University of Chicago Political Science Math Prefresher

Anton Strezhnev

9/12/2022

Table of contents

1 Overview

1.1 Introduction

The 2022 UChicago Math Prefresher for incoming Political Science graduate students will be held from September 12-14; September 19-21 and September 23rd. The course is designed as a brief review of math fundamentals – calculus, optimization, probability theory and linear algebra among other topics – as well as an introduction to programming in the R statistical computing language. The course is entirely optional and there are no grades or assignments but we encourage all incoming graduate students to attend if they are able.

1.2 Course Booklet

The course notes for the math and programming sections as well as all practice problems are available on this website and can be accessed by navigating the menus in the sidebar.

1.3 Schedule

The prefresher will run for a total of seven days September 12-14, September 19-21 and September 23rd, with breaks for the APSA conference and the new student orientation. Each day will run from around 9am to 4pm with many breaks in between. We will be meeting in room 407 of Pick Hall.

The morning will focus on math instruction. We will have two one hour sessions from 9:30am - 10:30am and 10:45am-11:45am, with a ~15 minute break in between. These sessions will involve a combination of lectures and working through practice problems.

We will break for lunch from 12:00pm-1:00pm. On September 13th and Spetember 19th, we will have a catered lunch with a faculty member guest. Otherwise, you are free to explore the campus for various lunch options.

The afternoon will focus on coding instruction with lecture/demonstration from 1:30pm-2:45pm. After a short break you will work together on a variety of coding exercises from 3:00-3:30pm. In the last 30 minutes we will regroup to wrap up and discuss any questions on the material.

1.4 Software

As the afternoons of the prefresher will involve instruction in coding, you should be sure to bring a laptop and a charging cable. In addition, prior to the start of the prefresher, please make sure to have installed the following on your computer:

- R (version 4.2.1 or higher)
- RStudio Desktop Open Source License (this is the primary IDE or integrated development environment in which we will be working)
- LaTeX: This is primarily to allow you to generate PDF documents using RMarkdown. We will use the TinyTeX LaTeX distribution which is designed to be minimalist and tailored specifically for R users. After installing R and RStudio, open up an instance of R, install the 'tinytex' package and run the `install_tinytex()` command

```
install.packages('tinytex')
tinytex::install_tinytex()
```

We will also spend some time discussing document preparation and typesetting using LaTeX and Markdown. For the former, we will be using the popular cloud platform Overleaf, which allows for collaborative document editing and streamlines a lot of the irritating parts of typesetting in LaTeX. You should register for an account using your university e-mail as all University of Chicago students and faculty have access to an Overleaf Pro account for free.

You are also welcome to install a LaTeX editor on your local machine to work alongside the TinyTeX distribution or any other TeX distribution that you prefer such as TexMaker

1.5 Acknowledgments

This prefresher draws heavily on the wonderful materials that have been developed by over 20 years of instructors at the Harvard Government Math Prefresher that have been so generously distributed under the GPL 3.0 License. Special thanks to Shiro Kuriwaki, Yon Soo Park, and Connor Jerzak for their efforts in converting the original prefresher materials into the easily distributed Markdown format.

2 Sets, Operations, and Functions

2.1 Sets

Sets are the fundamental building blocks of mathematics. Events are not inherently numerical: the onset of war or the stock market crashing is not inherently a number. Sets can define such events, and we wrap math around so that we have a transparent language to communicate about those events. Combining sets with operations, relations, metrics, measures, etc… allows us to define useful mathematical structures. For example, the set of *real numbers* ($\mathbb{R}$) has a notion of *order* as well as defined *operations* of addition and multiplication.

Set : A set is any well defined collection of elements. If x is an element of S, $x \in S$.

Examples:

1. The set of choices available to a player in Rock-Paper-Scissors $\{\text{Rock}, \text{Paper}, \text{Scissors}\}$
2. The set of possible outcomes of a roll of a six-sided die $\{1, 2, 3, 4, 5, 6\}$
3. The set of all natural numbers $\mathbb{N}$
4. The set of all real numbers $\mathbb{R}$

Common mathematical notation relevant to sets:

- $\in$ = "is an element of"; $\notin$ = "is not an element of"
- $\forall$ = "for all" (univeral quantifier)
- $\exists$ = "there exists" (existential quantifier)
- $:$ = "such that"

Subset: If every element of set A is also in set B, then A is a *subset* of B. $A \subseteq B$. If, in addition to being a subset of B, A is not equal to B, A is a *proper subset* $A \subset B$.

Empty Set: a set with no elements. $S = \{\}$. It is denoted by the symbol $\emptyset$.

Cardinality: The cardinality of a set S, typically written $|S|$ is the number of members of S.

Many sets are infinite. For example, $\mathbb{N}$ the set of natural numbers $\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$ - Sets with cardinality less than $|\mathbb{N}|$ are *countable* - Sets with the same cardinality as $mathbbN|$ are *countably infinite* - Sets with greater cardinality than $|\mathbb{N}|$ are *uncountably infinite* (e.g. the real numbers).

Set operations:

1. **Union**: The union of two sets A and B, $A \cup B$, is the set containing all of the elements in A or B. $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$
2. **Intersection**: The intersection of sets A and B, $A \cap B$, is the set containing all of the elements in both A and B. $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$
3. **Complement**: If set A is a subset of S, then the complement of A, denoted A^C, is the set containing all of the elements in S that are not in A.

Properties of set operations:

- **Commutative**: $A \cup B = B \cup A$; $A \cap B = B \cap A$
- **Associative**: $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$
- **Distributive**: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- **de Morgan's laws**: $(A \cup B)^C = A^C \cap B^C$; $(A \cap B)^C = A^C \cup B^C$
- **Disjointness**: Sets are disjoint when they do not intersect, such that $A \cap B = \emptyset$. A collection of sets is pairwise disjoint (**mutually exclusive**) if, for all $i \neq j$, $A_i \cap A_j = \emptyset$. A collection of sets form a partition of set S if they are pairwise disjoint and they cover set S, such that $\bigcup_{i=1}^{k} A_i = S$.

Example 2.1.

Sets

Let set A be $\{1, 2, 3, 4\}$, B be $\{3, 4, 5, 6\}$, and C be $\{5, 6, 7, 8\}$. Sets A, B, and C are all subsets of the S which is $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

Write out the following sets:

1. $A \cup B$
2. $C \cap B$
3. B^c
4. $A \cap (B \cup C)$

Exercise 2.1.

Sets

Suppose you had a pair of four-sided dice. You sum the results from a single toss.

What is the set of possible outcomes?

Consider subsets $A = \{2, 8\}$ and $B = \{2, 3, 7\}$ of the sample space you found. What is

1. A^c
2. $(A \cup B)^c$

2.2 Metric spaces

A *metric space* is a set that has a notion of *distance* - called a "metric" - defined between any two elements (sometimes referred to as "points").

The distance function $d(x, y)$ defines the distance between element x and element y

- The real numbers $\mathbb{R}$ have a single distance function: $d(x, y) = |x - y|$
- In higher-dimensional real space (e.g. $\mathbb{R}^2$), we can define multiple distance metrics between $x = (x_1, x_2)$ and $y = (y_1, y_2)$
  - "Euclidean" distance: $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$
  - "Taxicab" distance: $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$
  - Chebyshev distance: $d(x, y) = \max\{|x_1 - y_1| + |x_2 - y_2|\}$
- All of these generalize to $\mathbb{R}^n$

A metric is a function that satisfies the following axioms

1. A distance between a point and itself is zero $d(x, x) = 0$
2. The distance between two points is strictly positive $d(x, y) > 0 \forall x \neq y$
3. Distance from x to y is the same as the distance from y to x ($d(x, y) = d(y, x)$)
4. The "triangle inequality" holds: $d(x, z) \leq d(x, y) + d(y, z)$

Once we have a metric space, we can define some additional useful concepts

Ball: A ball of radius r centered at x_0 is a set that contains all points with a distance less than r from x_0.

Sphere: A sphere of radius r centered at x_0 is the set that contains all points with a distance exactly r from x_0.

Interior Point: The point x is an interior point of the set S if x is in S and if there is some ϵ-ball around x that contains only points in S. The **interior** of S is the collection of all interior points in S. The interior can also be defined as the union of all open sets in S.

- If the set S is circular, the interior points are everything inside of the circle, but not on the circle's rim.
- Example: The interior of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 < 4\}$.

Boundary Point: The point $\mathbf{x}$ is a boundary point of the set S if every ϵ-ball around $\mathbf{x}$ contains both points that are in S and points that are outside S. The **boundary** is the collection of all boundary points.

- If the set S is circular, the boundary points are everything on the circle's rim.
- Example: The boundary of $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 = 4\}$.

Open: A set S is open if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.

- If the set S is circular and open, the points contained within the set get infinitely close to the circle's rim, but do not touch it.
- Example: $\{(x, y) : x^2 + y^2 < 4\}$

Closed: A set S is closed if it contains all of its boundary points.

- Alternatively: A set is closed if its complement is open.
- If the set S is circular and closed, the set contains all points within the rim as well as the rim itself.
- Example: $\{(x, y) : x^2 + y^2 \leq 4\}$
- Note: a set may be neither open nor closed. Example: $\{(x, y) : 2 < x^2 + y^2 \leq 4\}$

2.3 Operators; Sum and Product notation

Addition (+), Subtraction (-), multiplication and division are basic operations of arithmetic. In statistics or calculus, we will often want to add a *sequence* of numbers that can be expressed as a pattern without needing to write down all its components. For example, how would we express the sum of all numbers from 1 to 100 without writing a hundred numbers?

