Mathematics for Political Scientists

Master Political Science

Carlos Gueiros

University of Mannheim

August 29 - September 2, 2022

Introduction

Course Objectives

What is this course about?

Course Objectives

What is this course about?

- Recap of your high-school / Abitur knowledge in mathematics.
- Introduction to the fundamentals in math that are necessary for your understanding of statistics and game theory.
- Overcome possible reservations against the use of mathematics.
- A refresher and starting point for future individual learning.

Course Objectives

What is this course about?

- Recap of your high-school / Abitur knowledge in mathematics.
- Introduction to the fundamentals in math that are necessary for your understanding of statistics and game theory.
- Overcome possible reservations against the use of mathematics.
- A refresher and starting point for future individual learning.

What is this course not about?

Course Objectives

What is this course about?

- Recap of your high-school / Abitur knowledge in mathematics.
- Introduction to the fundamentals in math that are necessary for your understanding of statistics and game theory.
- Overcome possible reservations against the use of mathematics.
- A refresher and starting point for future individual learning.

What is this course not about?

- It is not a mathematical freak show!
- It does not introduce into advanced mathematical techniques.

Math and applications in political science

Why is math important to social scientists?

Math and applications in political science

Why is math important to social scientists?

- Mathematics allows for orderly and systematic communication. Ideas expressed mathematically can be more carefully defined and more directly communicated than narrative language, which is susceptible to vagueness and misinterpretation.
- Mathematics is an effective way to describe and model our world.

Math and applications in political science

Why is math important to social scientists?

- Mathematics allows for orderly and systematic communication. Ideas expressed mathematically can be more carefully defined and more directly communicated than narrative language, which is susceptible to vagueness and misinterpretation.
- Mathematics is an effective way to describe and model our world.

Applications

Math and applications in political science

Why is math important to social scientists?

- Mathematics allows for orderly and systematic communication. Ideas expressed mathematically can be more carefully defined and more directly communicated than narrative language, which is susceptible to vagueness and misinterpretation.
- Mathematics is an effective way to describe and model our world.

Applications

- Game Theory, Decision Theory
- Computer Simulation, Agent-Based Modeling
- Statistics, Econometrics
- Empirical Analyses in any field

Syllabus

I Set Theory
- introduction, functions, binary relations

II Analysis
- derivatives, optimization, integration

III Linear Algebra
- vectors, matrices, linear equations

IV Probability Theory
- combinatorics, conditional probabilities, distributions

Organization

- ILIAS and Course Website[1]: Syllabus, slides, exercises and extra materials.
- Lecture + Slides
- Exercises
  - on the board
  - independent work and self-study
  - group work
- Active participation

[1] https://math-refresher-22.netlify.app/

Organization

- ILIAS and Course Website[1]: Syllabus, slides, exercises and extra materials.
- Lecture + Slides
- Exercises
  - on the board
  - independent work and self-study
  - group work
- Active participation

This course is voluntary!

[1] https://math-refresher-22.netlify.app/

Schedule

Date	Day	Time	Room
29.08.2022	Monday	14:00 - 17:15	B 317 Seminarraum
30.08.2022	Tuesday	9:15 - 17:15	B 317 Seminarraum
31.08.2022	Wednesday	9:15 - 13:15	B 317 Seminarraum
01.09.2022	Thursday	9:15 - 17:15	B 317 Seminarraum
02.09.2022	Friday	9:15 - 12:30	B 317 Seminarraum

General Readings

Recommended:

General

- Gill (2006): Essential Mathematics for Political and Social Research.
- Moore/Siegel (2013): A Mathematics Course for Political and Social Research. *An introductory mathematics course aimed at social scientists, provides good intuitions for basic concepts and applications. It has accompanying video lectures on Youtube.*
- Simon/Blume (1994) *A comprehensive treatment of mathematics for students of economics for both undergraduate and more advanced level.*
- Sydsaeter/Hammond (2008) *Another standard mathematics textbook for economics undergraduates.*

Specific Readings

- Calculus
  - Spivak (2006) *A classic standard textbook for a first class in Calculus for mathematics students at undergraduate level.*
- Probability Theory
  - DeGroot/Schervish (2011) *A comprehensive standard treatment of probability and statistics for mathematics undergraduate students. Intuitive and (relatively) rigorous at the same time with lots of exercises.*
- Linear Algebra
  - Lay (2011) *A standard introduction for mathematics undergraduates.*
  - The Matrix Cookbook[2]
    An overview over some more advanced matrix calculus.

[2] http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf

Set Theory

Motivation

Explanations of political outcomes often begin with the presumption that such outcomes are the result of purposive decisions made by relevant individuals (e.g. voters, legislators) or groups of individuals (e.g. political parties, interest groups, nation states)

Motivation

Explanations of political outcomes often begin with the presumption that such outcomes are the result of purposive decisions made by relevant individuals (e.g. voters, legislators) or groups of individuals (e.g. political parties, interest groups, nation states)

Fundamental to these kind of explanations are the concepts of ‘choice’ and ‘preferences’.

Motivation

Explanations of political outcomes often begin with the presumption that such outcomes are the result of purposive decisions made by relevant individuals (e.g. voters, legislators) or groups of individuals (e.g. political parties, interest groups, nation states)

Fundamental to these kind of explanations are the concepts of 'choice' and 'preferences'.

Set Theory is fundamental to the formalization of these concepts.

Motivation

Explanations of political outcomes often begin with the presumption that such outcomes are the result of purposive decisions made by relevant individuals (e.g. voters, legislators) or groups of individuals (e.g. political parties, interest groups, nation states)

Fundamental to these kind of explanations are the concepts of 'choice' and 'preferences'.

Set Theory is fundamental to the formalization of these concepts.

Set Theory is fundamental to the understanding of many other fields of mathematics, e.g. the concept of 'functions'.

What Is a Set?

Definition (Set)

A **set** is a collection of distinct objects, where the objects therein are called **elements** or **members**.

What Is a Set?

Definition (Set)

A **set** is a collection of distinct objects, where the objects therein are called **elements** or **members**.

For example $A = \{1, 2, 3\}$ is a set, and 1 is an element of A (write $1 \in A$), whereas 4 is not an element of A ($4 \notin A$).

What Is a Set?

Definition (Set)

A **set** is a collection of distinct objects, where the objects therein are called **elements** or **members**.

For example $A = \{1, 2, 3\}$ is a set, and 1 is an element of A (write $1 \in A$), whereas 4 is not an element of A ($4 \notin A$).

If a set does not contain any elements, we call it an **empty set**. The shorthand for an empty set is $\emptyset$ or $\{\}$.

Example: Sets of Numbers

Symbol	Explanation	Example
$\mathbb{N}$	set of natural numbers	$1, 2, 3, 4, \ldots$
$\mathbb{Z}$	set of integers	$-2, -1, 0, 1, 2, \ldots$
$\mathbb{Q}$	set of rational numbers (fractions)	$-\frac{9}{7}, -1, 0, \frac{1}{2}, 1, \ldots$
$\mathbb{R}$	set of real numbers	fractions plus e.g. π or e
$\mathbb{R}^+$	set of positive real numbers	
$\mathbb{C}$	set of complex numbers	$\sqrt{-1}$

Relations of Sets

A set itself can, furthermore, be part of another set. E.g. $A = \{1, 2, 3\}$ is part of $B = \{1, 2, 3, 4\}$. We then say that A is a **subset** of B and write $A \subseteq B$.

Relations of Sets

A set itself can, furthermore, be part of another set. E.g. $A = \{1, 2, 3\}$ is part of $B = \{1, 2, 3, 4\}$. We then say that A is a **subset** of B and write $A \subseteq B$. In particular it is true for every set A that $A \subseteq A$.

Relations of Sets

A set itself can, furthermore, be part of another set. E.g. $A = \{1, 2, 3\}$ is part of $B = \{1, 2, 3, 4\}$. We then say that A is a **subset** of B and write $A \subseteq B$. In particular it is true for every set A that $A \subseteq A$.

If A is a subset of B, but not equal to B (like in the example above), we call A a **proper** or **strict subset** of B and write $A \subset B$.

Relations of Sets

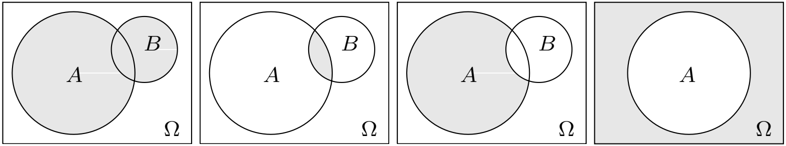
A set itself can, furthermore, be part of another set. E.g. $A = \{1, 2, 3\}$ is part of $B = \{1, 2, 3, 4\}$. We then say that A is a **subset** of B and write $A \subseteq B$. In particular it is true for every set A that $A \subseteq A$.

If A is a subset of B, but not equal to B (like in the example above), we call A a **proper** or **strict subset** of B and write $A \subset B$.

If two sets do not have any element in common, these sets are said to be **disjoint**. E.g. $A = \{1, 2, 3\}$ and $C = \{4, 5\}$ are disjoint.

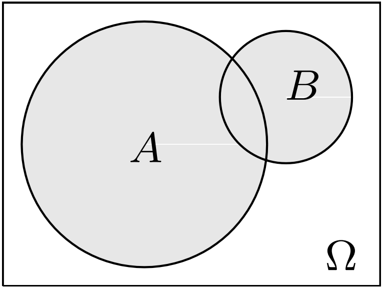
Operations on Sets I

We can visualize operations on sets using so called **Venn diagrams**.



Operations on Sets II

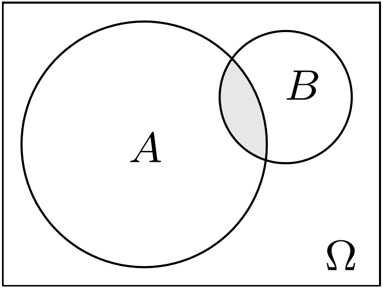
A **union** contains all elements that are either in A or B or in both. Formally, this is $A \cup B = \{x | x \in A \text{ or } x \in B \text{ or both}\}$.



If $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A \cup B = \{1, 2, 3, 4\}$.

Operations on Sets III

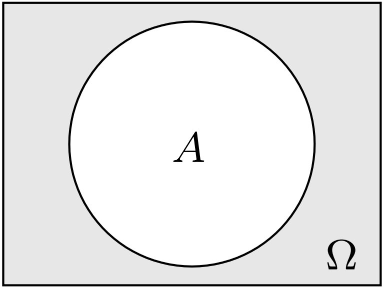
An **intersection** contains all elements that are both in A and B.
Formally, this is $A \cap B = \{x | x \in A \text{ and } x \in B\}$.



If $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A \cap B = \{3\}$.

Operations on Sets IV

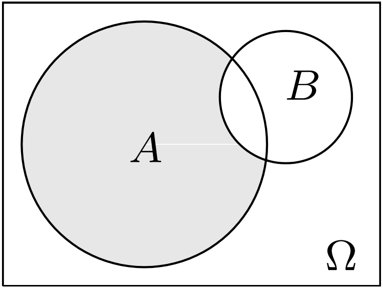
Let there be a **universal set** Ω with the subset A. The **complement** of A is every element of Ω that is not an element of A.
Formally, this is $A^C = \{x | x \notin A \text{ (and } x \in \Omega)\}$.



If $A = \{1, 2, 3\}$ and $\Omega = \{1, 2, 3, 4, 5\}$, then $A^C = \{4, 5\}$.

Operations on Sets V

We can also form **differences** of sets.
$A \setminus B = \{x | x \in A \text{ and } x \notin B\}$.



If $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2\}$, then $A \setminus B = \{3, 4, 5\}$.

Cardinality

The **cardinality** of a set is a measure of the number of elements in the set.
Usually denoted with $|A|$ (alternatives: $n(A)$, $card(A)$ or $\#A$).

If $A = \{1, 2, 3, 4, 5\}$, then $|A| = 5$.

Summary of definitions

$\emptyset$	empty set
$\cup$	union of two sets
$\cap$	intersection of two sets
$\subseteq$	is a subset of
$\subset$	is a strict subset of
$\supseteq$	is a superset of
$\supset$	is a strict superset of

Useful Notation

$\in$	is an element of
$\forall$	for all
$\exists$	there exists
$\Rightarrow$	implies
$\Leftrightarrow$, iff	if and only if
: or $s.t.$	such that
$\equiv$	equivalent to
$\sim$ or $\neg$	not
$\setminus$	without

Laws of Set Theory

Commutative
$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Laws of Set Theory

Commutative

$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Associative

$(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$

Laws of Set Theory

Commutative
$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Associative
$(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$

Idempotent
$A \cap A = A$ and $A \cup A = A$

Laws of Set Theory

Commutative
$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Associative
$(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$

Idempotent
$A \cap A = A$ and $A \cup A = A$

Distributive
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Laws of Set Theory

Commutative
$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Associative
$(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$

Idempotent
$A \cap A = A$ and $A \cup A = A$

Distributive
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

De Morgan's Laws
$(A \cup B)^C = A^C \cap B^C$ and $(A \cap B)^C = A^C \cup B^C$

Laws of Set Theory

Commutative
$A \cup B = B \cup A$ and $A \cap B = B \cap A$

Associative
$(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$

Idempotent
$A \cap A = A$ and $A \cup A = A$

Distributive
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

De Morgan's Laws
$(A \cup B)^C = A^C \cap B^C$ and $(A \cap B)^C = A^C \cup B^C$
$A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$ and $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

Spaces

Remember: $\mathbb{R}^1$ is the set of real numbers extending from $-\infty$ to ∞, the real number line.

$\mathbb{R}^n$ is an n-dimensional space **("Euclidean space")**, where each of the n axes extends from $-\infty$ to ∞.

Examples:

- $\mathbb{R}^1$ ($\mathbb{R}$) is a line.
- $\mathbb{R}^2$ is a plane.
- $\mathbb{R}^3$ is a 3D-space.

Points in $\mathbb{R}^n$ are ordered n-tuples, where each element of the n-tuple represents the coordinate along that dimension.

Interval Notation for $\mathbb{R}^1$

Open interval: $(a, b) \equiv \{x \in \mathbb{R}^1 : a < x < b\}$

Closed interval: $[a, b] \equiv \{x \in \mathbb{R}^1 : a \leq x \leq b\}$

Half open, half closed interval: $(a, b] \equiv \{x \in \mathbb{R}^1 : a < x \leq b\}$

Neighborhoods: Intervals, Disks, and Balls

We need a formal construct for what it means to be "near" a point $\mathbf{c}$ in $\mathbb{R}^n$. We call this the **neighborhood** of $\mathbf{c}$ and represent it by an open interval, disk, or ball, depending on whether n is one, two, or more dimensions, respectively. Given the point $\mathbf{c}$, these are defined as

Neighborhoods: Intervals, Disks, and Balls

We need a formal construct for what it means to be "near" a point **c** in $\mathbb{R}^n$. We call this the **neighborhood** of **c** and represent it by an open interval, disk, or ball, depending on whether n is one, two, or more dimensions, respectively. Given the point **c**, these are defined as

- ϵ-**interval** in $\mathbb{R}^1$: $\{x : |x - c| < \epsilon\}$
  The open interval $(c - \epsilon, c + \epsilon)$.

