \int_{-\infty}^{y} f(t)dt$

If $f(y)$ is a density function, then (1) $f(y) \geq 0$ for any value of y, and (2) $\int_{-\infty}^{\infty} f(y)dy = 1$

These are the continuous equivalents of the discrete rules that (1) $0 \leq p(y) \leq 1$ for any value of y, and (2) $\sum_y p(y) = 1$

Also, $\Pr(a \leq Y \leq b) = \int_a^b f(y)dy = F(b) - F(a)$ (the latter equality follows from the first FTC)

The expected value of a continuous random variable Y is $E(Y) = \int_{-\infty}^{\infty} yf(y)dy$ provided that the integral exists.

Compare this definition to that of a discrete random variable. Basically, the density function f for a continuous random variable is the "equivalent" of the probability function p for a discrete random variable, although $f(y)$ does *not* give the probability that $Y = y$, which is technically 0 for any value of y

16 Homework Problems

1. Consider the sets $A = \{a, b, c, d\}$, $B = \{b, d, e\}$, and $C = \{c, x, y\}$. Answer the following questions, in each case listing the elements in alphabetical order.

   (a) What is $A \cap C$?

   (b) What is $C \cup B$?

   (c) What is $B \setminus A$?

   (d) List every subset of B.

2. Find the set of real numbers for which the inequality $2 - 5x \leq 12$ holds. That is, solve for x.

3. Find the set of real numbers for which the following inequality holds. (Recall that $| a |$ denotes the absolute value of a.)

   (a) $| x - 3 | \leq 2$

   (b) $3 < | x + 2 |$

4. Find the distance between the points $(-2, 1)$ and $(3, 4)$ in the Cartesian plane. (Recall that the distance formula is $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.)

5. Sketch the graph of the linear equation $y = 2x + 1$. (A neat hand-drawn figure is fine.)

6. Write the equation, in the form $y = mx + c$, of the straight line passing through the points $(1, 5)$ and $(-3, 0)$. (That is, find the values of m and c.)

7. (a bit more challenging; try your best) Use the ϵ-δ definition of a limit (definition 5 in the notes above) to prove that $\lim_{x \to 2}(3x - 2) = 4$.

8. Find the following limit, using whatever method you want. $\lim_{x \to 2}(4x^2 + 3)$

9. The same for the following limit. $\lim_{x \to 1} \frac{x^2+x+2}{x+1}$

10. The same for the following limit. $\lim_{x \to 0}(1 + \frac{1}{x^2})$

11. (looks difficult, but is not really; try your best) Use the Intermediate Value Theorem (Theorem 10 in the above notes) to show that the polynomial function $f(x) = x^3+2x-1$ has a 0 (i.e., a point where $f(x) = 0$) in the interval $[0, 1]$.

12. (looks difficult, but is not really; try your best) Use definition 14 (in the above notes) to find the derivative $f'(x)$ of the function $f(x) = 3x + 2$. (That is, your solution should actually use definition 14, and should not just apply some differentiation rule.)

13. Find the derivative $f'(x)$ of each of the following functions, using whichever differentiation rule you want to (unless otherwise indicated).

    (a) $f(x) = 2x^3 - x^2 + 3x$. Also find $f''(x)$. Find the value of $f'(3)$ and $f''(3)$.

    (b) $f(x) = x^2 - \frac{4}{x}$. Also find $f''(x)$. Find the value of $f'(1)$ and $f''(1)$.

    (c) $f(x) = (x^2 - 2x + 1)(x^3 - 1)$ (use the product rule)

    (d) $f(x) = \frac{3x-2}{2x-3}$ (use the quotient rule)

    (e) $f(x) = (9 - x^2)^{2/3}$ (use the chain rule)

14. (looks difficult, but is not really; try your best) Use implicit differentiation to find $\frac{dy}{dx}$ for the following equation. $x^2 - y^2 = 16$.

15. For the following function, find the intervals of x on which the function is strictly increasing and strictly decreasing. Use the first and second order conditions to find

the relative maxima and relative minima of the function, and evaluate the value of the function at these points. Based on your answers, provide a graph showing what this function looks like (a neat hand-drawn figure will be fine). $f(x) = -2x^2 + 4x + 3$

16. Calculate the following indefinite integral. $\int (x^3 + 2)dx$

17. Use the Fundamental Theorem of Calculus (Theorem 46 in the above notes) to calculate the following definite integral. $\int_1^2 (5x^4 + 5)dx$

18. Use L'Hopital's Rule (Theorem 31 in the above notes) to evaluate the following limit. $\lim_{x\to\infty} \frac{3x^2-2x+1}{2x^2+3}$

19. For the function $f(x, y) = x^2 - 3xy + y^2$, find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, $\frac{\partial^2 f}{\partial y \partial x}$, and $\frac{\partial^2 f}{\partial x \partial y}$. Find the (x, y) points (if any) at which $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$ both hold (this may involve solving 2 equations in 2 unknowns). For each such point (if any), calculate d in the second-partials test (Theorem 35 in the above notes), and apply the test to determine whether each of those points is a relative minimum, a relative maximum, a saddle point, or whether the test gives no information as to this.

20. In any given speech, candidate Trump can choose to speak x words, where $x \in [0, \infty)$. Each word spoken increases his support by 120 units (where "unit" is a deliberately vague concept on my part). That is, $s(x) = 120x$. However, each word spoken makes him increasingly hoarse, and at an exponential rate: $h(x) = 3x^2$. His utility function is his support minus his hoarseness: $U(x) = s(x) - h(x)$.

    (a) Find the optimal number of words he should speak at any given speech, i.e., find the maximizer of $U(x)$.

(b) Now suppose that his speech cannot exceed 15 words, i.e., $x \in [0, 15]$. Find the solution to this constrained maximization problem, i.e., find the maximizer on this range (using any method you want).

21. Consider the function $z = f(x, y) = -3x^2 + 2y$. Find the maximum of this function subject to the equality constraint $3x - y = 10$:

(a) by direct substitution.

(b) using the Method of Lagrange Multipliers.