For this we use the summation operator $\sum$ and the product operator $\prod$.

Summation:

$$\sum_{i=1}^{100} x_i = x_1 + x_2 + x_3 + \cdots + x_{100}$$

The bottom of the $\sum$ symbol indicates an index (here, i), and its start value 1. At the top is where the index ends. The notion of "addition" is part of the $\sum$ symbol. The content to the right of the summation is the meat of what we add. While you can pick your favorite index, start, and end values, the content must also have the index.

A few important features of sums:

- $\sum\limits_{i=1}^{n} cx_i = c\sum\limits_{i=1}^{n} x_i$
- $\sum\limits_{i=1}^{n}(x_i + y_i) = \sum\limits_{i=1}^{n} x_i + \sum\limits_{i=1}^{n} y_i$
- $\sum\limits_{i=1}^{n} c = nc$

Product:

$$\prod_{i=1}^{n} x_i = x_1 x_2 x_3 \cdots x_n$$

Properties:

- $\prod\limits_{i=1}^{n} cx_i = c^n \prod\limits_{i=1}^{n} x_i$
- $\prod\limits_{i=k}^{n} cx_i = c^{n-k+1} \prod\limits_{i=k}^{n} x_i$
- $\prod\limits_{i=1}^{n}(x_i + y_i) = \text{a total mess}$
- $\prod\limits_{i=1}^{n} c = c^n$

Other Useful Operations

Factorials!:

$$x! = x \cdot (x-1) \cdot (x-2) \cdots (1)$$

Modulo: Tells you the remainder when you divide the first number by the second.

- $17 \mod 3 = 2$
- $100 \% 30 = 10$

Example 2.2.

Operators

1. $\sum_{i=1}^{5} i =$
2. $\prod_{i=1}^{5} i =$
3. $14 \mod 4 =$
4. $4! =$

Exercise 2.2.

Operators

Let $x_1 = 4, x_2 = 3, x_3 = 7, x_4 = 11, x_5 = 2$

1. $\sum\limits_{i=1}^{3} (7)x_i$
2. $\sum\limits_{i=1}^{5} 2$
3. $\prod\limits_{i=3}^{5} (2)x_i$

2.4 Introduction to Functions

A **function** is a mapping, or transformation, that relates members of one set to members of another set. For instance, if you have two sets: set A and set B, a function from A to B maps every value a in set A such that $f(a) \in B$. Functions can be "many-to-one", where many values or combinations of values from set A produce a single output in set B, or they can be "one-to-one", where each value in set A corresponds to a single value in set B. A function by definition has a single function value for each element of its domain. This means, there cannot be "one-to-many" mapping.

Dimensionality: $\mathbf{R}^1$ is the set of all real numbers extending from $-\infty$ to $+\infty$ — i.e., the real number line. $\mathbf{R}^n$ is an n-dimensional space, where each of the n axes extends from $-\infty$ to $+\infty$.

- $\mathbf{R}^1$ is a one dimensional line.
- $\mathbf{R}^2$ is a two dimensional plane.
- $\mathbf{R}^3$ is a three dimensional space.

Points in $\mathbf{R}^n$ are ordered n-tuples (just means an combination of n elements where order matters), where each element of the n-tuple represents the coordinate along that dimension.

For example:

- $\mathbf{R}^1$: (3)
- $\mathbf{R}^2$: (-15, 5)

- $\mathbf{R}^3$: (86, 4, 0)

Examples of mapping notation:

Function of one variable: $f : \mathbf{R}^1 \to \mathbf{R}^1$

- $f(x) = x + 1$. For each x in $\mathbf{R}^1$, $f(x)$ assigns the number $x + 1$.

Function of two variables: $f : \mathbf{R}^2 \to \mathbf{R}^1$.

- $f(x, y) = x^2 + y^2$. For each ordered pair (x, y) in $\mathbf{R}^2$, $f(x, y)$ assigns the number $x^2 + y^2$.

We often use variable x as input and another y as output, e.g. $y = x + 1$

Example 2.3.

Functions

For each of the following, state whether they are one-to-one or many-to-one functions.

1. For $x \in [0, \infty]$, $f : x \to x^2$ (this could also be written as $f(x) = x^2$).
2. For $x \in [-\infty, \infty]$, $f : x \to x^2$.

Exercise 2.3.

Functions

For each of the following, state whether they are one-to-one or many-to-one functions.

1. For $x \in [-3, \infty]$, $f : x \to x^2$.
2. For $x \in [0, \infty]$, $f : x \to \sqrt{x}$

Some functions are defined only on proper subsets of $\mathbf{R}^n$.

- **Domain**: the set of numbers in X at which $f(x)$ is defined.
- **Range**: elements of Y assigned by $f(x)$ to elements of X, or $f(X) = \{y : y = f(x), x \in X\}$ Most often used when talking about a function $f : \mathbf{R}^1 \to \mathbf{R}^1$.
- **Image**: same as range, but more often used when talking about a function $f : \mathbf{R}^n \to \mathbf{R}^1$.

Some General Types of Functions

Monomials: $f(x) = ax^k$

a is the coefficient. k is the degree.

Examples: $y = x^2$, $y = -\frac{1}{2}x^3$

Polynomials: sum of monomials.

Examples: $y = -\frac{1}{2}x^3 + x^2$, $y = 3x + 5$

The degree of a polynomial is the highest degree of its monomial terms. Also, it's often a good idea to write polynomials with terms in decreasing degree.

2.5 Logarithms and Exponents

Exponential Functions: Example: $y = 2^x$

Relationship of logarithmic and exponential functions:

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

The log function can be thought of as an inverse for exponential functions. a is referred to as the "base" of the logarithm.

Common Bases: The two most common logarithms are base 10 and base e.

1. Base 10: $y = \log_{10}(x) \iff 10^y = x$. The base 10 logarithm is often simply written as "$\log(x)$" with no base denoted.
2. Base e: $y = \log_e(x) \iff e^y = x$. The base e logarithm is referred to as the "natural" logarithm and is written as "$\ln(x)$".

Properties of exponential functions:

- $a^x a^y = a^{x+y}$
- $a^{-x} = 1/a^x$
- $a^x/a^y = a^{x-y}$
- $(a^x)^y = a^{xy}$
- $a^0 = 1$

Properties of logarithmic functions (any base):

Generally, when statisticians or social scientists write $\log(x)$ they mean $\log_e(x)$. In other words: $\log_e(x) \equiv \ln(x) \equiv \log(x)$

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

and

$$a^{\log_a(x)} = x$$

- $\log(xy) = \log(x) + \log(y)$
- $\log(x^y) = y\log(x)$
- $\log(1/x) = \log(x^{-1}) = -\log(x)$
- $\log(x/y) = \log(x \cdot y^{-1}) = \log(x) + \log(y^{-1}) = \log(x) - \log(y)$
- $\log(1) = \log(e^0) = 0$

Change of Base Formula: Use the change of base formula to switch bases as necessary:

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

Example:

$$\log_{10}(x) = \frac{\ln(x)}{\ln(10)}$$

You can use logs to go between sum and product notation. This will be particularly important when you're learning how to optimize likelihood functions.

$$
\begin{aligned}
\log\left(\prod_{i=1}^{n} x_i\right) &= \log(x_1 \cdot x_2 \cdot x_3 \cdots x_n) \\
&= \log(x_1) + \log(x_2) + \log(x_3) + \cdots + \log(x_n) \\
&= \sum_{i=1}^{n} \log(x_i)
\end{aligned}
$$

Therefore, you can see that the log of a product is equal to the sum of the logs. We can write this more generally by adding in a constant, c:

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

Example 2.4.

Logarithms

Evaluate each of the following logarithms

1. $\log_4(16)$
2. $\log_2(16)$

Simplify the following logarithm. By "simplify", we actually really mean - use as many of the logarithmic properties as you can.

3. $\log_4(x^3y^5)$

Exercise 2.4. Evaluate each of the following logarithms

1. $\log_{\frac{3}{2}}(\frac{27}{8})$

Simplify each of the following logarithms. By "simplify", we actually really mean - use as many of the logarithmic properties as you can.

2. $\log(\frac{x^9y^5}{z^3})$
3. $\ln\sqrt{xy}$

2.6 Graphing Functions

What can a graph tell you about a function?

- Is the function increasing or decreasing? Over what part of the domain?
- How "fast" does it increase or decrease?
- Are there global or local maxima and minima? Where?
- Are there inflection points?
- Is the function continuous?
- Is the function differentiable?
- Does the function tend to some limit?
- Other questions related to the substance of the problem at hand.

2.7 Solving for Variables and Finding Roots

Sometimes we're given a function $y = f(x)$ and we want to find how x varies as a function of y. Use algebra to move x to the left hand side (LHS) of the equation and so that the right hand side (RHS) is only a function of y.

Example 2.5.

Solving

Solve for x:

1. $y = 3x + 2$
2. $y = e^x$

Solving for variables is especially important when we want to find the **roots** of an equation: those values of variables that cause an equation to equal zero. Especially important in finding equilibria and in doing maximum likelihood estimation.

Procedure: Given $y = f(x)$, set $f(x) = 0$. Solve for x.

Multiple Roots:

$$f(x) = x^2 - 9 \quad \Longrightarrow \quad 0 = x^2 - 9 \quad \Longrightarrow \quad 9 = x^2 \quad \Longrightarrow \quad \pm\sqrt{9} = \sqrt{x^2} \quad \Longrightarrow \quad \pm 3 = x$$

Quadratic Formula: For quadratic equations $ax^2 + bx + c = 0$, use the quadratic formula:

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

Exercise 2.5.