Neighborhoods: Intervals, Disks, and Balls

We need a formal construct for what it means to be "near" a point **c** in $\mathbb{R}^n$. We call this the **neighborhood** of **c** and represent it by an open interval, disk, or ball, depending on whether n is one, two, or more dimensions, respectively. Given the point **c**, these are defined as

- ϵ-**interval** in $\mathbb{R}^1$: $\{x : |x - c| < \epsilon\}$
  The open interval $(c - \epsilon, c + \epsilon)$.
- ϵ-**disk** in $\mathbb{R}^2$: $\{x : ||x - c|| < \epsilon\}$
  The open interior of the circle centered at **c** with radius ϵ.

Neighborhoods: Intervals, Disks, and Balls

We need a formal construct for what it means to be "near" a point **c** in $\mathbb{R}^n$. We call this the **neighborhood** of **c** and represent it by an open interval, disk, or ball, depending on whether n is one, two, or more dimensions, respectively. Given the point **c**, these are defined as

- ϵ-**interval** in $\mathbb{R}^1$: $\{x : |x - c| < \epsilon\}$
  The open interval $(c - \epsilon, c + \epsilon)$.
- ϵ-**disk** in $\mathbb{R}^2$: $\{x : ||x - c|| < \epsilon\}$
  The open interior of the circle centered at **c** with radius ϵ.
- ϵ-**ball** in $\mathbb{R}^n$: $\{x : ||x - c|| < \epsilon\}$
  The open interior of the sphere centered at **c** with radius ϵ.

Interior and Boundary Points

Definition (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.

Interior and Boundary Points

Definition (Interior Point)

The point $\mathbf{x}$ is an interior point of the set S if $\mathbf{x}$ is in S and if there is some ϵ-ball around $\mathbf{x}$ that contains only points in S. The **interior** of S is the collection of all interior points in S.
Example: The interior of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 < 4\}$.

Interior and Boundary Points

Definition (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.
Example: The interior of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 < 4\}$.

Definition (Boundary Point)

The point **x** is a boundary point of the set S if every ϵ-ball around **x** contains both points that are in S and points that are outside S. The **boundary** of S is the collection of all boundary points.

Interior and Boundary Points

Definition (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.
Example: The interior of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 < 4\}$.

Definition (Boundary Point)

The point **x** is a boundary point of the set S if every ϵ-ball around **x** contains both points that are in S and points that are outside S. The **boundary** of S is the collection of all boundary points.
Example: The boundary of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 = 4\}$.

Open and Closed Sets, Closure

Definition (Open Set)

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

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.
Example: $\{(x, y) : x^2 + y^2 \leq 4\}$

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.
Example: $\{(x, y) : x^2 + y^2 \leq 4\}$

Note: a set may be neither open nor closed.

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.
Example: $\{(x, y) : x^2 + y^2 \leq 4\}$

Note: a set may be neither open nor closed.
Example: $\{(x, y) : 2 < x^2 + y^2 \leq 4\}$

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.
Example: $\{(x, y) : x^2 + y^2 \leq 4\}$

Note: a set may be neither open nor closed.
Example: $\{(x, y) : 2 < x^2 + y^2 \leq 4\}$

Definition (Closure)

The **closure** of set S is the smallest closed set that contains S.

Open and Closed Sets, Closure

Definition (Open Set)

A set S is called **open** if for each point $\mathbf{x}$ in S, there exists an open ϵ-ball around $\mathbf{x}$ completely contained in S.
Example: $\{(x, y) : x^2 + y^2 < 4\}$

Definition (Closed Set)

A set S is called **closed** if it contains all of its boundary points.
Example: $\{(x, y) : x^2 + y^2 \leq 4\}$

Note: a set may be neither open nor closed.
Example: $\{(x, y) : 2 < x^2 + y^2 \leq 4\}$

Definition (Closure)

The **closure** of set S is the smallest closed set that contains S.
Example: The closure of $\{(x, y) : x^2 + y^2 < 4\}$ is
$\{(x, y) : x^2 + y^2 \leq 4\}$

Bounded Set

Sometimes the definition of a closed set is not sufficient. Consider the following case: the set $(-\infty, 0] \cup [1, \infty)$

Bounded Set

Sometimes the definition of a closed set is not sufficient. Consider the following case: the set $(-\infty, 0] \cup [1, \infty)$ is a closed set because its complement $(0, 1)$ is open. However, there is no upper bound to this set.

Bounded Set

Sometimes the definition of a closed set is not sufficient. Consider the following case: the set $(-\infty, 0] \cup [1, \infty)$ is a closed set because its complement $(0, 1)$ is open. However, there is no upper bound to this set.

Definition (Boundedness)

A set $A \subset \mathbb{R}^n$ is **bounded** if it can be contained within an ϵ-ball. That is, there will always be a real-valued number or vector that is outside the set.
Example: any interval that does not have ∞ or $-\infty$ as endpoints; any disk in a plane with finite radius.

Compact Set

Definition (Compact Set)

A set $A \subset \mathbb{R}^n$ is **compact** if it is closed and bounded.

Convexity

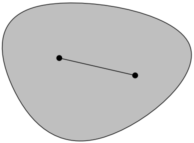
Definition (Convex Set)

A set A in $\mathbb{R}^n$ is said to be **convex** iff for each $x, y \in A$, the line segment $\lambda x + (1 - \lambda)y$ for $\lambda \in (0, 1)$ belongs to A. That is, all points on a line connecting two points in the set are in the set.

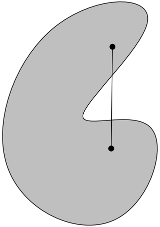
Convexity

Definition (Convex Set)

A set A in $\mathbb{R}^n$ is said to be **convex** iff for each $x, y \in A$, the line segment $\lambda x + (1 - \lambda)y$ for $\lambda \in (0, 1)$ belongs to A. That is, all points on a line connecting two points in the set are in the set.



set is convex



set is not convex

Why Bother with This?

These formal definitions are rather abstract and meaningless at first glance. However, they constitute some very important fundamentals, which ease the life of a scientist. Why is that?

Why Bother with This?

These formal definitions are rather abstract and meaningless at first glance. However, they constitute some very important fundamentals, which ease the life of a scientist. Why is that?

In many applications we can show that some results hold if a set is compact. For example, in game theory we know that (under certain very general assumptions about rationality of persons) amongst a set of possible choices there will always be some alternative which is preferred the most by a person if the set of choices is compact.

Why Bother with This?

These formal definitions are rather abstract and meaningless at first glance. However, they constitute some very important fundamentals, which ease the life of a scientist. Why is that?

In many applications we can show that some results hold if a set is compact. For example, in game theory we know that (under certain very general assumptions about rationality of persons) amongst a set of possible choices there will always be some alternative which is preferred the most by a person if the set of choices is compact.

In addition, if we know that this set is also convex, we then know that there will be exactly one most preferred alternative.

Why Bother with This?

These formal definitions are rather abstract and meaningless at first glance. However, they constitute some very important fundamentals, which ease the life of a scientist. Why is that?

In many applications we can show that some results hold if a set is compact. For example, in game theory we know that (under certain very general assumptions about rationality of persons) amongst a set of possible choices there will always be some alternative which is preferred the most by a person if the set of choices is compact.

In addition, if we know that this set is also convex, we then know that there will be exactly one most preferred alternative.

Beyond this example there are many other applications in political science that use the notion of compact sets.

Set Theory

Functions

What is a function?

Definition (Function)

A **function** or **map**, denoted by $f : X \mapsto Y$, has 3 parts:

- A set X to map from. This set is called the domain of f.
- A set Y to map to. This set is called the co-domain of f.
- A rule for every element $x \in X$, assigning it to some element $y \in Y$. This is written $f(x) = y$

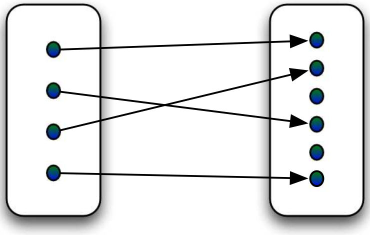
Examples:

- $$\begin{aligned} f &: \{1,2,3\} \to \{3,4,5\} \\ &: x \mapsto x+2 \end{aligned}$$
- $$\begin{aligned} &f : \{1,2\} \to \{1,3\} \\ &f(1) = 1, f(2) = 3 \end{aligned}$$

Linking Sets: Injection, Bijection, and Surjection

Definition (Injection)

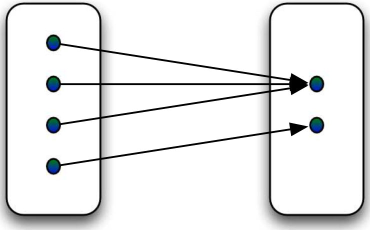
A function f is called **injective** if for every $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$. Verbally, every element of the codomain Y is linked to at most one element of the domain X.



Linking Sets: Injection, Bijection, and Surjection

Definition (Surjection)

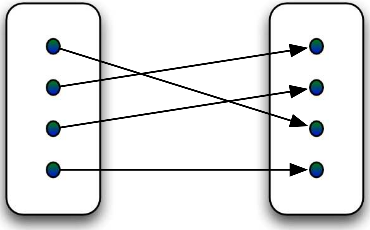
A function f is called **surjective** if for every $y \in Y$ there is an $x \in X$ with $f(x) = y$. Verbally, every element of the codomain Y is linked to at least one element of the domain X.



Linking Sets: Injection, Bijection, and Surjection

Definition (Bijection)

A function f is called **bijective** if it is injective and surjective, i.e. every element of the domain X is linked to one and only one element of the codomain Y and vice versa.



Set Theory

Binary Relations

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?
- How does a voting rule affect the outcome?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?
- How does a voting rule affect the outcome?
- Do certain voting rules provide incentives for individuals to misrepresent their preferences (to vote strategically)?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?
- How does a voting rule affect the outcome?
- Do certain voting rules provide incentives for individuals to misrepresent their preferences (to vote strategically)?
- How does agenda-setting affect the collective choice?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?
- How does a voting rule affect the outcome?
- Do certain voting rules provide incentives for individuals to misrepresent their preferences (to vote strategically)?
- How does agenda-setting affect the collective choice?
- What is this thing called 'democracy'?

Motivation

In positive political theory we are interested in individual and collective decision-making. One might ask:

- What must be true for a decision to be rational?
- How can individual decisions be aggregated?
- How does a voting rule affect the outcome?
- Do certain voting rules provide incentives for individuals to misrepresent their preferences (to vote strategically)?
- How does agenda-setting affect the collective choice?
- What is this thing called 'democracy'?

Binary relations are essential tools to formalize concepts like 'preferences' and 'choice'.

Definition

Definition (Binary Relation)

A binary relation R is a subset of $S \times S$ of **ordered pairs** of elements of S. R compares two elements of S, x and y, with each other. Write xRy.

Examples

1. Party Members
   $S = \{\text{CDU,SPD,Greens,FDP}\}$
   $R = \{(x, y) \in S \times S | \text{"Has more members"}\}$

Examples

1. Party Members
   $S = \{$CDU,SPD,Greens,FDP$\}$
   $R = \{(x, y) \in S \times S |$"Has more members"$\}$
   $R = \{$(SPD,CDU),(SPD, Greens),(SPD,FDP),(CDU,Greens), (CDU,FDP),(Greens,FDP)$\}$

Examples

1. Party Members
   $S = \{$CDU,SPD,Greens,FDP$\}$
   $R = \{(x, y) \in S \times S |$"Has more members"$\}$
   $R = \{$(SPD,CDU),(SPD, Greens),(SPD,FDP),(CDU,Greens),(CDU,FDP),(Greens,FDP)$\}$
2. Coalition Preferences
   $S = \{$(CDU,SPD),(CDU,FDP),(SPD,FDP)$\}$
   $R = \{(x, y) \in S \times S |$"FDP prefers"$\}$

Examples

1. Party Members
   $S = \{$CDU,SPD,Greens,FDP$\}$
   $R = \{(x, y) \in S \times S |$"Has more members"$\}$
   $R = \{$(SPD,CDU),(SPD, Greens),(SPD,FDP),(CDU,Greens),
   (CDU,FDP),(Greens,FDP)$\}$
2. Coalition Preferences
   $S = \{$(CDU,SPD),(CDU,FDP),(SPD,FDP)$\}$
   $R = \{(x, y) \in S \times S |$"FDP prefers"$\}$
   $R = \{$[(CDU,FDP),(SPD,FDP)],[(CDU,FDP),(CDU,SPD)],
   [(SPD,FDP),(CDU,SPD)]$\}$

Properties of Relations

Let R be a relation on S.

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$
- **Transitive:** $\forall x, y, z \in S, (xRyRz) \implies (xRz)$

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$
- **Transitive:** $\forall x, y, z \in S, (xRyRz) \implies (xRz)$
- **Complete:** $\forall x, y \in S, x \neq y,$ either xRy or yRx

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$
- **Transitive:** $\forall x, y, z \in S, (xRyRz) \implies (xRz)$
- **Complete:** $\forall x, y \in S, x \neq y,$ either xRy or yRx
- **Symmetric:** $\forall x, y \in S, (xRy) \implies (yRx)$

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$
- **Transitive:** $\forall x, y, z \in S, (xRyRz) \implies (xRz)$
- **Complete:** $\forall x, y \in S, x \neq y,$ either xRy or yRx
- **Symmetric:** $\forall x, y \in S, (xRy) \implies (yRx)$
- **Asymmetric:** $\forall x, y \in S, (xRy) \implies \neg(yRx)$

Properties of Relations

Let R be a relation on S.

- **Reflexive:** $\forall x \in S, xRx$
- **Transitive:** $\forall x, y, z \in S, (xRyRz) \implies (xRz)$
- **Complete:** $\forall x, y \in S, x \neq y$, either xRy or yRx
- **Symmetric:** $\forall x, y \in S, (xRy) \implies (yRx)$
- **Asymmetric:** $\forall x, y \in S, (xRy) \implies \neg(yRx)$
- **Antisymmetric:** $\forall x, y, \in S, [(xRy) \wedge (yRx)] \implies (x = y)$

Orders

- **Weak:** reflexive, transitive, complete

Orders

- **Weak:** reflexive, transitive, complete
- **Strict:** transitive, complete, antisymmetric

Orders

- **Weak:** reflexive, transitive, complete
- **Strict:** transitive, complete, antisymmetric
- **Weak partial:** reflexive, transitive, antisymmetric

Orders

- **Weak:** reflexive, transitive, complete
- **Strict:** transitive, complete, antisymmetric
- **Weak partial:** reflexive, transitive, antisymmetric
- **Strict partial:** transitive, asymmetric

Orders

- **Weak:** reflexive, transitive, complete
- **Strict:** transitive, complete, antisymmetric
- **Weak partial:** reflexive, transitive, antisymmetric
- **Strict partial:** transitive, asymmetric
- **Equivalence relation:** reflexive, transitive, symmetric

Preferences

Definition (Preferences)

A relation R is called a **preference relation** if and only if R is reflexive, transitive and complete (a weak order).

Maximal Elements

Definition (Maximal Elements)

Let R be a weak or partial order on Ω and $S \subseteq \Omega$. Then, the set of "R-maximal elements of S is"

$$M(S, R) = \{s \in S : \forall t \in S, sRt\}$$

Choice function

Definition (Choice function)

Let $\mathcal{X}$ be the family of all nonempty subsets of Ω. A choice function is a map $c\colon \mathcal{X} \to \mathcal{X}$ such that for all $S \in \mathcal{X}$, $c(S) \subseteq S$.