Roots

Solve for x:

1. $f(x) = 3x + 2 = 0$
2. $f(x) = x^2 + 3x - 4 = 0$
3. $f(x) = e^{-x} - 10 = 0$

3 Limits

Solving limits, i.e. finding out the value of functions as its input moves closer to some value, is important for the social scientist's mathematical toolkit for two related tasks. The first is for the study of calculus, which will be in turn useful to show where certain functions are maximized or minimized. The second is for the study of statistical inference, which is the study of inferring things about things you cannot see by using things you can see.

Example: The Central Limit Theorem

Perhaps the most important theorem in statistics is the Central Limit Theorem,

Theorem 3.1 (Central Limit Theorem)**.** *For any series of independent and identically distributed random variables $X_1, X_2, \cdots$, we know the distribution of its sum even if we do not know the distribution of X. The distribution of the sum is a Normal distribution.*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Normal(0,1)$$

where μ is the mean of X and σ is the standard deviation of X. The arrow is read as "converges in distribution to". $Normal(0,1)$ indicates a Normal Distribution with mean 0 and variance 1.

That is, the limit of the distribution of the lefthand side is the distribution of the righthand side.

The sign of a limit is the arrow "$\rightarrow$". Although we have not yet covered probability so we have not described what distributions and random variables are, it is worth foreshadowing the Central Limit Theorem. The Central Limit Theorem is powerful because it gives us a *guarantee* of what would happen if $n \rightarrow \infty$, which in this case means we collected more data.

Example: The Law of Large Numbers

A finding that perhaps rivals the Central Limit Theorem is the (Weak) Law of Large Numbers:

Theorem 3.2 ((Weak) Law of Large Numbers)**.** *For any draw of identically distributed independent variables with mean μ, the sample average after n draws, $\bar{X}_n$, converges in probability to the true mean as $n \to \infty$:*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

A shorthand of which is $\bar{X}_n \xrightarrow{p} \mu$, where the arrow is read as "converges in probability to".

Intuitively, the more data, the more accurate is your guess. For example, Figure 3.1 shows how the sample average from many coin tosses converges to the true value : 0.5.

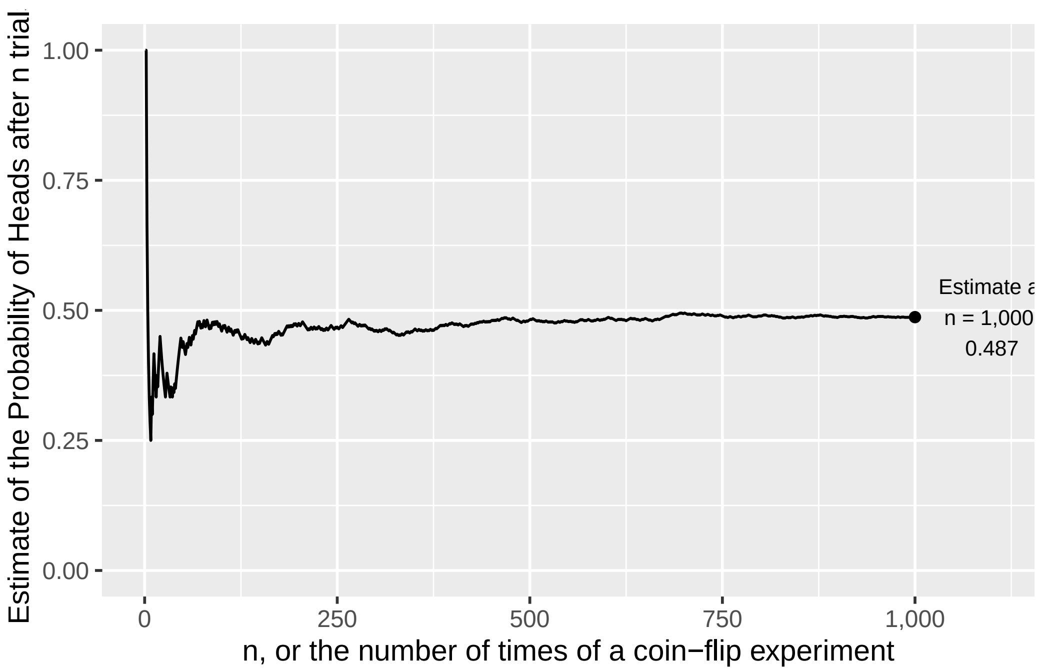

Figure 3.1: As the number of coin tosses goes to infinity, the average probabiity of heads converges to 0.5

3.1 Sequences

We need a couple of steps until we get to limit theorems in probability. First we will introduce a "sequence", then we will think about the limit of a sequence, then we will think about the limit of a *function*.

A **sequence** $\{x_n\} = \{x_1, x_2, x_3, \ldots, x_n\}$ is an ordered set of real numbers, where x_1 is the first term in the sequence and y_n is the nth term. Generally, a sequence is infinite, that is it extends to $n = \infty$. We can also write the sequence as $\{x_n\}_{n=1}^{\infty}$

where the subscript and superscript are read together as "from 1 to infinity."

Example 3.1.

Sequences

How do these sequences behave?

1. $\{A_n\} = \left\{2 - \frac{1}{n^2}\right\}$
2. $\{B_n\} = \left\{\frac{n^2+1}{n}\right\}$
3. $\{C_n\} = \left\{(-1)^n \left(1 - \frac{1}{n}\right)\right\}$

We find the sequence by simply "plugging in" the integers into each n. The important thing is to get a sense of how these numbers are going to change.

Graphing helps you make this point more clearly. See the sequence of $n = 1, ...20$ for each of the three examples in Figure 3.2.

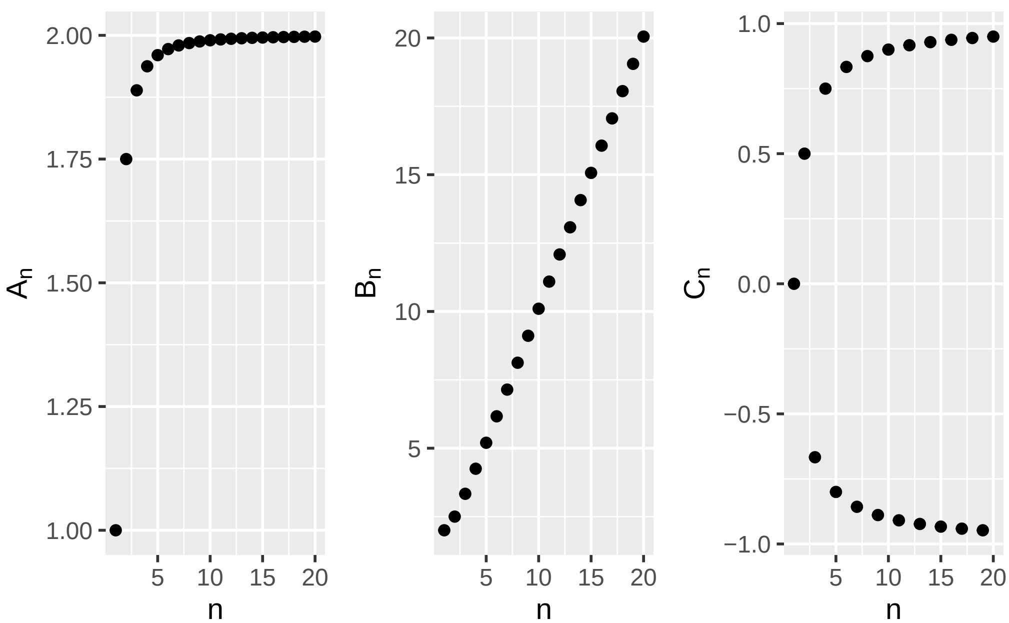

Figure 3.2: Behavior of Some Sequences

3.2 The Limit of a Sequence

The notion of "converging to a limit" is the behavior of the points in Example -@#exm-seqbehav. In some sense, that's the counterfactual we want to know. What happens as $n \to \infty$?

1. Sequences like 1 above that converge to a limit.
2. Sequences like 2 above that increase without bound.
3. Sequences like 3 above that neither converge nor increase without bound — alternating over the number line.

Definition: Limit The sequence $\{y_n\}$ has the limit L, which we write as $\lim\limits_{n\to\infty} y_n = L$, if for any $\epsilon > 0$ there is an integer N (which depends on ϵ) with the property that $|y_n - L| < \epsilon$ for each $n > N$. $\{y_n\}$ is said to converge to L. If the above does not hold, then $\{y_n\}$ diverges.

We can also express the behavior of a sequence as bounded or not:

1. Bounded: if $|y_n| \leq K$ for all n
2. Monotonically Increasing: $y_{n+1} > y_n$ for all n
3. Monotonically Decreasing: $y_{n+1} < y_n$ for all n

A limit is *unique*: If $\{y_n\}$ converges, then the limit L is unique.

If a sequence converges, then the sum of such sequences also converges. Let $\lim\limits_{n\to\infty} y_n = y$ and $\lim\limits_{n\to\infty} z_n = z$. Then

1. $\lim\limits_{n\to\infty} [ky_n + \ell z_n] = ky + \ell z$
2. $\lim\limits_{n\to\infty} y_n z_n = yz$
3. $\lim\limits_{n\to\infty} \frac{y_n}{z_n} = \frac{y}{z}$, provided $z \neq 0$

This looks reasonable enough. The harder question, obviously is when the parts of the fraction *don't* converge. If $\lim_{n\to\infty} y_n = \infty$ and $\lim_{n\to\infty} z_n = \infty$, What is $\lim_{n\to\infty} y_n - z_n$? What is $\lim_{n\to\infty} \frac{y_n}{z_n}$?

It is nice for a sequence to converge in limit. We want to know if complex-looking sequences converge or not. The name of the game here is to break that complex sequence up into sums of simple fractions where n only appears in the denominator: $\frac{1}{n}$, $\frac{1}{n^2}$, and so on. Each of these will converge to 0, because the denominator gets larger and larger. Then, because of the properties above, we can then find the final sequence.

Example 3.2.

Ratios

Find the limit of $\lim_{n\to\infty} \frac{n+3}{n}$

Solution. At first glance, $n+3$ and n both grow to ∞, so it looks like we need to divide infinity by infinity. However, we can express this fraction as a sum, then the limits apply separately:

$$\lim_{n\to\infty} \frac{n+3}{n} = \lim_{n\to\infty} \left(1 + \frac{3}{n}\right) = \underbrace{\lim_{n\to\infty} 1}_{1} + \underbrace{\lim_{n\to\infty} \left(\frac{3}{n}\right)}_{0}$$

so, the limit is actually 1.

After some practice, the key to intuition is whether one part of the fraction grows "faster" than another. If the denominator grows faster to infinity than the numerator, then the fraction will converge to 0, even if the numerator will also increase to infinity. In a sense, limits show how not all infinities are the same.

Exercise 3.1.

Limits

Find the following limits of sequences, then explain in English the intuition for why that is the case.

1. $\lim_{n\to\infty} \frac{2n}{n^2+1}$
2. $\lim_{n\to\infty} (n^3 - 100n^2)$

3.3 Limits of a Function

We've now covered functions and just covered limits of sequences, so now is the time to combine the two.

A function f is a compact representation of some behavior we care about. Like for sequences, we often want to know if $f(x)$ approaches some number L as its independent variable x moves to some number c (which is usually 0 or $\pm\infty$). If it does, we say that the limit of $f(x)$, as x approaches c, is L: $\lim_{x\to c} f(x) = L$. Unlike a sequence, x is a continuous number, and we can move in decreasing order as well as increasing.

For a limit L to exist, the function $f(x)$ must approach L from both the left (increasing) and the right (decreasing).

Definition 3.1.

Limits of a function

Let $f(x)$ be defined at each point in some open interval containing the point c. Then L equals $\lim_{x \to c} f(x)$ if for any (small positive) number ϵ, there exists a corresponding number $\delta > 0$ such that if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

A neat, if subtle result is that $f(x)$ does not necessarily have to be defined at c for $\lim_{x \to c}$ to exist.

Properties: Let f and g be functions with $\lim_{x \to c} f(x) = k$ and $\lim_{x \to c} g(x) = \ell$.

1. $\lim_{x \to c}[f(x) + g(x)] = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$
2. $\lim_{x \to c} kf(x) = k \lim_{x \to c} f(x)$
3. $\lim_{x \to c} f(x)g(x) = \left[\lim_{x \to c} f(x)\right] \cdot \left[\lim_{x \to c} g(x)\right]$
4. $\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{\lim_{x \to c} f(x)}{\lim_{x \to c} g(x)}$, provided $\lim_{x \to c} g(x) \neq 0$.

Simple limits of functions can be solved as we did limits of sequences. Just be careful which part of the function is changing.

Example 3.3.

Limits of a function

Find the limit of the following functions.

1. $\lim_{x \to c} k$
2. $\lim_{x \to c} x$
3. $\lim_{x \to 2}(2x - 3)$
4. $\lim_{x \to c} x^n$

Limits can get more complex in roughly two ways. First, the functions may become large polynomials with many moving pieces. Second, the functions may become discontinuous.

The function can be thought of as a more general or "smooth" version of sequences. For example,

Example 3.4.

Limits of ratios

Find the limit of

$$\lim_{x\to\infty} \frac{(x^4+3x-99)(2-x^5)}{(18x^7+9x^6-3x^2-1)(x+1)}$$

Now, the functions will become a bit more complex:

Exercise 3.2.

Limits of a function

Solve the following limits of functions

1. $\lim_{x \to 0} |x|$
2. $\lim_{x \to 0} \left(1 + \frac{1}{x^2}\right)$

So there are a few more alternatives about what a limit of a function could be:

1. Right-hand limit: The value approached by $f(x)$ when you move from right to left.
2. Left-hand limit: The value approached by $f(x)$ when you move from left to right.
3. Infinity: The value approached by $f(x)$ as x grows infinitely large. Sometimes this may be a number; sometimes it might be ∞ or $-\infty$.
4. Negative infinity: The value approached by $f(x)$ as x grows infinitely negative. Sometimes this may be a number; sometimes it might be ∞ or $-\infty$.

The distinction between left and right becomes important when the function is not determined for some values of x. What are those cases in the examples below?

3.4 Continuity

To repeat a finding from the limits of functions: $f(x)$ does not necessarily have to be defined at c for $\lim_{x \to c}$ to exist. Functions that have breaks in their lines are called discontinuous. Functions that have no breaks are called continuous. Continuity is a concept that is more fundamental to, but related to that of “differentiability”, which we will cover next in calculus.

Definition 3.2.

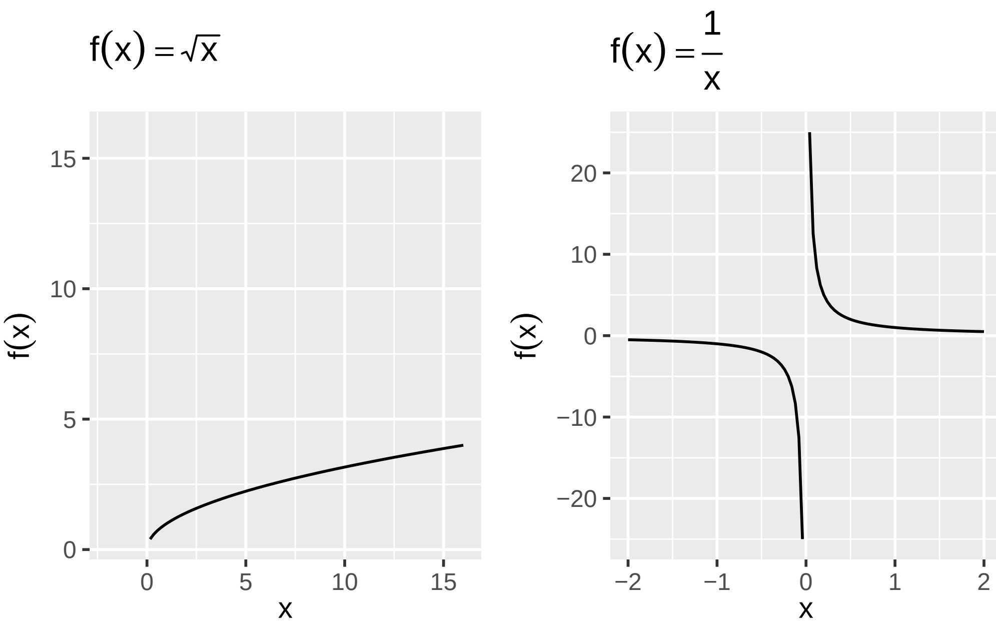

Figure 3.3: Functions which are not defined in some areas

Continuity

Suppose that the domain of the function f includes an open interval containing the point c. Then f is continuous at c if $\lim_{x \to c} f(x)$ exists and if $\lim_{x \to c} f(x) = f(c)$. Further, f is continuous on an open interval (a, b) if it is continuous at each point in the interval.

To prove that a function is continuous for all points is beyond this practical introduction to math, but the general intuition can be grasped by graphing.

Example 3.5.

Continuity

For each function, determine if it is continuous or discontinuous.

1. $f(x) = \sqrt{x}$
2. $f(x) = e^x$
3. $f(x) = 1 + \frac{1}{x^2}$
4. $f(x) = \text{floor}(x)$.

The floor is the smaller of the two integers bounding a number. So $\text{floor}(x = 2.999) = 2$, $\text{floor}(x = 2.0001) = 2$, and $\text{floor}(x = 2) = 2$.

Solution. In Figure 3.4, we can see that the first two functions are continuous, and the next two are discontinuous. $f(x) = 1 + \frac{1}{x^2}$ is discontinuous at $x = 0$, and $f(x) = \text{floor}(x)$ is discontinuous at each whole number.