Defining Rationality

Definition (Rationality)

Given a choice function c, a choice is **rational** if and only if there $\exists$ a weak order R on Ω such that $\forall S \subseteq \Omega$, $c(S) = M(S, R)$. R is said to be a preference relation that **rationalizes** the choice function $c()$.

Defining Utility

Definition (Utility Function)

A **utility function** for an individual is a function that maps every element in S into the reals, $u : S \rightarrow \mathbb{R}$ such that $\forall a, b \in S, aRb \Leftrightarrow u(a) \geq u(b)$.

Analysis I

Rules for Exponentials and Fractions

- $x^0 = 1$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$
- $\frac{1}{x^a} = x^{-a}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$
- $\frac{1}{x^a} = x^{-a}$
- $\left(\frac{x}{y}\right)^a = \left(\frac{x^a}{y^a}\right) = x^a \cdot y^{-a}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$
- $\frac{1}{x^a} = x^{-a}$
- $\left(\frac{x}{y}\right)^a = \left(\frac{x^a}{y^a}\right) = x^a \cdot y^{-a}$
- $x^{\left(\frac{a}{b}\right)} = (x^a)^{\frac{1}{b}} = \sqrt[b]{x^a}$

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$
- $\frac{1}{x^a} = x^{-a}$
- $\left(\frac{x}{y}\right)^a = \left(\frac{x^a}{y^a}\right) = x^a \cdot y^{-a}$
- $x^{\left(\frac{a}{b}\right)} = (x^a)^{\frac{1}{b}} = \sqrt[b]{x^a}$
- For a^b we say "a raised to the b-th power," "a raised to the power/exponent (of) b," or more briefly "a to the b."

Rules for Exponentials and Fractions

- $x^0 = 1$
- $x^a = \underbrace{x \cdot x \cdot x \ldots \cdot x}_{a \text{ factors}}$
- $x^a \cdot x^b = x^{a+b}$
- $(x^a)^b = x^{a \cdot b}$
- $(xy)^a = x^a y^a$
- $\frac{1}{x^a} = x^{-a}$
- $\left(\frac{x}{y}\right)^a = \left(\frac{x^a}{y^a}\right) = x^a \cdot y^{-a}$
- $x^{\left(\frac{a}{b}\right)} = (x^a)^{\frac{1}{b}} = \sqrt[b]{x^a}$
- For a^b we say "a raised to the b-th power," "a raised to the power/exponent (of) b," or more briefly "a to the b."
- For $\frac{a}{b}$ we say "a divided by b," "a by b," or "a over b."

Binomial Theorem

- $(a+b)^2 = a^2 + 2ab + b^2$

Binomial Theorem

- $(a+b)^2 = a^2 + 2ab + b^2$
- $(a-b)^2 = a^2 - 2ab + b^2$

Binomial Theorem

- $(a+b)^2 = a^2 + 2ab + b^2$
- $(a-b)^2 = a^2 - 2ab + b^2$
- $(a+b)(a-b) = a^2 - b^2$

Binomial Theorem

- $(a+b)^2 = a^2 + 2ab + b^2$
- $(a-b)^2 = a^2 - 2ab + b^2$
- $(a+b)(a-b) = a^2 - b^2$
- and universally stated: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$; $n \in \mathbb{N}$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$
- $\log_a(xy) = \log_a(x) + \log_a(y)$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$
- $\log_a(xy) = \log_a(x) + \log_a(y)$
- $\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$
- $\log_a(xy) = \log_a(x) + \log_a(y)$
- $\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$
- $\log_a(x^y) = y \log_a(x)$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$
- $\log_a(xy) = \log_a(x) + \log_a(y)$
- $\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$
- $\log_a(x^y) = y \log_a(x)$
- $\log_a(a^x) = x$ and $a^{\log_a(x)} = x$

Logarithms

- For those with a German background: Please note that in English texts the expression *log* without specification of a base is equal to *ln*, i.e. the natural logarithm!
- $\log_a(1) = 0$
- $\log_a(xy) = \log_a(x) + \log_a(y)$
- $\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$
- $\log_a(x^y) = y \log_a(x)$
- $\log_a(a^x) = x$ and $a^{\log_a(x)} = x$
- Read $\log_a b$ as "the logarithm of b to the base a" or "the base-a logarithm of b"

Quadratic Expressions

Equations of the form $ax^2 + bx + c = 0$ can be solved using the quadratic formula (in German the so-called "Mitternachtsformel")

$$x_{1|2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Equations with one variable

Assume that we want to solve the following equation for x.

$$2\sqrt{x} - 3 = 1$$

Equations with one variable

Assume that we want to solve the following equation for x.

$$\begin{aligned} 2\sqrt{x} - 3 &= 1 \quad \big| + 3 \quad \text{we can add } \ldots \\ 2\sqrt{x} &= 4 \end{aligned}$$

Equations with one variable

Assume that we want to solve the following equation for x.

$$\begin{aligned} 2\sqrt{x} - 3 &= 1 \quad | + 3 \quad \text{we can add } \ldots \\ 2\sqrt{x} &= 4 \quad | : 2 \quad \ldots\text{divide}\ldots \\ \sqrt{x} &= 2 \end{aligned}$$

Equations with one variable

Assume that we want to solve the following equation for x.

$$\begin{aligned} 2\sqrt{x} - 3 &= 1 \quad | + 3 & \text{we can add } \ldots \\ 2\sqrt{x} &= 4 \quad | : 2 & \ldots\text{divide}\ldots \\ \sqrt{x} &= 2 \quad | a^2 & \ldots\text{raise to the power...} \\ x &= 4 \end{aligned}$$

Equations with one variable

Assume that we want to solve the following equation for x.

$$\begin{aligned} 2\sqrt{x} - 3 &= 1 \quad | + 3 && \text{we can add } \ldots \\ 2\sqrt{x} &= 4 \quad | : 2 && \ldots\text{divide}\ldots \\ \sqrt{x} &= 2 \quad | a^2 && \ldots\text{raise to the power...} \\ x &= 4 && \ldots\text{and much more} \end{aligned}$$

Equations with several variables

In political science applications solving for one variable oftentimes is not enough. So let us now consider the solution of two simultaneous equations with two variables.

$$\begin{aligned} 2x + 3y &= 4 \quad (1) \\ x - 2y &= 5 \quad (2) \end{aligned}$$

Equations with several variables

In political science applications solving for one variable oftentimes is not enough. So let us now consider the solution of two simultaneous equations with two variables.

$$\begin{aligned} 2x + 3y &= 4 \quad (1) \\ x - 2y &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x

$$x = 2y + 5 \quad (2)'$$

Equations with several variables

In political science applications solving for one variable oftentimes is not enough. So let us now consider the solution of two simultaneous equations with two variables.

$$\begin{aligned} 2x + 3y &= 4 \quad (1) \\ x - 2y &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x and insert this into (1):

$$\begin{aligned} x &= 2y + 5 && \text{(2)'} \\ 4y + 10 + 3y &= 4 && \text{(2)' in (1)} \end{aligned}$$

Equations with several variables

In political science applications solving for one variable oftentimes is not enough. So let us now consider the solution of two simultaneous equations with two variables.

$$\begin{aligned} 2x + 3y &= 4 \quad (1) \\ x - 2y &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x and insert this into (1):

$$\begin{aligned} x &= 2y + 5 && (2)' \\ 4y + 10 + 3y &= 4 && (2)' \text{ in } (1) \end{aligned}$$

This gives $y = -\frac{6}{7}$.

Equations with several variables

In political science applications solving for one variable oftentimes is not enough. So let us now consider the solution of two simultaneous equations with two variables.

$$\begin{aligned} 2x + 3y &= 4 \quad (1) \\ x - 2y &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x and insert this into (1):

$$\begin{aligned} x &= 2y + 5 \quad && (2)' \\ 4y + 10 + 3y &= 4 \quad && (2)' \text{ in } (1) \end{aligned}$$

This gives $y = -\frac{6}{7}$. Inserting this into (2)' gives $x = \frac{23}{7}$.

Analysis I

Derivatives

Motivation

- What is the relationship between the level of democracy and economic growth?

Motivation

- What is the relationship between the level of democracy and economic growth?
- for linear relationships, the information is directly available from the equation - the slope m

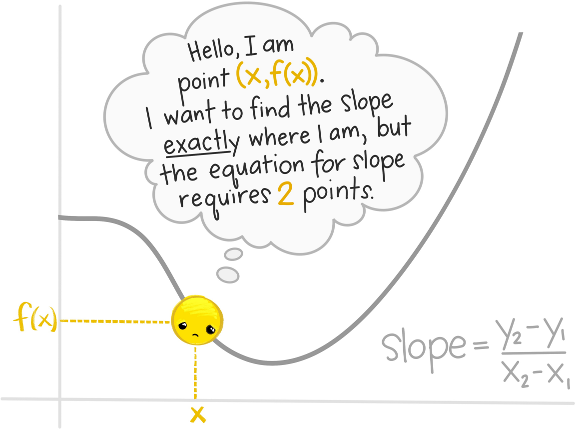
Motivation

- What is the relationship between the level of democracy and economic growth?
- for linear relationships, the information is directly available from the equation - the slope m
- What do we do when we have a non-linear function?

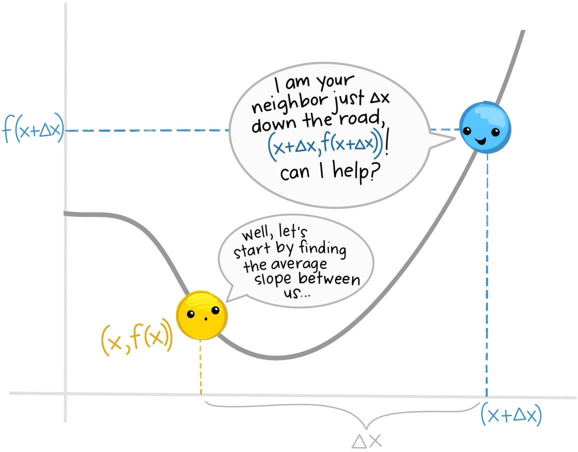
Motivation

- What is the relationship between the level of democracy and economic growth?
- for linear relationships, the information is directly available from the equation - the slope m
- What do we do when we have a non-linear function?
- What is the slope m at some point x_0?

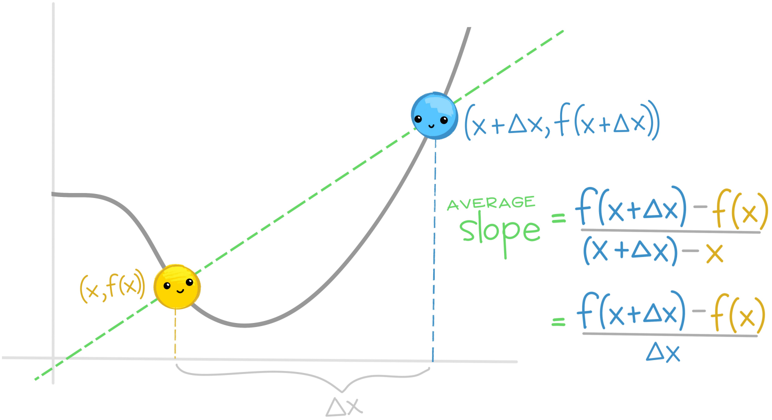
What is a derivative? I



What is a derivative? II



What is a derivative? III

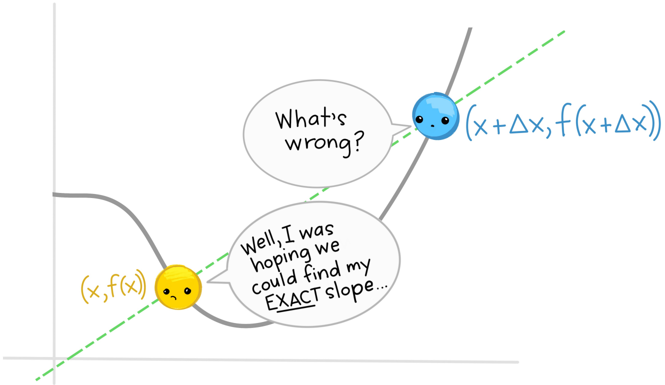


What is a derivative? IV

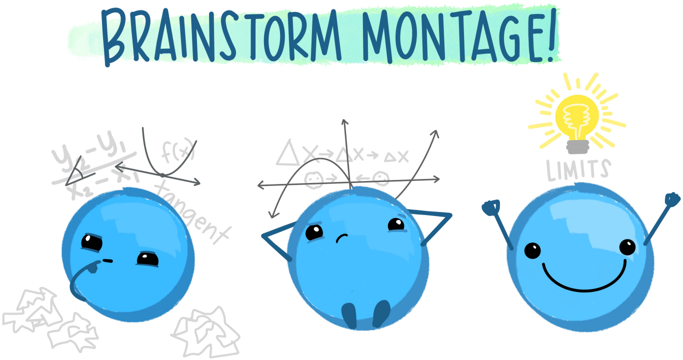
So: the average slope between ANY 2 POINTS on function $f(x)$ separated by Δx is

$$m = \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

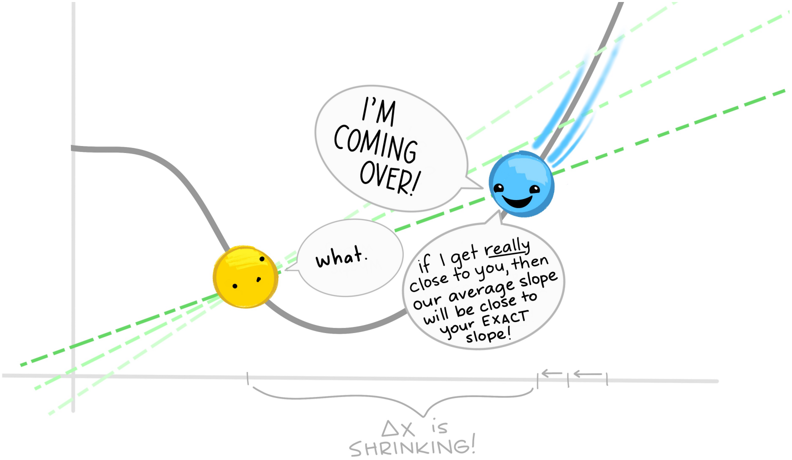
What is a derivative? V



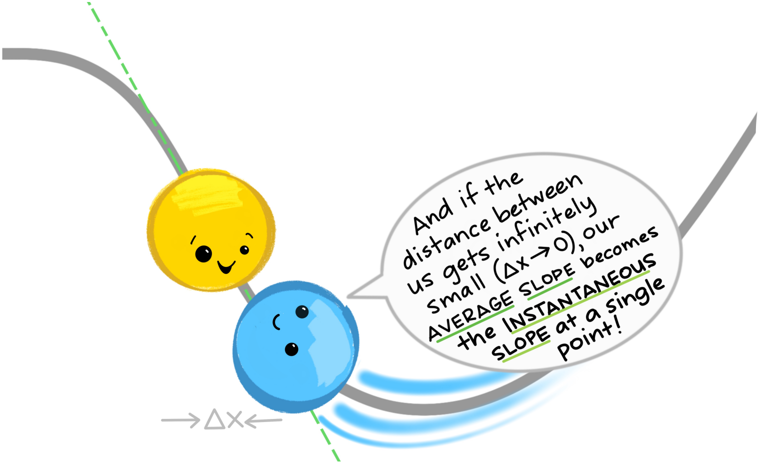
What is a derivative? VI



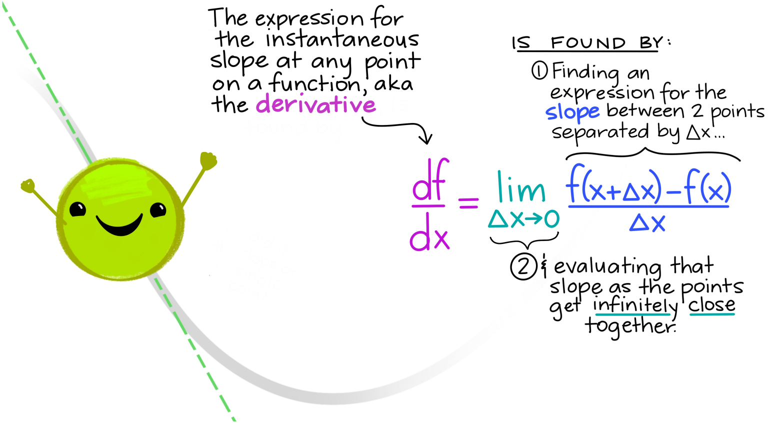
What is a derivative? VII



What is a derivative? VIII



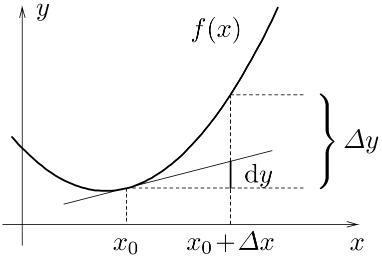
What is a derivative? IX



What is a derivative? X

We want to estimate the slope of a function at point x_0.