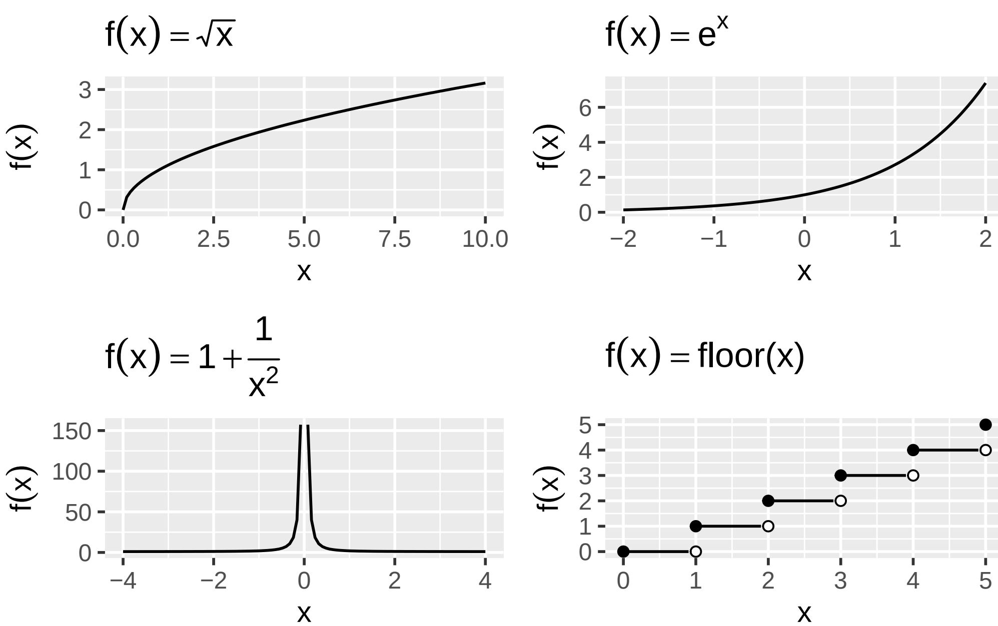

Figure 3.4: Continuous and Discontinuous Functions

Some properties of continuous functions:

1. If f and g are continuous at point c, then $f+g$, $f-g$, $f \cdot g$, $|f|$, and αf are continuous at point c also. f/g is continuous, provided $g(c) \neq 0$.
2. Boundedness: If f is continuous on the closed bounded interval $[a, b]$, then there is a number K such that $|f(x)| \leq K$ for each x in $[a, b]$.
3. Max/Min: If f is continuous on the closed bounded interval $[a, b]$, then f has a maximum and a minimum on $[a, b]$. They may be located at the end points.

Exercise

Let $f(x) = \frac{x^2+2x}{x}$.

1. Graph the function. Is it defined everywhere?
2. What is the functions limit at $x \to 0$?

4 Calculus

Calculus is a fundamental part of any type of statistics exercise. Although you may not be taking derivatives and integral in your daily work as an analyst, calculus undergirds many concepts we use: maximization, expectation, and cumulative probability.

Example: The Mean is a Type of Integral

The average of a quantity is a type of weighted mean, where the potential values are weighted by their likelihood, loosely speaking. The integral is actually a general way to describe this weighted average when there are conceptually an infinite number of potential values.

If X is a continuous random variable, its expected value $E(X)$ – the center of mass – is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

where $f(x)$ is the probability density function of X.

This is a continuous version of the case where X is discrete, in which case

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

even more concretely, if the potential values of X are finite, then we can write out the expected value as a weighted mean, where the weights is the probability that the value occurs.

$$E(X) = \sum_{x} \left(\underbrace{x}_{\text{value}} \cdot \underbrace{P(X = x)}_{\text{weight, or PMF}} \right)$$

4.1 Derivatives

The derivative of f at x is its rate of change at x: how much $f(x)$ changes with a change in x. The rate of change is a fraction — rise over run — but because not all lines are straight and the rise over run formula will give us different values depending on the range we examine, we need to take a limit (Section -Chapter 3).

Definition 4.1.

Derivative

Let f be a function whose domain includes an open interval containing the point x. The derivative of f at x is given by

$$\frac{d}{dx}f(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{(x+h) - x} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

There are a two main ways to denote a derivate:

- Leibniz Notation: $\frac{d}{dx}(f(x))$
- Prime or Lagrange Notation: $f'(x)$

If $f(x)$ is a straight line, the derivative is the slope. For a curve, the slope changes by the values of x, so the derivative is the slope of the line tangent to the curve at x. See, For example, Figure -Figure 4.1

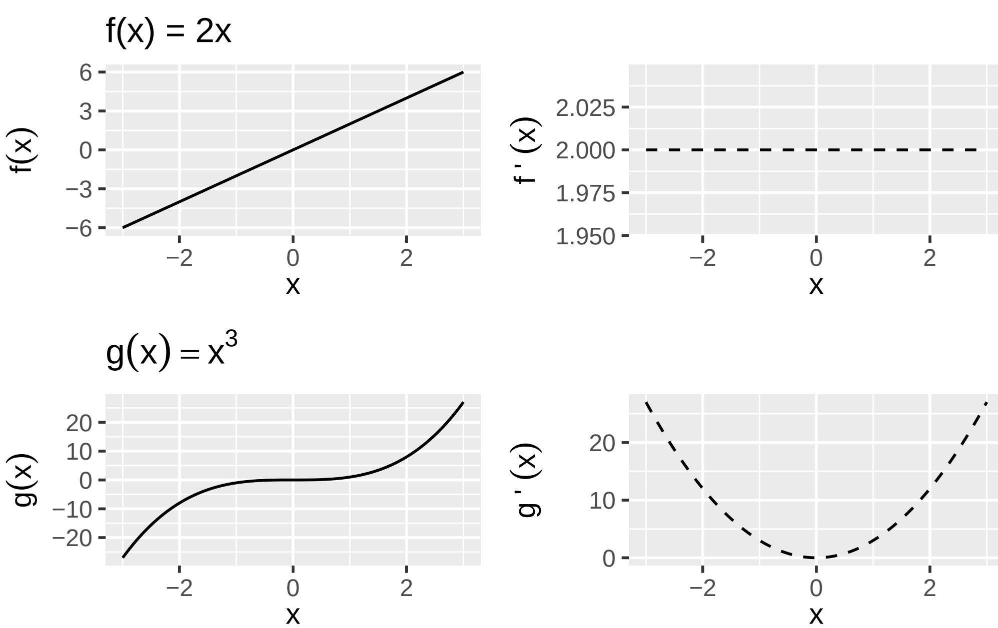

Figure 4.1: The Derivative as a Slope

If $f'(x)$ exists at a point x_0, then f is said to be **differentiable** at x_0. That also implies that $f(x)$ is continuous at x_0.

Properties of derivatives

Suppose that f and g are differentiable at x and that α is a constant. Then the functions $f \pm g$, αf, fg, and f/g (provided $g(x) \neq 0$) are also differentiable at x. Additionally,

Constant rule:

$$[kf(x)]' = kf'(x)$$

Sum rule:

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

With a bit more algebra, we can apply the definition of derivatives to get a formula for of the derivative of a product and a derivative of a quotient.

Product rule:

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

Quotient rule:

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

Finally, one way to think of the power of derivatives is that it takes a function a notch down in complexity. The power rule applies to any higher-order function:

Power rule:

$$[x^k]' = kx^{k-1}$$

For any real number k (that is, both whole numbers and fractions). The power rule is proved **by induction**, a neat method of proof used in many fundamental applications to prove that a general statement holds for every possible case, even if there are countably infinite cases. We'll show a simple case where k is an integer here.

Proposition 4.1.

Power Rule

$$\left[x^k\right]' = kx^{k-1}$$

for any integer k.

Proof. First, consider the first case (the base case) of $k = 1$. We can show by the definition of derivatives (setting $f(x) = x^1 = 1$) that

$$[x^1]' = \lim_{h \to 0} \frac{(x+h) - x}{(x+h) - x} = 1.$$

Because 1 is also expressed as $1x^{1-1}$, the statement we want to prove holds for the case $k = 1$.

Now, *assume* that the statement holds for some integer m. That is, assume

$$\left[x^m\right]' = mx^{m-1}$$

Then, for the case $m + 1$, using the product rule above, we can simplify

$$\begin{aligned}
\left[x^{m+1}\right]' &= [x^m \cdot x]' \\
&= (x^m)' \cdot x + (x^m) \cdot (x)' \\
&= mx^{m-1} \cdot x + x^m \quad \because \text{by previous assumption} \\
&= mx^m + x^m \\
&= (m+1)x^m \\
&= (m+1)x^{(m+1)-1}
\end{aligned}$$

Therefore, the rule holds for the case $k = m + 1$ once we have assumed it holds for $k = m$. Combined with the first case, this completes proof by induction – we have now proved that the statement holds for all integers $k = 1, 2, 3, \cdots$.

To show that it holds for real fractions as well, we can prove expressing that exponent by a fraction of two integers.

□

These “rules” become apparent by applying the definition of the derivative above to each of the things to be “derived”, but these come up so frequently that it is best to repeat until it is muscle memory.

Exercise 4.1.

Derivatives

For each of the following functions, find the first-order derivative $f'(x)$.

1. $f(x) = c$
2. $f(x) = x$
3. $f(x) = x^2$
4. $f(x) = x^3$
5. $f(x) = \frac{1}{x^2}$
6. $f(x) = (x^3)(2x^4)$
7. $f(x) = x^4 - x^3 + x^2 - x + 1$
8. $f(x) = (x^2 + 1)(x^3 - 1)$
9. $f(x) = 3x^2 + 2x^{1/3}$
10. $f(x) = \frac{x^2+1}{x^2-1}$

4.2 Higher-Order Derivatives (Derivatives of Derivatives of Derivatives)

The first derivative is applying the definition of derivatives on the function, and it can be expressed as

$$f'(x), \quad y', \quad \frac{d}{dx}f(x), \quad \frac{dy}{dx}$$

We can keep applying the differentiation process to functions that are themselves derivatives. The derivative of $f'(x)$ with respect to x, would then be

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

and we can therefore call it the **Second derivative:**

$$f''(x), \quad y'', \quad \frac{d^2}{dx^2}f(x), \quad \frac{d^2y}{dx^2}$$

Similarly, the derivative of $f''(x)$ would be called the third derivative and is denoted $f'''(x)$. And by extension, the **nth derivative** is expressed as $\frac{d^n}{dx^n}f(x)$, $\frac{d^ny}{dx^n}$.

Example 4.1.

Succession of derivatives

$$
\begin{aligned}
f(x) &= x^3 \\
f'(x) &= 3x^2 \\
f''(x) &= 6x \\
f'''(x) &= 6 \\
f''''(x) &= 0
\end{aligned}
$$

Earlier, in Section -Section 4.1, we said that if a function differentiable at a given point, then it must be continuous. Further, if $f'(x)$ is itself continuous, then $f(x)$ is called continuously differentiable. All of this matters because many of our findings about optimization (Section @ref(optim)) rely on differentiation, and so we want our function to be differentiable in as many layers. A function that is continuously differentiable infinitly is called "smooth". Some examples: $f(x) = x^2$, $f(x) = e^x$.

4.3 Composite Functions and the Chain Rule

As useful as the above rules are, many functions you'll see won't fit neatly in each case immediately. Instead, they will be functions of functions. For example, the difference between $x^2 + 1^2$ and $(x^2 + 1)^2$ may look trivial, but the sum rule can be easily applied to the former, while it's actually not obvious what do with the latter.

Composite functions are formed by substituting one function into another and are denoted by

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

To form $f[g(x)]$, the range of g must be contained (at least in part) within the domain of f. The domain of $f \circ g$ consists of all the points in the domain of g for which $g(x)$ is in the domain of f.

Example 4.2.

Composite functions

Let $f(x) = \log x$ for $0 < x < \infty$ and $g(x) = x^2$ for $-\infty < x < \infty$.

Then

$$(f \circ g)(x) = \log x^2, -\infty < x < \infty - \{0\}$$

Also

$$(g \circ f)(x) = [\log x]^2, 0 < x < \infty$$

Notice that $f \circ g$ and $g \circ f$ are not the same functions.

With the notation of composite functions in place, now we can introduce a helpful additional rule that will deal with a derivative of composite functions as a chain of concentric derivatives.

Chain Rule:

Let $y = (f \circ g)(x) = f[g(x)]$. The derivative of y with respect to x is

$$\frac{d}{dx}\{f[g(x)]\} = f'[g(x)]g'(x)$$

We can read this as: "the derivative of the composite function y is the derivative of f evaluated at $g(x)$, times the derivative of g."

The chain rule can be thought of as the derivative of the "outside" times the derivative of the "inside", remembering that the derivative of the outside function is evaluated at the value of the inside function.

- The chain rule can also be written as

  $$\frac{dy}{dx} = \frac{dy}{dg(x)}\frac{dg(x)}{dx}$$

  This expression does not imply that the $dg(x)$'s cancel out, as in fractions. They are part of the derivative notation and you can't separate them out or cancel them.)

Example 4.3.

Composite Exponent

Find $f'(x)$ for $f(x) = (3x^2 + 5x - 7)^6$.

The direct use of a chain rule is when the exponent of is itself a function, so the power rule could not have applied generaly:

Generalized Power Rule:

If $f(x) = [g(x)]^p$ for any rational number p,

$$f'(x) = p[g(x)]^{p-1} g'(x)$$

4.4 Derivatives of natural logs and the exponent

Natural logs and exponents (they are inverses of each other; see Section @ref(logexponents)) crop up everywhere in statistics. Their derivative is a special case from the above, but quite elegant.

Theorem 4.1.

Derivative of Exponents/Logs

The functions e^x *and the natural logarithm* $\log(x)$ *are continuous and differentiable in their domains, and their first derivative is*

$$(e^x)' = e^x$$

$$\log(x)' = \frac{1}{x}$$

Also, when these are composite functions, it follows by the generalized power rule that

$$\left(e^{g(x)}\right)' = e^{g(x)} \cdot g'(x)$$

$$(\log g(x))' = \frac{g'(x)}{g(x)}, \quad \textit{if} \;\; g(x) > 0$$

Derivatives of natural exponential function (e)

To repeat the main rule in Theorem @ref(thm:derivexplog), the intuition is that

1. Derivative of e^x is itself: $\frac{d}{dx}e^x = e^x$ (See Figure 4.2)
2. Same thing if there were a constant in front: $\frac{d}{dx}\alpha e^x = \alpha e^x$
3. Same thing no matter how many derivatives there are in front: $\frac{d^n}{dx^n}\alpha e^x = \alpha e^x$
4. Chain Rule: When the exponent is a function of x, remember to take derivative of that function and add to product. $\frac{d}{dx}e^{g(x)} = e^{g(x)}g'(x)$

Example 4.4.

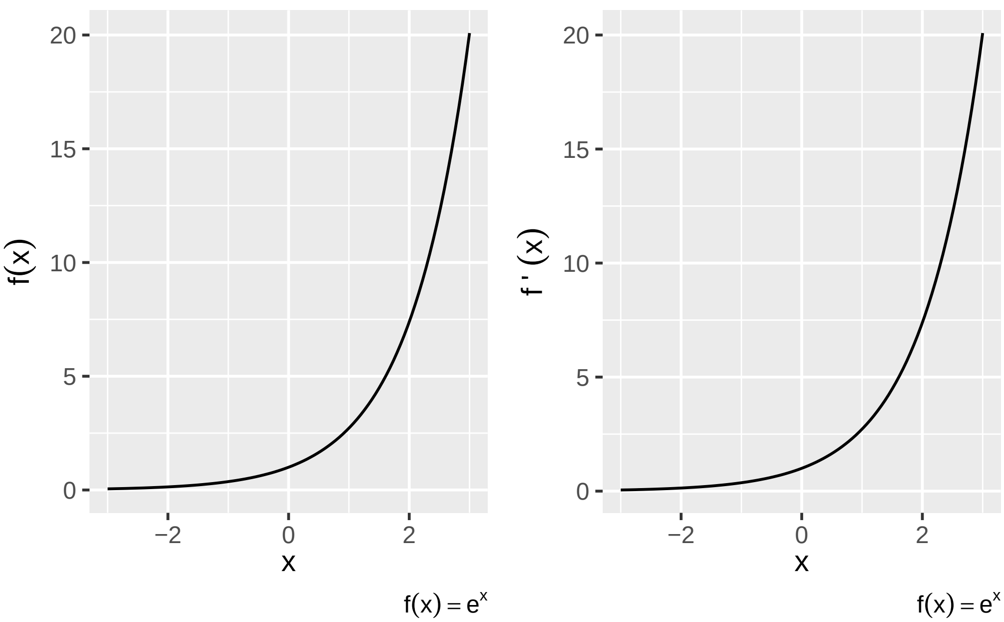

Figure 4.2: Derivative of the Exponential Function

Derivatives of exponents

Find the derivative for the following.

1. $f(x) = e^{-3x}$
2. $f(x) = e^{x^2}$
3. $f(x) = (x-1)e^x$

Derivatives of logarithms

The natural log is the mirror image of the natural exponent and has mirroring properties, again, to repeat the theorem,

1. log prime x is one over x: $\frac{d}{dx} \log x = \frac{1}{x}$ (Figure 4.3)
2. Exponents become multiplicative constants: $\frac{d}{dx} \log x^k = \frac{d}{dx} k \log x = \frac{k}{x}$
3. Chain rule again: $\frac{d}{dx} \log u(x) = \frac{u'(x)}{u(x)}$
4. For any positive base b, $\frac{d}{dx} b^x = (\log b)\left(b^x\right)$.

Example 4.5.

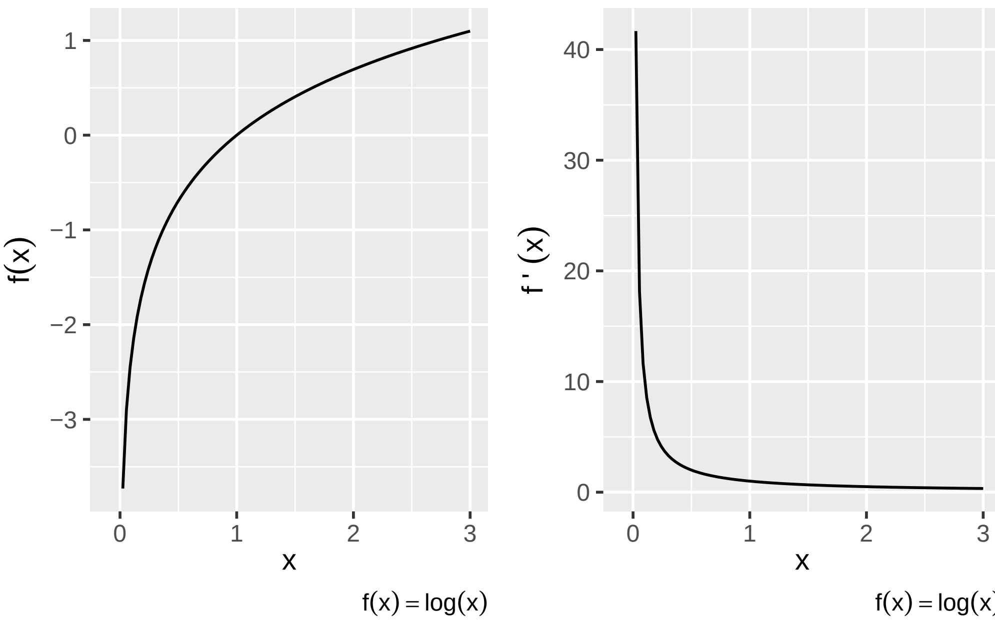

Figure 4.3: Derivative of the Natural Log

Derivatives of logs

Find dy/dx for the following.