- As a rough estimate we can form the difference quotient $\frac{\Delta y}{\Delta x}$.
- Decreasing Δx continuously brings us closer and closer to the true slope...
- In limit we approach the **derivative** at point x_0.



Illustrations by Allison Horst

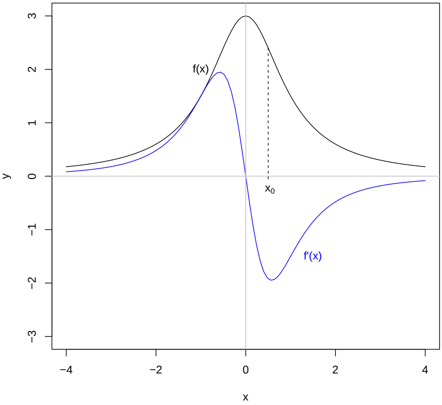
Intuition I

The derivative:

- is a measure of how a function changes as its input changes
- of a function at a chosen input value describes the best linear approximation of the function near that input value
- at a point equals the slope of the tangent line to the graph of the function at that point (linearization of a function for the multivariate case)

Intuition II

- $f(x) = \frac{3}{1+x^2}$
- $f'(x) = -\frac{6x}{(x^2+1)^2}$
- Observations:
  - slope is not a number anymore, but a function (it varies with x)
  - for any x, $f'(x)$ gives us the slope (a value)
  - e.g. $f'(x_0 = 0.5) =$ -1.92



Definition

Definition (Limit of a Function)

Assuming $x, p, c, L \in \mathbb{R}$, the limit of a real valued function f when x approaches p, denoted as $\lim_{x \to p} f(x) = L$, is L if $\forall \epsilon > 0 \exists c > 0, s.t. \forall x, 0 < |x - p| < c \implies |f(x) - L| < \epsilon$.

Note, that if $p = +\infty$ or $p = -\infty$, L is called the asymptote of the function.

Definition

Definition (Derivative)

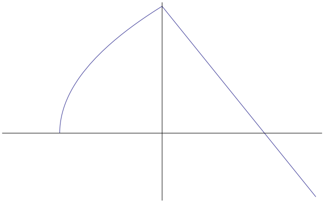
Let $(x_0, f(x_0))$ be a point on the graph of $y = f(x)$. The **derivative** of f at x_0, written $f'(x_0)$, $\frac{df}{dx}(x_0)$, $\frac{dy}{dx}(x_0)$ is the slope of the tangent line to the graph of f at $(x_0, f(x_0))$:

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

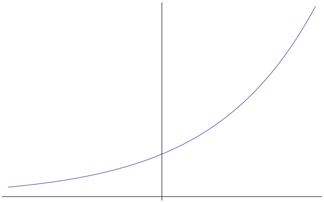
if this limit exists. If this limit exists for every point x in the domain of f, the function is differentiable.

Differentiability

- graph has to be 'smooth' (no gaps, holes, ...)
- if f is differentiable, it must be continuos (converse does not hold)



function is not differentiable



function is differentiable

Continuity

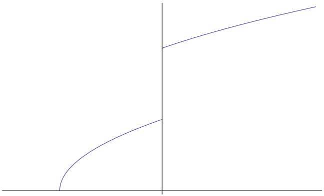
Definition (Continuity)

A function f is **continuous** at $x = a$ if $\lim_{x \to a} f(x) = f(a)$

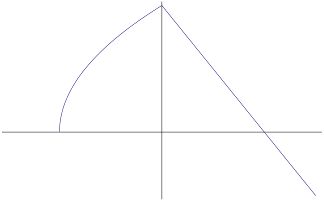
Continuity

Definition (Continuity)

A function f is **continuous** at $x = a$ if $\lim_{x \to a} f(x) = f(a)$

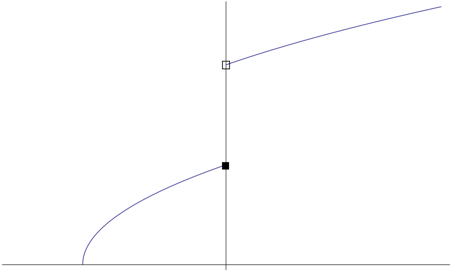


function is discontinuous

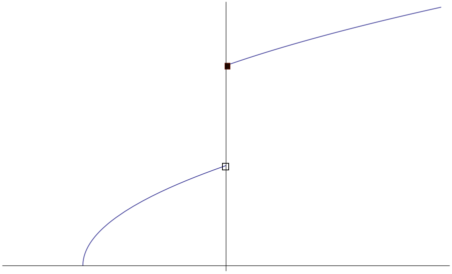


function is continuous

Semi-Continuity



function is lower
(semi-)continuous



function is upper
(semi-)continuous

Analysis I

Rules of Differentiation

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$
- $f(x) = ln(x)$, then $f'(x) = \frac{1}{x}$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$
- $f(x) = ln(x)$, then $f'(x) = \frac{1}{x}$
- $f(x) = log_a x$, then $f'(x) = \frac{1}{x \ln a}$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$
- $f(x) = ln(x)$, then $f'(x) = \frac{1}{x}$
- $f(x) = log_a x$, then $f'(x) = \frac{1}{x \ln a}$
- $f(x) = e^{ax}$, then $f'(x) = ae^{ax}$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$
- $f(x) = ln(x)$, then $f'(x) = \frac{1}{x}$
- $f(x) = log_a x$, then $f'(x) = \frac{1}{x \ln a}$
- $f(x) = e^{ax}$, then $f'(x) = ae^{ax}$
- $f(x) = a$, where a is a constant, e.g. 1, then $f'(x) = 0$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a$, then $f'(x) = ax^{a-1}$
- $f(x) = ln(x)$, then $f'(x) = \frac{1}{x}$
- $f(x) = log_a x$, then $f'(x) = \frac{1}{x \ln a}$
- $f(x) = e^{ax}$, then $f'(x) = ae^{ax}$
- $f(x) = a$, where a is a constant, e.g. 1, then $f'(x) = 0$
- $f(x) = a^x$, then $f'(x) = \log_a a^x$

Rules of Differentiation I

Rules for Common Functions

- $f(x) = x^a,$ then $f'(x) = ax^{a-1}$
- $f(x) = ln(x),$ then $f'(x) = \frac{1}{x}$
- $f(x) = log_a x,$ then $f'(x) = \frac{1}{x \ln a}$
- $f(x) = e^{ax},$ then $f'(x) = ae^{ax}$
- $f(x) = a$, where a is a constant, e.g. 1, then $f'(x) = 0$
- $f(x) = a^x,$ then $f'(x) = \log_a a^x$
- $f(x) = \frac{1}{x} = x^{-1},$ then $f'(x) = -\frac{1}{x^2}$

Rules of Differentiation II

Sum Rule

- $[f(x)+g(x)]' = f'(x)+g'(x)$
- Example:

$$\begin{aligned} h(x) &= 2x + x^2 \\ h'(x) &= 2 + 2x \end{aligned}$$

Rules of Differentiation II

Product Rule

- $[f(x) \cdot g(x)]' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$
- Example:

$$
\begin{aligned}
h(x) &= 2x \cdot \sqrt{x} \\
h'(x) &= 2 \cdot \sqrt{x} + 2x \cdot \frac{1}{2\sqrt{x}}
\end{aligned}
$$

Rules of Differentiation III

Quotient Rule

- $\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{(g(x))^2}$
- Example:

$$h(x) = \frac{3x}{2 - x^2}$$

$$h'(x) = \frac{3 \cdot (2 - x^2) - 3x \cdot (-2x)}{(2 - x^2)^2}$$

Rules of Differentiation III

Chain Rule

- $[f(g(x))]' = f'(g(x)) \cdot g'(x)$
- Example:

$$\begin{aligned} h(x) &= (5x-2)^3 \\ h'(x) &= 3(5x-2)^2 \cdot 5 \end{aligned}$$

Analysis I

Partial Derivatives

Motivation

- What if the relationship between the level of democracy does not only dependent on economic growth, but also on the political institutions?

Motivation

- What if the relationship between the level of democracy does not only dependent on economic growth, but also on the political institutions?
- We can generalize the concept of a derivative to the multivariate case

Motivation

- What if the relationship between the level of democracy does not only dependent on economic growth, but also on the political institutions?
- We can generalize the concept of a derivative to the multivariate case
- Partial derivates say something about the changes in y given a change in x_i holding all other arguments at some level

Partial Derivatives I

Definition (Partial Derivatives)

Let f be a multivariate function. Then for each variable x_i at each set of points $(x_1^0, ..., x_n^0)$ in the domain of f:

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

is called the partial derivative, if the limit exists.

Note, that we usually write $\frac{\partial f}{\partial x}$ for partial derivatives and $\frac{df}{dy}$ for derivatives.

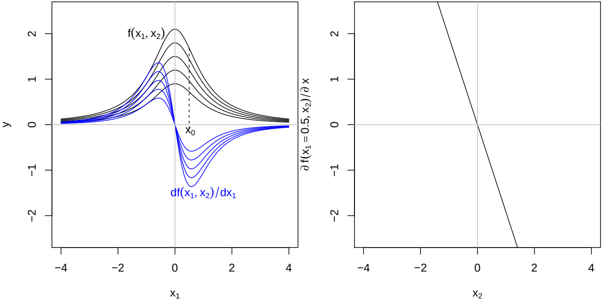
Partial Derivatives II

Example:

$$
\begin{aligned}
f(x_1, x_2) &= x_1^2 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_1} &= 2x_1 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_2} &= x_1^2 \cdot \frac{1}{x_2}
\end{aligned}
$$

Intuition

- $f(x_1, x_2) = \frac{3x_2}{1+x_1^2}$
- $\frac{\partial f(x_1,x_2)}{\partial x_1} = \frac{-6x_1x_2}{(x_1^2+1)^2}$
- Observations:
  - slope varies not only with x_1, but also with x_2
  - e.g. $\frac{\partial f(x_1=0.5,x_2)}{\partial x_1} = -1.92x_2$



Second-order Partial Derivatives

Reconsider the example from the last slide

$$
\begin{aligned}
f(x_1, x_2) &= x_1^2 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_1} &= 2x_1 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_2} &= x_1^2 \cdot \frac{1}{x_2}
\end{aligned}
$$

Second-order Partial Derivatives

Reconsider the example from the last slide

$$\begin{aligned} f(x_1, x_2) &= x_1^2 \cdot \ln x_2 \\ \frac{\partial f}{\partial x_1} &= 2x_1 \cdot \ln x_2 \\ \frac{\partial f}{\partial x_2} &= x_1^2 \cdot \frac{1}{x_2} \\ \frac{\partial^2 f}{\partial x_1^2} &= 2 \cdot \ln x_2 \end{aligned}$$

Second-order Partial Derivatives

Reconsider the example from the last slide

$$
\begin{aligned}
f(x_1, x_2) &= x_1^2 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_1} &= 2x_1 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_2} &= x_1^2 \cdot \frac{1}{x_2} \\
\frac{\partial^2 f}{\partial x_1^2} &= 2 \cdot \ln x_2 \\
\frac{\partial^2 f}{\partial x_2^2} &= -x_1^2 \cdot \frac{1}{x_2^2}
\end{aligned}
$$

Second-order derivatives describe how the slope of the first derivative changes given changes in x.

Mixed Partial Derivatives I

Reconsider the example from the last slide

$$
\begin{aligned}
f(x_1, x_2) &= x_1^2 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_1} &= 2x_1 \cdot \ln x_2 \\
\frac{\partial f}{\partial x_2} &= x_1^2 \cdot \frac{1}{x_2} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} &= 2x_1 \cdot \frac{1}{x_2}
\end{aligned}
$$

Mixed Partial Derivatives II

Theorem (Young's Theorem)

Suppose that all the m^{th}-order partial derivatives of the function $f(x_1, x_2, \ldots, x_n)$ are continuous. If any of them involve differentiating with respect to each of the variables the same number of times, then they are necessarily equal.

In the case of $f(x_1, x_2)$, that implies for example:

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} \equiv \frac{\partial^2 f}{\partial x_2 \partial x_1}$$

Hessian Matrix I

Because of the importance of the second-order partial derivatives for constrained optimization there does exist a special of collecting them, the so-called **Hessian Matrix**

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{pmatrix}$$

Application

- Estimation of covariance matrix
- Optimization in maximum likelihood
- ...

Analysis II

Analysis II

Optimization

Motivation for Optimization

In decision theory we are interested in the decision-making process of an individual.

Motivation for Optimization

In decision theory we are interested in the decision-making process of an individual.

Let us assume, we have a specified utility function of a person $u(x) = -(x + \sqrt{a})^2$.

Motivation for Optimization

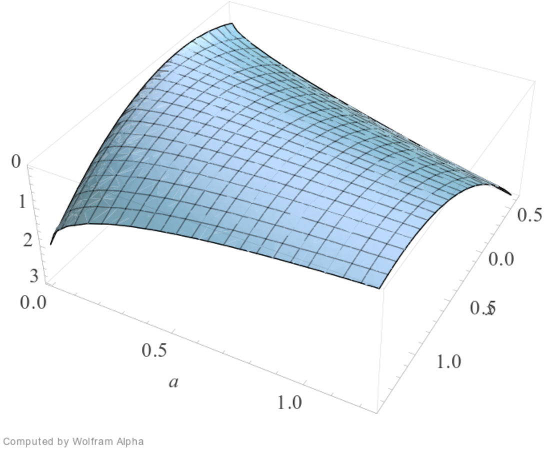
In decision theory we are interested in the decision-making process of an individual.

Let us assume, we have a specified utility function of a person $u(x) = -(x + \sqrt{a})^2$.

We want to know the optimal choice the person can take. How do we do this?

Motivation for Optimization

$$u(x) = -(x + \sqrt{a})^2.$$



Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

$$\frac{df}{dx} \equiv 0$$

Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

$$\frac{df}{dx} \equiv 0$$

- We derive the first line because we know that in an extreme point the slope of the function (i.e. its first derivative) equals zero.

Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

$$\frac{df}{dx} \equiv 0$$

- We derive the first line because we know that in an extreme point the slope of the function (i.e. its first derivative) equals zero.
- In our case $\frac{df}{dx} = -2x - 2\sqrt{a}$.

Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

$$\frac{df}{dx} \equiv 0$$

- We derive the first line because we know that in an extreme point the slope of the function (i.e. its first derivative) equals zero.
- In our case $\frac{df}{dx} = -2x - 2\sqrt{a}$.
- Solving the equality gives us $x^\star = -\sqrt{a}$.

Single Variable Optimization - FOC

The first step to get an answer to this problem is to search for the so-called **first-order condition (FOC)**:

$$\frac{df}{dx} \equiv 0$$

- We derive the first line because we know that in an extreme point the slope of the function (i.e. its first derivative) equals zero.
- In our case $\frac{df}{dx} = -2x - 2\sqrt{a}$.
- Solving the equality gives us $x^\star = -\sqrt{a}$.
- So now we know that at this point the function either has a (local) maximum/minimum (or a saddle point).

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

- Local maximum if $\frac{d^2f}{dx^2}(x^\star) < 0$, i.e. the function is concave

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

- Local maximum if $\frac{d^2 f}{dx^2}(x^\star) < 0$, i.e. the function is concave
- Local minimum if $\frac{d^2 f}{dx^2}(x^\star) > 0$, i.e. the function is convex

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

- Local maximum if $\frac{d^2f}{dx^2}(x^\star) < 0$, i.e. the function is concave
- Local minimum if $\frac{d^2f}{dx^2}(x^\star) > 0$, i.e. the function is convex
- Saddle point if $\frac{d^2f}{dx^2}(x^\star) = 0$ and $\frac{d^3f}{dx^3} \neq 0$.

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

- Local maximum if $\frac{d^2 f}{dx^2}(x^\star) < 0$, i.e. the function is concave
- Local minimum if $\frac{d^2 f}{dx^2}(x^\star) > 0$, i.e. the function is convex
- Saddle point if $\frac{d^2 f}{dx^2}(x^\star) = 0$ and $\frac{d^3 f}{dx^3} \neq 0$.

In our example $\frac{d^2 f}{dx^2}(-\sqrt{a}) = -2$. So we have a local maximum.

Single Variable Optimization - SOC

Now we need to specify which of the three possibilities applies. We do this by checking the **second-order condition**.

- Local maximum if $\frac{d^2f}{dx^2}(x^\star) < 0$, i.e. the function is concave
- Local minimum if $\frac{d^2f}{dx^2}(x^\star) > 0$, i.e. the function is convex
- Saddle point if $\frac{d^2f}{dx^2}(x^\star) = 0$ and $\frac{d^3f}{dx^3} \neq 0$.

In our example $\frac{d^2f}{dx^2}(-\sqrt{a}) = -2$. So we have a local maximum.

Controlling for the other parts of the function, we find that this is also a global maximum.

Convex, Concave, and Inflection Point

- A function is called **convex** if $\frac{d^2f}{dx^2} \geq 0$.

Convex, Concave, and Inflection Point

- A function is called **convex** if $\frac{d^2f}{dx^2} \geq 0$.
- A function is called **concave** if $\frac{d^2f}{dx^2} \leq 0$.

Convex, Concave, and Inflection Point

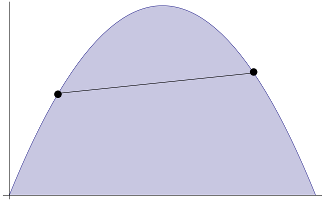
- A function is called **convex** if $\frac{d^2f}{dx^2} \geq 0$.
- A function is called **concave** if $\frac{d^2f}{dx^2} \leq 0$.
- A point a as called **inflection point** if $\frac{d^2f}{dx^2} = 0$ and $\frac{d^2f}{dx^2}$ changes sign at a.

Convex, Concave, and Inflection Point

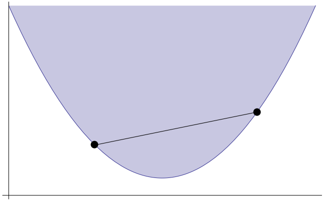
- A function is called **convex** if $\frac{d^2 f}{dx^2} \geq 0$.
- A function is called **concave** if $\frac{d^2 f}{dx^2} \leq 0$.
- A point a as called **inflection point** if $\frac{d^2 f}{dx^2} = 0$ and $\frac{d^2 f}{dx^2}$ changes sign at a.
- If a is an inflection point and $\frac{df}{dx} = 0$, then it is a **saddle point**.

More General Definition of Concavity/Convexity

A function is called concave (convex) if the line segment joining any two points on the graph is below (above) the graph, or on the graph.



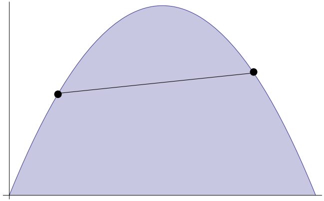
concave



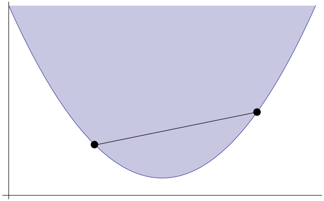
convex

More General Definition of Concavity/Convexity

A function is called concave (convex) if the line segment joining any two points on the graph is below (above) the graph, or on the graph.



concave

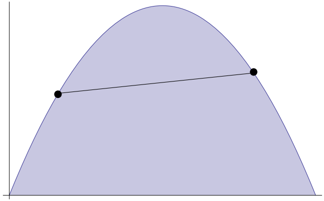


convex

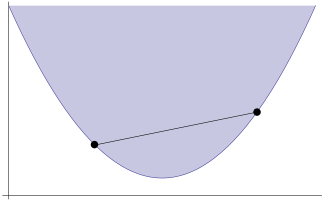
We can derive the concavity/convexity of functions from the concept of convex sets.

More General Definition of Concavity/Convexity

A function is called concave (convex) if the line segment joining any two points on the graph is below (above) the graph, or on the graph.



concave

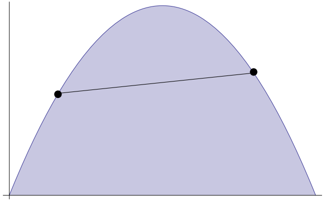


convex

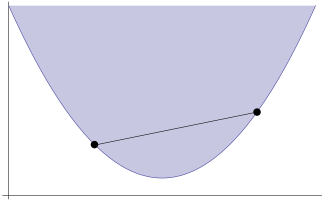
We can derive the concavity/convexity of functions from the concept of convex sets. A function is called **convex** if the set of all points which are on or above its graph is a convex set.

More General Definition of Concavity/Convexity

A function is called concave (convex) if the line segment joining any two points on the graph is below (above) the graph, or on the graph.



concave



convex

We can derive the concavity/convexity of functions from the concept of convex sets. A function is called **convex** if the set of all points which are on or above its graph is a convex set. Conversely, a function is called **concave** if the set of all points which are on or below its graph is a convex set.

Bivariate Optimization I

Consider a C^2 function (i.e. a function that is both continuous and twice differentiable) $f(x, y)$ in a convex set S.

Bivariate Optimization I

Consider a C^2 function (i.e. a function that is both continuous and twice differentiable) $f(x, y)$ in a convex set S.

Fist-order condition

- Find the first-order partial derivatives and equate them to zero.
- Solve the two-equation system for the values of x and y.
- $(x^\star, y^\star)$ is the stationary point.

Bivariate Optimization I

Consider a C^2 function (i.e. a function that is both continuous and twice differentiable) $f(x, y)$ in a convex set S.

Fist-order condition

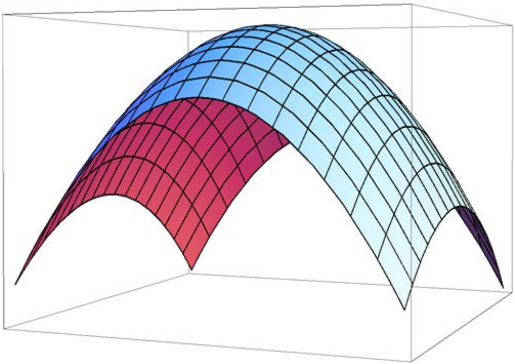
- Find the first-order partial derivatives and equate them to zero.
- Solve the two-equation system for the values of x and y.
- $(x^\star, y^\star)$ is the stationary point.

Second-order condition

- If for all (x, y) in S, $\frac{\partial^2 f}{\partial x^2} \leq 0$, $\frac{\partial^2 f}{\partial y^2} \leq 0$, and $\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \geq 0$ then $(x^\star, y^\star)$ is a maximum point for $f(x, y)$ in S.
- If for all (x, y) in S, $\frac{\partial^2 f}{\partial x^2} \geq 0$, $\frac{\partial^2 f}{\partial y^2} \geq 0$, and $\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \leq 0$ then $(x^\star, y^\star)$ is a minimum point for $f(x, y)$ in S.

Bivariate Optimization II

Consider the function $f(x, y) = -0.5(x - 1)^2 - y^2$.



Bivariate Optimization III

Function $f(x, y) = -0.5(x - 1)^2 - y^2$.

Bivariate Optimization III

Function $f(x, y) = -0.5(x-1)^2 - y^2$.

The first order condition

$$\begin{aligned} \frac{\partial f}{\partial x} &= -x + 1 \equiv 0 \\ \frac{\partial f}{\partial y} &= -2y \equiv 0 \end{aligned}$$

gives us a stationary point at $x = 1, y = 0$.

Bivariate Optimization IV

The second order condition

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= -1 < 0 \\ \frac{\partial^2 f}{\partial y^2} &= -2 < 0 \\ \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 &= (-1) \cdot (-2) - 0 \geq 0\end{aligned}$$

tells us that we have a maximum at $x = 1, y = 0$.

Extreme Value Theorem/Weierstrass Theorem

Theorem (Extreme Value Theorem/Weierstrass Theorem)

Suppose the function $f(\mathbf{x})$ *is continuous throughout a nonempty, closed and bounded set* S *in* $\mathbb{R}^n$*. Then there exists a point* $\mathbf{d}$ *in* S *where* f *has a minimum and a point* $\mathbf{c}$ *in* S *where* f *has a maximum. That is,*

$$f(\mathbf{d}) \leq f(\mathbf{x}) \leq f(\mathbf{c}) \text{ for all } \mathbf{x} \in S.$$

Extreme Value Theorem/Weierstrass Theorem

Theorem (Extreme Value Theorem/Weierstrass Theorem)

Suppose the function $f(\mathbf{x})$ is continuous throughout a nonempty, closed and bounded set S in $\mathbb{R}^n$. Then there exists a point $\mathbf{d}$ in S where f has a minimum and a point $\mathbf{c}$ in S where f has a maximum. That is,

$$f(\mathbf{d}) \leq f(\mathbf{x}) \leq f(\mathbf{c}) \text{ for all } \mathbf{x} \in S.$$

You will find the Weierstrass Theorem on page 20 of McCarty and Meirowitz (2007).

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

- ...what happens to the likelihood of the outbreak of civil war if the ethnic diversity of the country increases.

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

- ...what happens to the likelihood of the outbreak of civil war if the ethnic diversity of the country increases.
- ...how the level of voter turnout changes as party polarization changes.

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

- ...what happens to the likelihood of the outbreak of civil war if the ethnic diversity of the country increases.
- ...how the level of voter turnout changes as party polarization changes.
- ...how party cohesiveness changes as the level of electoral competitiveness changes?

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

- ...what happens to the likelihood of the outbreak of civil war if the ethnic diversity of the country increases.
- ...how the level of voter turnout changes as party polarization changes.
- ...how party cohesiveness changes as the level of electoral competitiveness changes?
- ...government public goods provision changes as the size of the winning coalition changes?

Comparative Statics I

Testable predictions of formal models are typically based on **comparative statics**. For example, a researcher might ask...

- ...what happens to the likelihood of the outbreak of civil war if the ethnic diversity of the country increases.
- ...how the level of voter turnout changes as party polarization changes.
- ...how party cohesiveness changes as the level of electoral competitiveness changes?
- ...government public goods provision changes as the size of the winning coalition changes?

More generally: How do changes in the parameters of a model affect the model's solution?

Comparative Statics II

Recall the optimal choice $x^\star = -\sqrt{a}$ of the person with the utility function $u(x) = -(x+a)^2$. How does the optimal choice change as the value of a changes?

Comparative Statics II

Recall the optimal choice $x^\star = -\sqrt{a}$ of the person with the utility function $u(x) = -(x+a)^2$. How does the optimal choice change as the value of a changes?

$$\frac{dx^\star}{da} = \frac{1}{2\sqrt{a}}$$

Comparative Statics II

Recall the optimal choice $x^\star = -\sqrt{a}$ of the person with the utility function $u(x) = -(x+a)^2$. How does the optimal choice change as the value of a changes?

$$\frac{dx^\star}{da} = \frac{1}{2\sqrt{a}}$$

An increase of one unit a increases $u(x)$ by $\frac{1}{2\sqrt{a}}$ units, **ceteris paribus**.

Optimization Under Constraints - Problem

So far we have considered decision problems in general. But what about situations in which an agent has to make her decision under given constraints?

Optimization Under Constraints - Problem

So far we have considered decision problems in general. But what about situations in which an agent has to make her decision under given constraints?

Let us consider the following example:
We as a city can decide to allocate our budget between cultural (c) and social (s) affairs. The overall utility function of our city is given by $f(x) = \frac{1}{2}s^2 + (c - \frac{1}{3})^2$. Our budget is constrained as $c + s = 2$.

Optimization Under Constraints - Problem

So far we have considered decision problems in general. But what about situations in which an agent has to make her decision under given constraints?

Let us consider the following example:
We as a city can decide to allocate our budget between cultural (c) and social (s) affairs. The overall utility function of our city is given by $f(x) = \frac{1}{2}s^2 + (c - \frac{1}{3})^2$. Our budget is constrained as $c + s = 2$.

A method to solve such problems is the so-called **Lagrangian multiplier method**.

Lagrangian Multiplier Method I

In order to solve the maximization problem $\max f(x, y)$ subject to $g(x, y) = c$ we proceed the following way.

Lagrangian Multiplier Method I

In order to solve the maximization problem $\max f(x, y)$ subject to $g(x, y) = c$ we proceed the following way.

1. Write down the Lagrangian $\mathcal{L}(x, y) = f(x, y) - \lambda (g(x, y) - c)$, where λ is a constant.

Lagrangian Multiplier Method I

In order to solve the maximization problem $\max f(x, y)$ subject to $g(x, y) = c$ we proceed the following way.

1. Write down the Lagrangian $\mathcal{L}(x, y) = f(x, y) - \lambda\,(g(x, y) - c)$, where λ is a constant.
2. Differentiate $\mathcal{L}$ with respect to x and y, and equate the partial derivatives to 0.

Lagrangian Multiplier Method I

In order to solve the maximization problem $\max f(x,y)$ subject to $g(x,y) = c$ we proceed the following way.