1. $f(x) = \log(x^2 + 9)$
2. $f(x) = \log(\log x)$
3. $f(x) = (\log x)^2$
4. $f(x) = \log e^x$

Outline of Proof

We actually show the derivative of the log first, and then the derivative of the exponential naturally follows.

The general derivative of the log at any base a is solvable by the definition of derivatives.

$$(\log_a x)' = \lim_{h \to 0} \frac{1}{h} \log_a \left(1 + \frac{h}{x}\right)$$

Re-express $g = \frac{h}{x}$ and get

$$\begin{aligned}(\log_a x)' &= \frac{1}{x} \lim_{g \to 0} \log_a (1 + g)^{\frac{1}{g}} \\ &= \frac{1}{x} \log_a e\end{aligned}$$

By definition of e. As a special case, when $a = e$, then $(\log x)' = \frac{1}{x}$.

Now let's think about the inverse, taking the derivative of $y = a^x$.

$$\begin{aligned} y &= a^x \\ \Rightarrow \log y &= x \log a \\ \Rightarrow \frac{y'}{y} &= \log a \\ \Rightarrow y' &= y \log a \end{aligned}$$

Then in the special case where $a = e$,

$$(e^x)' = (e^x)$$

4.5 Partial Derivatives

What happens when there's more than variable that is changing?

> If you can do ordinary derivatives, you can do partial derivatives: just hold all the other input variables constant except for the one you're differentiating with respect to. (Joe Blitzstein's Math Notes)

Suppose we have a function f now of two (or more) variables and we want to determine the rate of change relative to one of the variables. To do so, we would find its partial derivative, which is defined similar to the derivative of a function of one variable.

Partial Derivative: Let f be a function of the variables $(x_1, \dots, x_n)$. The partial derivative of f with respect to x_i is

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

Only the ith variable changes — the others are treated as constants.

We can take higher-order partial derivatives, like we did with functions of a single variable, except now the higher-order partials can be with respect to multiple variables.

Example 4.6.

Partial derivatives

Notice that you can take partials with regard to different variables.

Suppose $f(x, y) = x^2 + y^2$. Then

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y) &= \\
\frac{\partial f}{\partial y}(x, y) &= \\
\frac{\partial^2 f}{\partial x^2}(x, y) &= \\
\frac{\partial^2 f}{\partial x \partial y}(x, y) &=
\end{aligned}$$

Exercise 4.2.

Partial derivatives

Let $f(x, y) = x^3y^4 + e^x - \log y$. What are the following partial derivatives?

$$
\begin{aligned}
\frac{\partial f}{\partial x}(x, y) &= \\
\frac{\partial f}{\partial y}(x, y) &= \\
\frac{\partial^2 f}{\partial x^2}(x, y) &= \\
\frac{\partial^2 f}{\partial x \partial y}(x, y) &=
\end{aligned}
$$

4.6 Taylor Series Approximation

A common form of approximation used in statistics involves derivatives. A Taylor series is a way to represent common functions as infinite series (a sum of infinite elements) of the function's derivatives at some point a.

For example, Taylor series are very helpful in representing nonlinear (read: difficult) functions as linear (read: manageable) functions. One can thus **approximate** functions by using lower-order, finite series known as **Taylor polynomials**. If $a = 0$, the series is called a Maclaurin series.

Specifically, a Taylor series of a real or complex function $f(x)$ that is infinitely differentiable in the neighborhood of point a is:

$$
\begin{aligned}
f(x) &= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots \\
&= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n
\end{aligned}
$$

Taylor Approximation: We can often approximate the curvature of a function $f(x)$ at point a using a 2nd order Taylor polynomial around point a:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + R_2$$

R_2 is the remainder (R for remainder, 2 for the fact that we took two derivatives) and often treated as negligible, giving us:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

The more derivatives that are added, the smaller the remainder R and the more accurate the approximation. Proofs involving limits guarantee that the remainder converges to 0 as the order of derivation increases.

4.7 The Indefinite Integration

So far, we've been interested in finding the derivative $f = F'$ of a function F. However, sometimes we're interested in exactly the reverse: finding the function F for which f is its derivative. We refer to F as the antiderivative of f.

Definition 4.2.

Antiderivative

The antiverivative of a function $f(x)$ is a differentiable function F whose derivative is f.

$$F' = f.$$

Another way to describe is through the inverse formula. Let DF be the derivative of F. And let $DF(x)$ be the derivative of F evaluated at x. Then the antiderivative is denoted by D^{-1} (i.e., the inverse derivative). If $DF = f$, then $F = D^{-1}f$.

This definition bolsters the main takeaway about integrals and derivatives: They are inverses of each other.

Exercise 4.3.

Antiderivative

Find the antiderivative of the following:

1. $f(x) = \frac{1}{x^2}$
2. $f(x) = 3e^{3x}$

We know from derivatives how to manipulate F to get f. But how do you express the procedure to manipulate f to get F? For that, we need a new symbol, which we will call indefinite integration.

:::{#def-indefint}

5 Indefinite Integral

The indefinite integral of $f(x)$ is written

$$\int f(x)dx$$

and is equal to the antiderivative of f.

Example 5.1.

Graphing

Draw the function $f(x)$ and its indefinite integral, $\int f(x)dx$

$$f(x) = (x^2 - 4)$$

Solution. The Indefinite Integral of the function $f(x) = (x^2 - 4)$ can, for example, be $F(x) = \frac{1}{3}x^3 - 4x$. But it can also be $F(x) = \frac{1}{3}x^3 - 4x + 1$, because the constant 1 disappears when taking the derivative.

Some of these functions are plotted in the bottom panel of Figure 5.1 as dotted lines.

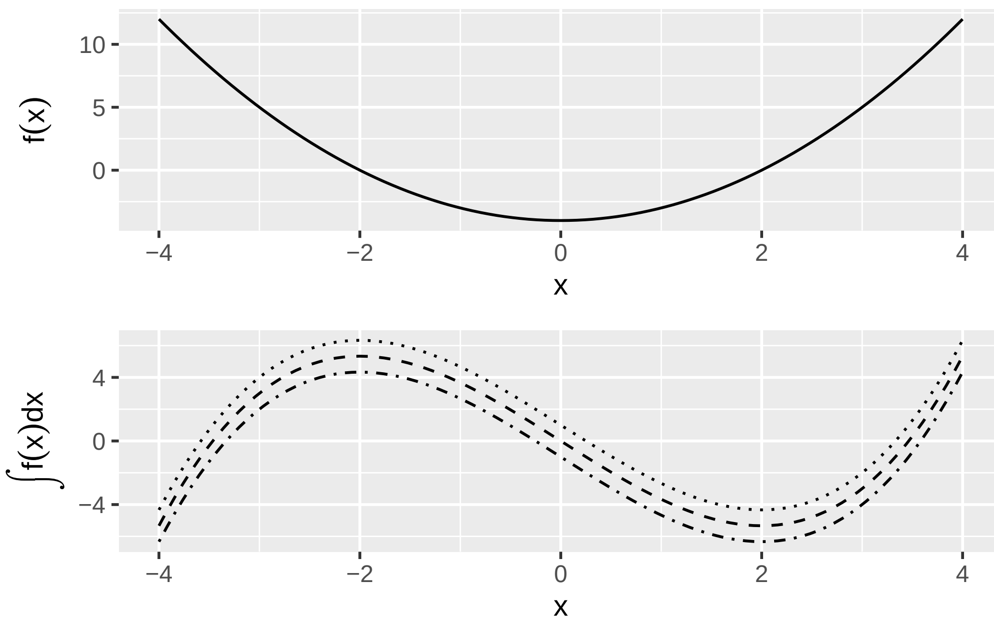

Figure 5.1: The Many Indefinite Integrals of a Function

Notice from these examples that while there is only a single derivative for any function, there are multiple antiderivatives: one for any arbitrary constant c. c just shifts the curve up or down on the y-axis. If more information is present about the antiderivative — e.g., that it passes through a particular point — then we can solve for a specific value of c.

Common Rules of Integration

Some common rules of integrals follow by virtue of being the inverse of a derivative.

1. Constants are allowed to slip out: $\int af(x)dx = a\int f(x)dx$
2. Integration of the sum is sum of integrations: $\int[f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$
3. Reverse Power-rule: $\int x^n dx = \frac{1}{n+1}x^{n+1} + c$
4. Exponents are still exponents: $\int e^x dx = e^x + c$
5. Recall the derivative of $\log(x)$ is one over x, and so: $\int \frac{1}{x}dx = \log x + c$
6. Reverse chain-rule: $\int e^{f(x)} f'(x)dx = e^{f(x)} + c$
7. More generally: $\int[f(x)]^n f'(x)dx = \frac{1}{n+1}[f(x)]^{n+1} + c$
8. Remember the derivative of a log of a function: $\int \frac{f'(x)}{f(x)}dx = \log f(x) + c$

Example 5.2.

Common Integration

Simplify the following indefinite integrals:

- (3x^2 dx)
- ((2x+1)dx)
- (e^x e$^{\{e}$x} dx)

5.1 The Definite Integral: The Area under the Curve

If there is a indefinite integral, there *must* be a definite integral. Indeed there is, but the notion of definite integrals comes from a different objective: finding the are a under a function. We will find, perhaps remarkably, that the formula we find to get the sum turns out to be expressible by the anti-derivative.

Suppose we want to determine the area $A(R)$ of a region R defined by a curve $f(x)$ and some interval $a \leq x \leq b$.