1. Write down the Lagrangian $\mathcal{L}(x,y) = f(x,y) - \lambda(g(x,y) - c)$, where λ is a constant.
2. Differentiate $\mathcal{L}$ with respect to x and y, and equate the partial derivatives to 0.
3. Solve the system of equations that the two partials form together with the constraint.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} \equiv 0 \\
\frac{\partial \mathcal{L}}{\partial y} &= \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} \equiv 0 \\
g(x,y) &= c
\end{aligned}$$

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

The system of equations

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

The system of equations

$$\frac{\partial \mathcal{L}}{\partial s} = s - \lambda \equiv 0$$

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

The system of equations

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial s} &= s - \lambda \equiv 0 \\
\frac{\partial \mathcal{L}}{\partial c} &= 2c - \frac{2}{3} - \lambda \equiv 0
\end{aligned}$$

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

The system of equations

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial s} &= s - \lambda \equiv 0 \\
\frac{\partial \mathcal{L}}{\partial c} &= 2c - \frac{2}{3} - \lambda \equiv 0 \\
s + c &= 2
\end{aligned}$$

Application to our problem

The Lagrangian

$$\mathcal{L}(s, c, \lambda) = \frac{1}{2}s^2 + \left(c - \frac{1}{3}\right)^2 - \lambda(s + c - 2)$$

The system of equations

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= s - \lambda \equiv 0 \\ \frac{\partial \mathcal{L}}{\partial c} &= 2c - \frac{2}{3} - \lambda \equiv 0 \\ s + c &= 2 \end{aligned}$$

If we solve the system of equations, we get $c = \frac{8}{9}$ and $s = \frac{10}{9}$.

Lagrangian Multiplier Method II

If we compare the Lagrangian method for constrained optimization to the unconstrained optimization, is still something missing?

Lagrangian Multiplier Method II

If we compare the Lagrangian method for constrained optimization to the unconstrained optimization, is still something missing?

Yes, theoretically we have to check for the second order-condition.

Lagrangian Multiplier Method II

If we compare the Lagrangian method for constrained optimization to the unconstrained optimization, is still something missing?

Yes, theoretically we have to check for the second order-condition.

You find the formulation in Sysdsæter/Hammond (2008) on pp. 506-507.

Advanced Constrained Optimization

There is much more to constrained optimization!

- Multivariable optimization (we need matrix algebra for that!).

Advanced Constrained Optimization

There is much more to constrained optimization!

- Multivariable optimization (we need matrix algebra for that!).
- Lagrangian for more than two variables.

Advanced Constrained Optimization

There is much more to constrained optimization!

- Multivariable optimization (we need matrix algebra for that!).
- Lagrangian for more than two variables.
- Lagrangian for more than one condition.

Advanced Constrained Optimization

There is much more to constrained optimization!

- Multivariable optimization (we need matrix algebra for that!).
- Lagrangian for more than two variables.
- Lagrangian for more than one condition.
- Optimization for inequalities
  $\max f(x, y)$ subject to $g(x, y) \leq c$
  "nonlinear programming" or "Kuhn-Tucker"

Advanced Constrained Optimization

There is much more to constrained optimization!

- Multivariable optimization (we need matrix algebra for that!).
- Lagrangian for more than two variables.
- Lagrangian for more than one condition.
- Optimization for inequalities
  $\max f(x, y)$ subject to $g(x, y) \leq c$
  "nonlinear programming" or "Kuhn-Tucker"

See Sysdsæter/Hammond (2008), Chapter 14.

Analysis II

Integration

Motivation I

- probability density functions (p.d.f) are fundamental to statistics

Motivation I

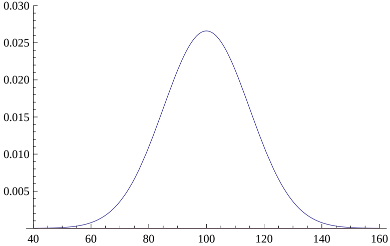
- probability density functions (p.d.f) are fundamental to statistics
- p.d.f. relate a particular event (x) to a probability (y)

Motivation I

- probability density functions (p.d.f) are fundamental to statistics
- p.d.f. relate a particular event (x) to a probability (y)
- when we are interested in calculating the probability for a range of events, we need to calculate the area under the curve

Motivation II

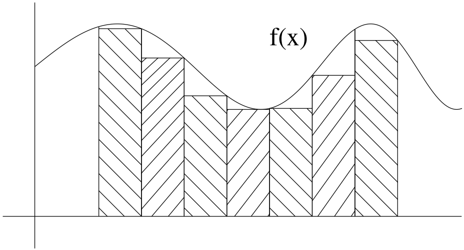
- We know that IQ test scores amongst people of the same age are distributed normally with mean 100 and standard deviation 15.
- What is the probability that a person has a score of more than 120?



It is the area below the normal p.d.f. for $x > 120$ (p$\approx$ 9.12%)

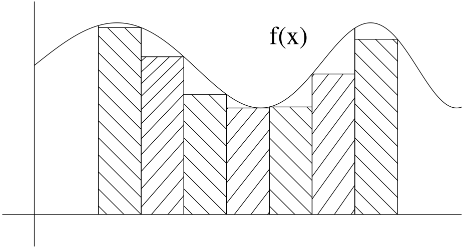
Intuition

- The **indefinite integral** $F(x)$ of a function $f(x)$ is the area between the function and the x-axis.



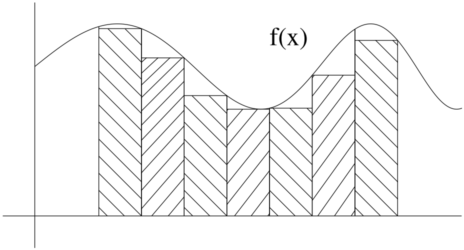
Intuition

- The **indefinite integral** $F(x)$ of a function $f(x)$ is the area between the function and the x-axis.
- We can think of this integral also as the sum of an infinite number of rectangles below the curve!



Intuition

- The **indefinite integral** $F(x)$ of a function $f(x)$ is the area between the function and the x-axis.
- We can think of this integral also as the sum of an infinite number of rectangles below the curve!
- Calculating an integral is the reverse process of taking a derivative. For this we sometimes refer to an integral as **antiderivative**.



Definition Integral

Definition (Riemann Integral)

Let f be a continuos function on a closed interval $[a, b]$. Let there be N equal subintervals, each of length $\delta = (b-a)/N$. Let $x_0, x_1, ..., x_N$ be the endpoints of these subintervals, e.i $x_0 = a, x_1 = a + \delta, x_2 = a + 2\delta, ...$ The sum

$$f(x_1)(x_1 - x_0) + f(x_2)(x_2 - x_1) + ... + f(x_N)(x_N - x_{N-1}) = \sum_{i=1}^{N} f(x_i)\delta$$

is the Riemann sum. Taking the limit gives the Riemann integral:

$$\lim_{\delta \to 0} \sum_{i=1}^{N} f(x_i)\delta = \int_a^b f(x)dx$$

Fundamental Theorem of Calculus

Theorem (Fundamental Theorem of Calculus (Part I))

Let f be a continuous real-valued function defined on a closed interval $[a, b]$. Let F be the function for all $x \in [a, b]$, by

$$F(x) = \int_a^x f(t)dt$$

Then, F is continuous on $[a, b]$, differentiable on the open interval (a, b), and

$$F'(x) = f(x)$$

for all $x \in (a, b)$.

Fundamental Theorem of Calculus

Theorem (Fundamental Theorem of Calculus (Part II))

Let f and F be real-valued functions defined on a closed interval $[a, b]$, such that the derivative of F is f. If f is (riemann) integrable on $[a, b]$ then

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

Note, that there are infinitely many functions F that have f as their derivative, obtained by adding to F an arbitrary constant. So, we write $\int f(x)dx = F(x) + c$, where c is an arbitrary constant.

Example

$$\int_1^4 x dx \quad =$$

Example

$$\int_1^4 x dx = \Big|_1^4 \frac{1}{2}x^2$$

Example

$$\int_1^4 x dx = \Big|_1^4 \frac{1}{2}x^2$$
$$= \frac{1}{2}4^2 - \frac{1}{2}1^2$$

Example

$$\begin{aligned}
\int_1^4 x dx &= \Big|_1^4 \frac{1}{2}x^2 \\
&= \frac{1}{2}4^2 - \frac{1}{2}1^2 \\
&= 7.5
\end{aligned}$$

Definite and Indefinite Integral

The difference between an indefinite and a definite integral is the interval of integration.

$\int f(x)dx$	indefinite integral
$\int_a^b f(x)dx$	definite integral

The numbers *a* and *b* are called, respectively, the lower and upper limit of integration.

Properties

Properties (I)

- $\int af(x)dx = a\int f(x)dx$

Properties

Properties (I)

- $\int af(x)dx = a\int f(x)dx$
- $\int [f(x) + g(x)]\, dx = \int f(x)dx + \int g(x)dx$

Properties (II)

Properties (II)

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

Properties (II)

Properties (II)

- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- $\int_a^a f(x)dx = 0$

Properties (II)

Properties (II)

- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- $\int_a^a f(x)dx = 0$
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$

Properties (II)

Properties (II)

- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- $\int_a^a f(x)dx = 0$
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$
- $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$

Properties (II)

Properties (II)

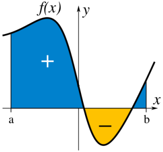
- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- $\int_a^a f(x)dx = 0$
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$
- $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$

Properties (II)

Properties (II)

- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- $\int_a^a f(x)dx = 0$
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$
- $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$

Caution: Areas between the function and the x-axis which are below the x-axis are subtracted!



Special Cases

Special Integrals

- $\int x^a dx = \frac{1}{a+1}x^{a+1} + c$, where $a \neq -1$

Special Cases

Special Integrals

- $\int x^a dx = \frac{1}{a+1}x^{a+1} + c$, where $a \neq -1$
- $\int \frac{1}{x-a} dx = \ln(x-a) + c$, where $x > a$

Special Cases

Special Integrals

- $\int x^a dx = \frac{1}{a+1} x^{a+1} + c$, where $a \neq -1$
- $\int \frac{1}{x-a} dx = \ln(x-a) + c$, where $x > a$
- $\int e^{ax} dx = \frac{1}{a} e^{ax} + c$, where $a \neq 0$

Special Cases

Special Integrals

- $\int x^a dx = \frac{1}{a+1} x^{a+1} + c$, where $a \neq -1$
- $\int \frac{1}{x-a} dx = \ln(x-a) + c$, where $x > a$
- $\int e^{ax} dx = \frac{1}{a} e^{ax} + c$, where $a \neq 0$
- $\int a^x dx = \frac{1}{\ln a} a^x + c$, where $a > 0$ and $a \neq 1$

Linear Algebra I

Motivation I

- A statistical model describes how some variables ($x_0...x_k$) generate another variable y given some parameters ($\beta_0...\beta_k$) and an error term ($\epsilon_1...\epsilon_n$), e.g. the linear regression model

Motivation I

- A statistical model describes how some variables ($x_0...x_k$) generate another variable y given some parameters ($\beta_0...\beta_k$) and an error term ($\epsilon_1...\epsilon_n$), e.g. the linear regression model
- to estimate the parameters, we essentially set up a system of n equations

Motivation I

- A statistical model describes how some variables ($x_0 ... x_k$) generate another variable y given some parameters ($\beta_0 ... \beta_k$) and an error term ($\epsilon_1 ... \epsilon_n$), e.g. the linear regression model
- to estimate the parameters, we essentially set up a system of n equations
- each equation describes how each of our n data point was generated, e.g.

Motivation I

- A statistical model describes how some variables $(x_0 ... x_k)$ generate another variable y given some parameters $(\beta_0 ... \beta_k)$ and an error term $(\epsilon_1 ... \epsilon_n)$, e.g. the linear regression model
- to estimate the parameters, we essentially set up a system of n equations
- each equation describes how each of our n data point was generated, e.g.

Motivation I

- A statistical model describes how some variables ($x_0 ... x_k$) generate another variable y given some parameters ($\beta_0 ... \beta_k$) and an error term ($\epsilon_1 ... \epsilon_n$), e.g. the linear regression model
- to estimate the parameters, we essentially set up a system of n equations
- each equation describes how each of our n data point was generated, e.g.

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_2 + \epsilon_2 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_n + \epsilon_n
\end{aligned}
$$

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
  - Is there a solution to a system?

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
  - Is there a solution to a system?
  - What is the solution set (e.g. the parameters)?

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
    - Is there a solution to a system?
    - What is the solution set (e.g. the parameters)?
    - How many solutions are there? What is the space of solutions?

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
  - Is there a solution to a system?
  - What is the solution set (e.g. the parameters)?
  - How many solutions are there? What is the space of solutions?
  - Can the system be described by a simpler system of equations?

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
  - Is there a solution to a system?
  - What is the solution set (e.g. the parameters)?
  - How many solutions are there? What is the space of solutions?
  - Can the system be described by a simpler system of equations?
  - ...

Motivation II

- Linear Algebra gives us the ability (among other things) to answer questions such as:
  - Is there a solution to a system?
  - What is the solution set (e.g. the parameters)?
  - How many solutions are there? What is the space of solutions?
  - Can the system be described by a simpler system of equations?
  - ...
- Matrix notation is a very efficient way to manipulate (simplify) systems of equations

Linear Algebra I

Vectors

Vector Spaces and Vectors

Definition (Vector Space)

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ Commutativity

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity
- $\mathbf{a} + -\mathbf{a} = \mathbf{0}$ — Existence of an additive inverse

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity
- $\mathbf{a} + -\mathbf{a} = \mathbf{0}$ — Existence of an additive inverse
- $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}$ — Distributivity of scalar sums

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity
- $\mathbf{a} + -\mathbf{a} = \mathbf{0}$ — Existence of an additive inverse
- $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}$ — Distributivity of scalar sums
- $(c + d)\mathbf{a} = c\mathbf{a} + d\mathbf{a}$ — Distributivity of vector sums

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity
- $\mathbf{a} + -\mathbf{a} = \mathbf{0}$ — Existence of an additive inverse
- $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}$ — Distributivity of scalar sums
- $(c + d)\mathbf{a} = c\mathbf{a} + d\mathbf{a}$ — Distributivity of vector sums
- $c(d\mathbf{a}) = (cd)\mathbf{a}$ — Associativity of scalar multiplication

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Vector Spaces and Vectors

Definition (Vector Space)

A vector space V is a nonempty set of objects, called **vectors** denoted with lower case bold letters, on which are defined two operations (addition, multiplication by real scalars), subject to eight axioms:

- $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ — Commutativity
- $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ — Associativity of vector addition
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ — Additive identity
- $\mathbf{a} + -\mathbf{a} = \mathbf{0}$ — Existence of an additive inverse
- $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}$ — Distributivity of scalar sums
- $(c + d)\mathbf{a} = c\mathbf{a} + d\mathbf{a}$ — Distributivity of vector sums
- $c(d\mathbf{a}) = (cd)\mathbf{a}$ — Associativity of scalar multiplication
- $1\mathbf{a} = \mathbf{a}$ — Multiplication identity

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in V \wedge c, d \in \mathcal{R}$

Euclidean Space

Euclidean Space

We focus on a special vector space:

- Euclidean space / Cartesian space - $\mathbb{R}^n$

Euclidean Space

We focus on a special vector space:

- Euclidean space / Cartesian space - $\mathbb{R}^n$
- Euclidean vector: collection of n real numbers either represented as row or column vector:

Euclidean Space

We focus on a special vector space:

- Euclidean space / Cartesian space - $\mathbb{R}^n$
- Euclidean vector: collection of n real numbers either represented as row or column vector:

Euclidean Space

We focus on a special vector space:

- Euclidean space / Cartesian space - $\mathbb{R}^n$
- Euclidean vector: collection of n real numbers either represented as row or column vector:

$$\mathbf{a} = (a_1, a_2, \ldots, a_n) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}'$$

Vector Spaces and Vectors

(continued)

- Terminology: a_i is an **element** or **component**; the vector's **dimension** is the equal to the number of components

Vector Spaces and Vectors

(continued)

- Terminology: a_i is an **element** or **component**; the vector's **dimension** is the equal to the number of components
- Interpretation of **a**:

Vector Spaces and Vectors

(continued)

- Terminology: a_i is an **element** or **component**; the vector's **dimension** is the equal to the number of components
- Interpretation of **a**:
  - line segment connecting the origin $(0, 0)$ with the point **a**

Vector Spaces and Vectors

(continued)

- Terminology: a_i is an **element** or **component**; the vector's **dimension** is the equal to the number of components
- Interpretation of $\mathbf{a}$:
  - line segment connecting the origin $(0, 0)$ with the point $\mathbf{a}$
  - the point $\mathbf{a}$

Vector Operations

Vector addition of vectors with the same dimension is defined as:

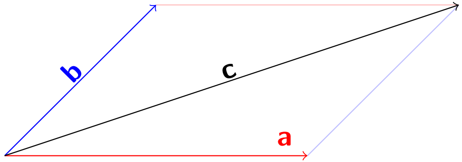
$$
\begin{aligned}
(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) &= (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n) \\
&= \mathbf{a} + \mathbf{b} = \mathbf{c}
\end{aligned}
$$

Vector Operations

Vector addition of vectors with the same dimension is defined as:

$$\begin{aligned}(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) &= (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n) \\ &= \mathbf{a} + \mathbf{b} = \mathbf{c}\end{aligned}$$

Graphically ($\mathbb{R}^2$):



Vector Operations

Scalar multiplication of a vector **a** and scalar α is defined as:

$$\alpha(a_1, a_2, \ldots, a_n) = (\alpha a_1, \alpha a_2, \ldots, \alpha a_n)$$

Vector Operations

Scalar multiplication of a vector **a** and scalar α is defined as:

$$\alpha(a_1, a_2, \ldots, a_n) = (\alpha a_1, \alpha a_2, \ldots, \alpha a_n)$$

Graphically ($\mathbb{R}^2$):

Vector Norm and Distance

The **norm** (length) of a vector $\mathbf{a} = (a_1, a_2, \dots a_n)$ is defined as:

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 \dots a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

A **normalized vector** has a norm of 1. A **zero vector** has a norm of 0 (note: $\|\mathbf{a}\| = 0 \iff a_i = 0 \forall i$).

Vector Norm and Distance

The **norm** (length) of a vector $\mathbf{a} = (a_1, a_2, \ldots a_n)$ is defined as:

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 \ldots a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

A **normalized vector** has a norm of 1. A **zero vector** has a norm of 0 (note: $\|\mathbf{a}\| = 0 \iff a_i = 0 \forall i$).

Application in $\mathbb{R}^2$: (Euclidean) distance between two points **a**, **b**

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \quad \text{(Theorem of Pythagoras)}$$

Vector Norm and Distance

The **norm** (length) of a vector $\mathbf{a} = (a_1, a_2, \ldots a_n)$ is defined as:

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 \ldots a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

A **normalized vector** has a norm of 1. A **zero vector** has a norm of 0 (note: $\|\mathbf{a}\| = 0 \iff a_i = 0 \forall i$).

Application in $\mathbb{R}^2$: (Euclidean) distance between two points **a**, **b**

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \quad \text{(Theorem of Pythagoras)}$$

Generalized to n-dimensions:

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{\sum_{i \in n} (a_i - b_i)^2}$$

Dot product

The **inner product** (dot product) of two vectors of equal dimension is defined as:

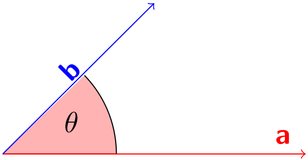
$$\mathbf{a} \cdot \mathbf{b} = a_1 \cdot b_1 + a_2 \cdot b_2 \ldots a_n \cdot b_n = \sum_{i=1}^{n} a_i b_i$$

Dot product

The **inner product** (dot product) of two vectors of equal dimension is defined as:

$$\mathbf{a} \cdot \mathbf{b} = a_1 \cdot b_1 + a_2 \cdot b_2 \ldots a_n \cdot b_n = \sum_{i=1}^{n} a_i b_i$$

Graphically ($\mathbb{R}^2$):



$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos(\theta)$, where θ is the **angle** between the vectors.

Properties

Properties of the Dot Product

If **a**, **b**, and **c** are *n*-vectors and α is a scalar, then

- $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$

Properties

Properties of the Dot Product

If **a**, **b**, and **c** are *n*-vectors and α is a scalar, then

- $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
- $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$

Properties

Properties of the Dot Product

If **a**, **b**, and **c** are *n*-vectors and α is a scalar, then

- $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
- $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$
- $(\alpha \mathbf{a}) \cdot \mathbf{b} = \mathbf{a}(\alpha \mathbf{b}) = \alpha(\mathbf{a} \cdot \mathbf{b})$

Properties

Properties of the Dot Product

If **a**, **b**, and **c** are *n*-vectors and α is a scalar, then

- $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
- $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$
- $(\alpha \mathbf{a}) \cdot \mathbf{b} = \mathbf{a}(\alpha \mathbf{b}) = \alpha(\mathbf{a} \cdot \mathbf{b})$
- $\mathbf{a} \cdot \mathbf{a} > 0 \Longleftrightarrow \mathbf{a} \neq \mathbf{0}$

Linear Algebra I

Matrices

Matrix

A **matrix A**, denoted with bold capital letters, is structured into I **rows** and J **columns**. It is said to have the **size** (dimension) $I \times J$. The cells in the matrix are called **elements**.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} \end{pmatrix}$$

Matrix Operations

Matrix Addition for two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same dimension corresponds to vector addition for each column (or row).

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} =$$

Matrix Operations

Matrix Addition for two matrices **A** and **B** with the same dimension corresponds to vector addition for each column (or row).

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 5 & 7 & 9 \\ 8 & 9 & 10 \end{pmatrix}$$

Matrix Operations

Matrix Addition for two matrices **A** and **B** with the same dimension corresponds to vector addition for each column (or row).

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 5 & 7 & 9 \\ 8 & 9 & 10 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} - \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} =$$

Matrix Operations

Matrix Addition for two matrices **A** and **B** with the same dimension corresponds to vector addition for each column (or row).

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 5 & 7 & 9 \\ 8 & 9 & 10 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} - \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 0 & 2 \\ 3 & 3 & 3 \\ 6 & 7 & 8 \end{pmatrix}$$

Matrix Operations

Scalar Multiplication for a matrix **A** with scalar α corresponds to scalar multiplication of a vector for each column (or row).

Example:

$$2 \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} =$$

Matrix Operations

Scalar Multiplication for a matrix **A** with scalar α corresponds to scalar multiplication of a vector for each column (or row).

Example:

$$2 \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}$$

Properties

Properties of Matrices (I)

1. $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$

Properties

Properties of Matrices (I)

1. $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$
2. $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$

Properties

Properties of Matrices (I)

1. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$

Properties

Properties of Matrices (I)

1. $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$
2. $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
3. $\mathbf{A}+\mathbf{0}=\mathbf{A}$
4. $\mathbf{A}+(-\mathbf{A})=\mathbf{0}$

Properties

Properties of Matrices (I)

1. $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$
2. $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
3. $\mathbf{A}+\mathbf{0}=\mathbf{A}$
4. $\mathbf{A}+(-\mathbf{A})=\mathbf{0}$
5. $(\alpha+\beta)\mathbf{A}=\alpha\mathbf{A}+\beta\mathbf{A}$

Properties

Properties of Matrices (I)

1. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$
4. $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$
5. $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$
6. $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$

Matrix Product

Matrix Product of two matrices **A** and **B** with dimension $w \times x$ and $y \times z$ is defined if the number of columns in **A** is equal to the number of rows in **B**, that is, $x = y$. The new matrix has dimension $w \times z$.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1x} \\ a_{21} & a_{22} & \cdots & a_{2x} \\ \vdots & \vdots & \ddots & \vdots \\ a_{w1} & a_{w2} & \cdots & a_{wx} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1z} \\ b_{21} & b_{22} & \cdots & b_{2z} \\ \vdots & \vdots & \ddots & \vdots \\ b_{y1} & b_{y2} & \cdots & b_{yz} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{y} a_{1i} b_{i1} & \sum_{i=1}^{y} a_{1i} b_{i2} & \cdots & \sum_{i=1}^{y} a_{1i} b_{iz} \\ \sum_{i=1}^{y} a_{2i} b_{i1} & \sum_{i=1}^{y} a_{2i} b_{i2} & \cdots & \sum_{i=1}^{y} a_{2i} b_{iz} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{y} a_{wi} b_{i1} & \sum_{i=1}^{y} a_{wi} b_{i2} & \cdots & \sum_{i=1}^{y} a_{wi} b_{iz} \end{pmatrix}$$

Matrix Product

Example:

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} =$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & & \\ & & \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & \\ & & \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ & & \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & & \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & 3 \cdot 9 + 4 \cdot 12 \\ & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & 3 \cdot 9 + 4 \cdot 12 \\ 5 \cdot 7 + 6 \cdot 10 & & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & 3 \cdot 9 + 4 \cdot 12 \\ 5 \cdot 7 + 6 \cdot 10 & 5 \cdot 8 + 6 \cdot 11 & \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & 3 \cdot 9 + 4 \cdot 12 \\ 5 \cdot 7 + 6 \cdot 10 & 5 \cdot 8 + 6 \cdot 11 & 5 \cdot 9 + 6 \cdot 12 \end{pmatrix}$$

Matrix Product

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \times \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 10 & 1 \cdot 8 + 2 \cdot 11 & 1 \cdot 9 + 2 \cdot 12 \\ 3 \cdot 7 + 4 \cdot 10 & 3 \cdot 8 + 4 \cdot 11 & 3 \cdot 9 + 4 \cdot 12 \\ 5 \cdot 7 + 6 \cdot 10 & 5 \cdot 8 + 6 \cdot 11 & 5 \cdot 9 + 6 \cdot 12 \end{pmatrix}$$

$$= \begin{pmatrix} 27 & 30 & 33 \\ 61 & 68 & 75 \\ 95 & 106 & 117 \end{pmatrix}$$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Note,

- $\mathbf{AB} \neq \mathbf{BA}$

Properties

Properties of Matrices (II)

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Note,

- $\mathbf{AB} \neq \mathbf{BA}$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) \neq (\mathbf{B} + \mathbf{C})\mathbf{A}$

Kronecker Product

If $\mathbf{A}$ is an $w \times x$ matrix and $\mathbf{B}$ is a $y \times z$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $wy \times xz$ block matrix.

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1x} \\ a_{21} & a_{22} & \cdots & a_{2x} \\ \vdots & \vdots & \ddots & \vdots \\ a_{w1} & a_{w2} & \cdots & a_{wx} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1z} \\ b_{21} & b_{22} & \cdots & b_{2z} \\ \vdots & \vdots & \ddots & \vdots \\ b_{y1} & b_{y2} & \cdots & b_{yz} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1x}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2x}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{w1}\mathbf{B} & a_{w2}\mathbf{B} & \cdots & a_{wx}\mathbf{B} \end{pmatrix}$$

Matrix Transposition

The **Transpose** is defined as a matrix where rows and columns are "interchanged". We denote the transpose of a matrix $\mathbf{A}$ by $\mathbf{A}^T$ or $\mathbf{A}'$.

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T =$$

Matrix Transposition

The **Transpose** is defined as a matrix where rows and columns are "interchanged". We denote the transpose of a matrix $\mathbf{A}$ by $\mathbf{A}^T$ or $\mathbf{A}'$.

Example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

Properties

Properties of Matrices (III)

1. $(\mathbf{A}')' = \mathbf{A}$

Properties

Properties of Matrices (III)

1. $(\mathbf{A}')' = \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

Properties

Properties of Matrices (III)

1. $(\mathbf{A}')' = \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
3. $(\alpha \mathbf{A})' = \alpha \mathbf{A}'$

Properties

Properties of Matrices (III)

1. $(\mathbf{A}')' = \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
3. $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$
4. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

Square Matrix

An $i \times j$ matrix $\mathbf{A}$ is called **square matrix** if $i = j$, that is, the numbers of rows and columns are the same.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Symmetric Matrix

A square matrix $\mathbf{A}$ is called **symmetric** if $\mathbf{A} = \mathbf{A}'$. That is, $\mathbf{A}$ is symmetric about its main diagonal. Another way to express this is $a_{ij} = a_{ji} \forall i, j$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 5 \end{pmatrix}' = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 5 \end{pmatrix}$$

Diagonal Matrix

A square symmetric matrix **A** is called **diagonal matrix** if $a_{ij} = 0 \; \forall \; i \neq j$. That is, every element is zero except for the elements on the main diagonal.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Identity Matrix

A square diagonal matrix **A** is called **identity matrix I** if the elements on the main diagonal are all equal to one.

$$\mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Triangular Matrix

A square matrix $\mathbf{A}$ is called upper (lower) **triangular matrix** if $a_{ij} = 0$ for all $i > j$ ($i < j$), that is, a matrix in which all entries below (above) the main diagonal are 0.

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 \\ 4 & 5 & 0 \\ 7 & 8 & 9 \end{pmatrix}$$

Idempotent Matrix

A square matrix $\mathbf{A}$ for which $\mathbf{A} \cdot \mathbf{A} = \mathbf{A}$ is called **idempotent**.

$$\begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix} \times \begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix} = \begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix}$$

The Hessian

Because of the importance of the second-order partial derivatives for constrained optimization there does exist a special way of collecting them, the so-called **Hessian matrix**.

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Trace

The **trace** of a matrix is the sum of the elements on the main diagonal.

$$\text{tr}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 15$$

Linear Algebra II

Linear Algebra II

Systems of Equations

Linear Systems of Equations I

Definition (Linear Equation)

Linear Systems of Equations I

Definition (Linear Equation)

A linear equation in the variables $x_1, ..., x_k$ is an equation that can be written in the form

$$b = a_1 x_1 + a_2 x_2 + ... + a_k x_k,$$

where b and the coefficients $a_1, ..., a_k$ are known, real (or complex) numbers. k is an integer.

Note, in statistics the 'coefficients' are usually the unknowns and the x are known (the data). Just a matter of notation.

Linear Systems of Equations II

Definition (Systems of linear equations)

Linear Systems of Equations II

Definition (Systems of linear equations)

A system of linear equations is a collection of n linear equations of the form:

$$
\begin{aligned}
b_1 &= a_{11}x_1 + a_{12}x_2 + ... + a_{1k}x_k \\
b_2 &= a_{21}x_1 + a_{22}x_2 + ... + a_{2k}x_k \\
\vdots &= \vdots \\
b_n &= a_{n1}x_1 + a_{n2}x_2 + ... + a_{nk}x_k
\end{aligned}
$$

Solving Systems of Linear Equations

1. Equation-by-equation substitution

Solving Systems of Linear Equations

1. Equation-by-equation substitution
2. Gaussian elimination

Solving Systems of Linear Equations

1. Equation-by-equation substitution
2. Gaussian elimination
3. Inverting the coefficient matrix

Solving Systems of Linear Equations

1. Equation-by-equation substitution
2. Gaussian elimination
3. Inverting the coefficient matrix
4. Cramer's Rule

Substitution

1. Solve one equation for one variable.

Substitution

1. Solve one equation for one variable.
2. Substitute the expression into the other $n-1$ equations, resulting in a new system of $n-1$ equations with $k-1$ variables.