One way to calculate the area would be to divide the interval $a \leq x \leq b$ into n subintervals of length Δx and then approximate the region with a series of rectangles, where the base of each rectangle is Δx and the height is $f(x)$ at the midpoint of that interval. $A(R)$ would then be approximated by the area of the union of the rectangles, which is given by

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

and is called a **Riemann sum**.

As we decrease the size of the subintervals Δx, making the rectangles “thinner,” we would expect our approximation of the area of the region to become closer to the true area. This allows us to express the area as a limit of a series:

$$A(R) = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i)\Delta x$$

Figure 5.2 shows that illustration. The curve depicted is $f(x) = -15(x-5) + (x-5)^3 + 50$. We want approximate the area under the curve between the x values of 0 and 10. We can do

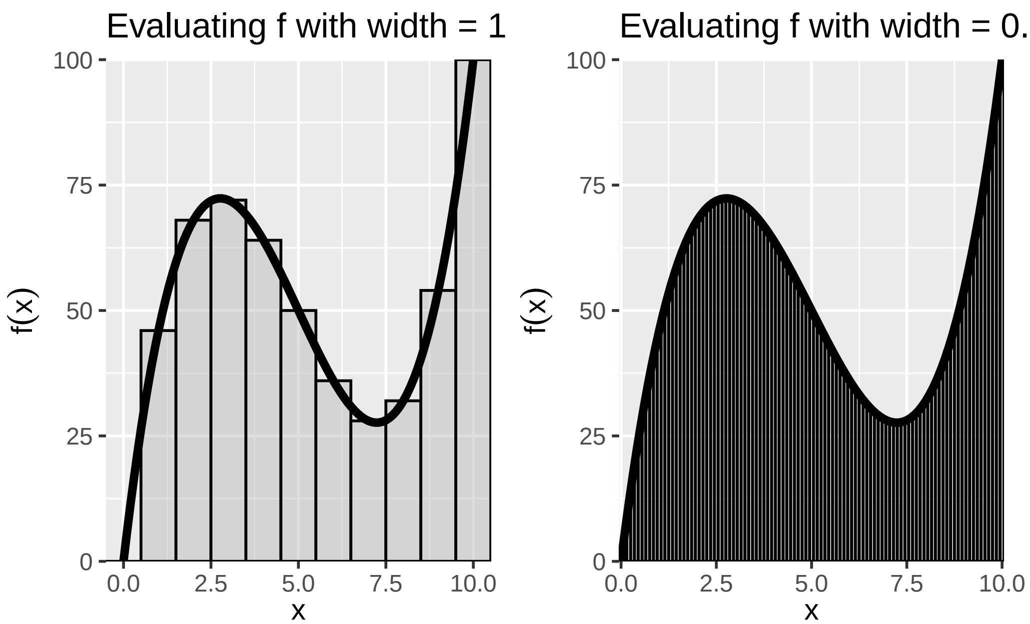

Figure 5.2: The Riemann Integral as a Sum of Evaluations

this in blocks of arbitrary width, where the sum of rectangles (the area of which is width times $f(x)$ evaluated at the midpoint of the bar) shows the Riemann Sum. As the width of the bars Δx becomes smaller, the better the estimate of $A(R)$.

This is how we define the "Definite" Integral:

Definition 5.1.

The Definite Integral (Riemann)

If for a given function f the Riemann sum approaches a limit as $\Delta x \to 0$, then that limit is called the Riemann integral of f from a to b. We express this with the $\int$, symbol, and write

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

The most straightforward of a definite integral is the definite integral. That is, we read

$$\int\limits_a^b f(x)dx$$

as the definite integral of f from a to b and we defined as the area under the "curve" $f(x)$ from point $x = a$ to $x = b$.

The fundamental theorem of calculus shows us that this sum is, in fact, the antiderivative.

Theorem 5.1.

First Fundamental Theorem of Calculus

Let the function f be bounded on $[a, b]$ and continuous on (a, b). Then, suggestively, use the symbol $F(x)$ to denote the definite integral from a to x:

$$F(x) = \int_a^x f(t)dt, \quad a \leq x \leq b$$

Then $F(x)$ has a derivative at each point in (a, b) and

$$F'(x) = f(x), \quad a < x < b$$

That is, the definite integral function of f is *the one of the antiderivatives of some f.*

This is again a long way of saying that that differentiation is the inverse of integration. But now, we've covered definite integrals.

The second theorem gives us a simple way of computing a definite integral as a function of indefinite integrals.

Theorem 5.2.

Second Fundamental Theorem of Calculus

Let the function f be bounded on $[a, b]$ and continuous on (a, b). Let F be any function that is continuous on $[a, b]$ such that $F'(x) = f(x)$ on (a, b). Then

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

So the procedure to calculate a simple definite integral $\int_a^b f(x)dx$ is then

1. Find the indefinite integral $F(x)$.
2. Evaluate $F(b) - F(a)$.

Example 5.3.

Definite Integral of a monomial

Solve $\int_1^3 3x^2 dx$.

Let $f(x) = 3x^2$.

Exercise 5.1.

Indefinite integrals

What is the value of $\int_{-2}^{2} e^x e^{e^x} dx$?

Common Rules for Definite Integrals

The area-interpretation of the definite integral provides some rules for simplification.

1. There is no area below a point:
$$\int_a^a f(x)dx = 0$$
2. Reversing the limits changes the sign of the integral:
$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$
3. Sums can be separated into their own integrals:
$$\int_a^b [\alpha f(x) + \beta g(x)]dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx$$
4. Areas can be combined as long as limits are linked:
$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$$

Exercise 5.2.

Definite integrals

Simplify the following definite intergrals.

1. $\int\limits_{1}^{1} 3x^2 dx =$
2. $\int\limits_{0}^{4} (2x+1) dx =$
3. $\int\limits_{-2}^{0} e^x e^{e^x} dx + \int\limits_{0}^{2} e^x e^{e^x} dx =$

5.2 Integration by Substitution

From the second fundamental theorem of calculus, we now that a quick way to get a definite integral is to first find the indefinite integral, and then just plug in the bounds.

Sometimes the integrand (the thing that we are trying to take an integral of) doesn't appear integrable using common rules and antiderivatives. A method one might try is **integration by substitution**, which is related to the Chain Rule.

Suppose we want to find the indefinite integral

$$\int g(x) dx$$

but $g(x)$ is complex and none of the formulas we have seen so far seem to apply immediately. The trick is to come up with a *new* function $u(x)$ such that

$$g(x) = f[u(x)]u'(x).$$

Why does an introduction of yet another function end of simplifying things? Let's refer to the antiderivative of f as F. Then the chain rule tells us that

$$\frac{d}{dx} F[u(x)] = f[u(x)]u'(x)$$

. So, $F[u(x)]$ is the antiderivative of g. We can then write

$$\int g(x) dx = \int f[u(x)]u'(x) dx = \int \frac{d}{dx} F[u(x)] dx = F[u(x)] + c$$

To summarize, the procedure to determine the indefinite integral $\int g(x)dx$ by the method of substitution:

1. Identify some part of $g(x)$ that might be simplified by substituting in a single variable u (which will then be a function of x).
2. Determine if $g(x)dx$ can be reformulated in terms of u and du.
3. Solve the indefinite integral.
4. Substitute back in for x

Substitution can also be used to calculate a definite integral. Using the same procedure as above,

$$\int_a^b g(x)dx = \int_c^d f(u)du = F(d) - F(c)$$

where $c = u(a)$ and $d = u(b)$.

Example 5.4. Integration by Substitution I

Solve the indefinite integral

$$\int x^2\sqrt{x+1}dx.$$

For the above problem, we could have also used the substitution $u = \sqrt{x+1}$. Then $x = u^2 - 1$ and $dx = 2udu$. Substituting these in, we get

$$\int x^2\sqrt{x+1}dx = \int (u^2-1)^2 u2udu$$

which when expanded is again a polynomial and gives the same result as above.

Another case in which integration by substitution is is useful is with a fraction.

Example 5.5.

Integration by Substitutiton II

Simplify

$$\int_0^1 \frac{5e^{2x}}{(1+e^{2x})^{1/3}}dx.$$

5.3 Integration by Parts

Another useful integration technique is **integration by parts**, which is related to the Product Rule of differentiation. The product rule states that

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

Integrating this and rearranging, we get

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$

or

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx$$

More easily remembered with the mnemonic "Ultraviolet Voodoo":

$$\int udv = uv - \int vdu$$

where $du = u'(x)dx$ and $dv = v'(x)dx$.

For definite integrals, this is simply

$$\int_a^b u\frac{dv}{dx}dx = uv|_a^b - \int_a^b v\frac{du}{dx}dx$$

Our goal here is to find expressions for u and dv that, when substituted into the above equation, yield an expression that's more easily evaluated.

Example 5.6.

Integration by parts

Simplify the following integrals. These seemingly obscure forms of integrals come up often when integrating distributions.

$$\int xe^{ax}dx$$

Solution. Let $u = x$ and $\frac{dv}{dx} = e^{ax}$. Then $du = dx$ and $v = (1/a)e^{ax}$. Substituting this into the integration by parts formula, we obtain

$$\begin{aligned}\int xe^{ax}dx &= uv - \int vdu \\ &= x\left(\frac{1}{a}e^{ax}\right) - \int \frac{1}{a}e^{ax}dx \\ &= \frac{1}{a}xe^{ax} - \frac{1}{a^2}e^{ax} + c\end{aligned}$$

Exercise 5.3.

Integration by parts

1. Integrate
$$\int x^n e^{ax} dx$$

2. Integrate
$$\int x^3 e^{-x^2} dx$$