Substitution

1. Solve one equation for one variable.
2. Substitute the expression into the other $n-1$ equations, resulting in a new system of $n-1$ equations with $k-1$ variables.
3. Repeat steps 1 and 2 until you get a solution to one variable.

Substitution

1. Solve one equation for one variable.
2. Substitute the expression into the other $n-1$ equations, resulting in a new system of $n-1$ equations with $k-1$ variables.
3. Repeat steps 1 and 2 until you get a solution to one variable.
4. Substitute the solution into previous equations.

Substitution

1. Solve one equation for one variable.
2. Substitute the expression into the other $n-1$ equations, resulting in a new system of $n-1$ equations with $k-1$ variables.
3. Repeat steps 1 and 2 until you get a solution to one variable.
4. Substitute the solution into previous equations.
5. Repeat

Substitution

Solve the following system of equations for $\{x_1, x_2\}$.

Substitution

Solve the following system of equations for $\{x_1, x_2\}$.

$$\begin{aligned} 2x_1 + 3x_2 &= 4 \quad (1) \\ x_1 - 2x_2 &= 5 \quad (2) \end{aligned}$$

Substitution

Solve the following system of equations for $\{x_1, x_2\}$.

$$\begin{aligned} 2x_1 + 3x_2 &= 4 \quad (1) \\ x_1 - 2x_2 &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x_1 and insert this into (1):

$$\begin{aligned} x_1 &= 2x_2 + 5 && \text{(2)'} \\ 4x_2 + 10 + 3x_2 &= 4 && \text{(2)' in (1)} \end{aligned}$$

This gives $x_2 = -\frac{6}{7}$.

Substitution

Solve the following system of equations for $\{x_1, x_2\}$.

$$\begin{aligned} 2x_1 + 3x_2 &= 4 \quad (1) \\ x_1 - 2x_2 &= 5 \quad (2) \end{aligned}$$

Solve equation (2) for x_1 and insert this into (1):

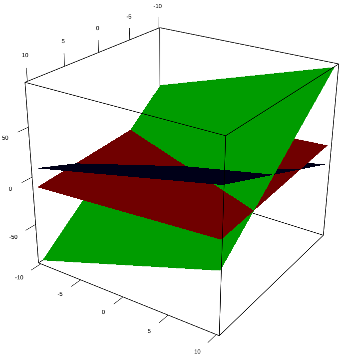
$$\begin{aligned} x_1 &= 2x_2 + 5 \quad && (2)' \\ 4x_2 + 10 + 3x_2 &= 4 && (2)' \text{ in } (1) \end{aligned}$$

This gives $x_2 = -\frac{6}{7}$.
Inserting this into (2)' gives $x_1 = \frac{23}{7}$.

Geometric Interpretation

- Example:

$$3x_1 + 2x_2 - x_3 = 1 \text{(blue plane)}$$
$$2x_1 - 2x_2 + 4x_3 = -2 \text{(red plane)}$$
$$-x_1 + \frac{1}{2}x_2 - \frac{1}{6}x_3 = 0 \text{(green plane)}$$



- Solution:

$$x_1 \approx 0.12$$
$$x_2 \approx 0.06$$
$$x_3 \approx -0.53$$

- intersection of the planes

Matrix Equations

The system of equations

$$3x_1 + 2x_2 - x_3 = 1$$
$$2x_1 - 2x_2 + 4x_3 = -2$$
$$-x_1 + \frac{1}{2}x_2 - \frac{1}{6}x_3 = 0$$

can be written as a matrix equation:

$$\begin{bmatrix} 3 & 2 & -1 \\ 2 & -2 & +4 \\ -1 & \frac{1}{2} & -\frac{1}{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$
$$\mathbf{Ax} = \mathbf{b}$$

Is there a solution to a system of linear equations?

Is there a solution to a system of linear equations?
The Determinant

Determinant I

Consider the following system of linear equations:

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

Determinant I

Consider the following system of linear equations:

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

or $\mathbf{Ax} = \mathbf{b}$.

Determinant I

Consider the following system of linear equations:

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

or $\mathbf{Ax} = \mathbf{b}$. Solving the system of linear equations yields

Determinant I

Consider the following system of linear equations:

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

or $\mathbf{Ax} = \mathbf{b}$. Solving the system of linear equations yields

$$x_1 = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11}a_{22} - a_{21}a_{12}}$$
$$x_2 = \frac{b_2 a_{11} - b_1 a_{21}}{a_{11}a_{22} - a_{21}a_{12}}$$

Note that the denominators are the same. These have to be nonzero for a unique solution to exist. The system would have none or an infinite number of solutions otherwise.

Determinant II

In this sense, the value of the denominator determines whether there is a unique solution to the equation system.

Determinant II

In this sense, the value of the denominator determines whether there is a unique solution to the equation system.

In fact, $a_{11}a_{22} - a_{21}a_{12}$ is called the **determinant** of the matrix **A**.

Determinant II

In this sense, the value of the denominator determines whether there is a unique solution to the equation system.

In fact, $a_{11}a_{22} - a_{21}a_{12}$ is called the **determinant** of the matrix $\mathbf{A}$.

$$|\mathbf{A}| = det(\mathbf{A}) = a_{11}a_{22} - a_{21}a_{12}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\det(\mathbf{A}) = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\det(\mathbf{A}) \quad = \quad \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\begin{aligned} \det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\ &= a_{11} \cdot a_{22} \cdot a_{33} \end{aligned}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\begin{aligned}\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\ &= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31}\end{aligned}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$
\begin{aligned}
\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\
&= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32}
\end{aligned}
$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\begin{aligned}\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\ &= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\ &\quad - a_{31} \cdot a_{22} \cdot a_{13}\end{aligned}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\begin{aligned}
\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\
&= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\
&\quad - a_{31} \cdot a_{22} \cdot a_{13} - a_{32} \cdot a_{23} \cdot a_{11}
\end{aligned}$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$
\begin{aligned}
\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\
&= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\
&\quad - a_{31} \cdot a_{22} \cdot a_{13} - a_{32} \cdot a_{23} \cdot a_{11} - a_{33} \cdot a_{21} \cdot a_{12}
\end{aligned}
$$

Determinant III

Sarrus's rule is a simple rule for calculating the determinant of 3×3 matrices.

$$\begin{aligned}
\det(\mathbf{A}) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\
&= a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\
&\quad - a_{31} \cdot a_{22} \cdot a_{13} - a_{32} \cdot a_{23} \cdot a_{11} - a_{33} \cdot a_{21} \cdot a_{12}
\end{aligned}$$

Determinants for **square matrices** of dimension larger then three

are not that easy to determine. However, there are procedures to calculate them. See Sydsæter/Hammond (2008), 580-582.

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(c \cdot \mathbf{A}) = c^n \cdot \det(\mathbf{A})$, where n is the dimension of $\mathbf{A}$

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(c \cdot \mathbf{A}) = c^n \cdot \det(\mathbf{A})$, where n is the dimension of $\mathbf{A}$
- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(c \cdot \mathbf{A}) = c^n \cdot \det(\mathbf{A})$, where n is the dimension of $\mathbf{A}$
- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(c \cdot \mathbf{A}) = c^n \cdot \det(\mathbf{A})$, where n is the dimension of $\mathbf{A}$
- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$
- In particular, if all the elements in a row (or column) of $\mathbf{A}$ are 0, then $\det(\mathbf{A}) = 0$.

Properties

Properties of Determinants (I)

- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(c \cdot \mathbf{A}) = c^n \cdot \det(\mathbf{A})$, where n is the dimension of $\mathbf{A}$
- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$
- In particular, if all the elements in a row (or column) of $\mathbf{A}$ are 0, then $\det(\mathbf{A}) = 0$.
- If two rows (or columns) of $\mathbf{A}$ are interchanged, the determinant changes sign, but the absolute value remains unchanged.

Properties

Properties of Determinants (I)

- If two rows (or columns) are proportional, i.e. multiples of each other, then $\det(\mathbf{A}) = 0$.

Properties

Properties of Determinants (I)

- If two rows (or columns) are proportional, i.e. multiples of each other, then $\det(\mathbf{A}) = 0$.
- The value of the determinant of $\mathbf{A}$ is unchanged if a multiple of one row (or column) is added to a different row (or column) of $\mathbf{A}$.

Properties

Properties of Determinants (I)

- If two rows (or columns) are proportional, i.e. multiples of each other, then $\det(\mathbf{A}) = 0$.
- The value of the determinant of $\mathbf{A}$ is unchanged if a multiple of one row (or column) is added to a different row (or column) of $\mathbf{A}$.

Properties

Properties of Determinants (I)

- If two rows (or columns) are proportional, i.e. multiples of each other, then $\det(\mathbf{A}) = 0$.
- The value of the determinant of $\mathbf{A}$ is unchanged if a multiple of one row (or column) is added to a different row (or column) of $\mathbf{A}$.

Note, $\det(\mathbf{A} + \mathbf{B}) \neq \det(\mathbf{A}) + \det(\mathbf{B})$.

Solving Systems of Linear Equations I

Solving Systems of Linear Equations I
Gaussian Elimination

Gaussian Elimination

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 \ldots a_{1k}x_k &= b_1 \\
a_{21}x_1 + a_{22}x_2 \ldots a_{2k}x_k &= b_2 \\
\vdots &= \vdots \\
a_{n1}x_1 + a_{n2}x_2 \ldots a_{nk}x_k &= b_n
\end{aligned}
$$

Gaussian Elimination

$$a_{11}x_1 + a_{12}x_2 ... a_{1k}x_k = b_1$$
$$a_{21}x_1 + a_{22}x_2 ... a_{2k}x_k = b_2$$
$$\vdots = \vdots$$
$$a_{n1}x_1 + a_{n2}x_2 ... a_{nk}x_k = b_n$$

Rewrite the system in an **augmented matrix $\mathbf{A}|\mathbf{b}$**

Gaussian Elimination

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 ... a_{1k}x_k &= b_1 \\
a_{21}x_1 + a_{22}x_2 ... a_{2k}x_k &= b_2 \\
\vdots &= \vdots \\
a_{n1}x_1 + a_{n2}x_2 ... a_{nk}x_k &= b_n
\end{aligned}
$$

Rewrite the system in an **augmented matrix $\mathbf{A}|\mathbf{b}$**

$$
\left(\begin{array}{cccc|c}
a_{11} & a_{12} & \dots & a_{1k} & b_1 \\
a_{21} & a_{22} & \dots & a_{2k} & b_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nk} & b_n
\end{array}\right)
$$

Gaussian Elimination

Gaussian Elimination

Using **elementary row operations**,

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,
- add one row onto another or

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,
- add one row onto another or
- multiply every factor of one row with a constant,

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,
- add one row onto another or
- multiply every factor of one row with a constant,

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,
- add one row onto another or
- multiply every factor of one row with a constant,

obtain a **row echelon form** of the augmented matrix:

Gaussian Elimination

Using **elementary row operations**,

- Swap rows,
- add one row onto another or
- multiply every factor of one row with a constant,

obtain a **row echelon form** of the augmented matrix:

$$\left(\begin{array}{ccccc|c} a_{11}^* & a_{12}^* & a_{13}^* & \dots & a_{1n}^* & b_1^* \\ 0 & a_{22}^* & a_{23}^* & \dots & a_{2n}^* & b_2^* \\ 0 & 0 & a_{33}^* & \dots & a_{3n}^* & b_3^* \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & a_{3n}^* & b_n^* \end{array}\right)$$

Iterated substitution gives you the solution vector $\mathbf{x}$ if it exists. Or..

Gaussian Elimination

continue to obtain a **reduced row echelon form** of the augmented matrix (Gauss-Jordan elimination):

Gaussian Elimination

continue to obtain a **reduced row echelon form** of the augmented matrix (Gauss-Jordan elimination):

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & \dots & 0 & \tilde{b}_1 \\ 0 & 1 & 0 & \dots & 0 & \tilde{b}_2 \\ 0 & 0 & 1 & \dots & 0 & \tilde{b}_3 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 & \tilde{b}_n \end{array}\right)$$

where $\tilde{\mathbf{b}}$ is the solution vector for $\mathbf{x}$.

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right)$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I}$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{}$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III}$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & .5 & 0 & 1.5 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & .5 & 0 & 1.5 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

$$\xrightarrow[2II]{-III}$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & .5 & 0 & 1.5 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

$$\xrightarrow[2II]{-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & .5 & 0 & 1.5 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

$$\xrightarrow[2II]{-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \xrightarrow[I-II]{.5I}$$

Gaussian Elimination

Example (from Wikipedia)

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right) \xrightarrow[II+1.5I]{III+I} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right)$$

$$\xrightarrow[III-4II]{} \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & .5 & .5 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow[II+.5III]{I-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & .5 & 0 & 1.5 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

$$\xrightarrow[2II]{-III} \left(\begin{array}{ccc|c} 2 & 1 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \xrightarrow[I-II]{.5I} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

Is there a solution to a system of linear equations?

Is there a solution to a system of linear equations?
The Matrix Rank

Rank

Definition (Matrix Rank)

The **rank** r of a $n \times n$ square matrix is equal to the number of nonzero rows after Gaussian elimination. A matrix possesses **full rank** if $r = n$.

Rank

Definition (Matrix Rank)

The **rank** r of a $n \times n$ square matrix is equal to the number of nonzero rows after Gaussian elimination. A matrix possesses **full rank** if $r = n$.

Note, the concept also applies to non-square matrices.

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix.

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix. A system $\mathbf{Ax} = \mathbf{b}$ has a unique solution if $\mathbf{A}$ has **full rank**.

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix. A system $\mathbf{Ax} = \mathbf{b}$ has a unique solution if $\mathbf{A}$ has **full rank**.

Let r represent the rank of a $n \times n$ matrix $\mathbf{A}_R$ (row echelon form).
Suppose $\mathbf{A}_R$ has r nonzero rows and $n - r$ zero rows.

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix. A system $\mathbf{Ax} = \mathbf{b}$ has a unique solution if $\mathbf{A}$ has **full rank**.

Let r represent the rank of a $n \times n$ matrix $\mathbf{A}_R$ (row echelon form).

Suppose $\mathbf{A}_R$ has r nonzero rows and $n - r$ zero rows.

Consequently, $r < n$.

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix. A system $\mathbf{Ax} = \mathbf{b}$ has a unique solution if $\mathbf{A}$ has **full rank**.
Let r represent the rank of a $n \times n$ matrix $\mathbf{A}_R$ (row echelon form).
Suppose $\mathbf{A}_R$ has r nonzero rows and $n - r$ zero rows.
Consequently, $r < n$.

- if any zero row of $\mathbf{A}_R$ is associated with a nonzero entry in $\mathbf{b}_R$, then the system of equations is **overdetermined** and has no solutions

Rank and Systems of Linear Equations

Let $\mathbf{A}$ be a $n \times n$ matrix. A system $\mathbf{Ax} = \mathbf{b}$ has a unique solution if $\mathbf{A}$ has **full rank**.
Let r represent the rank of a $n \times n$ matrix $\mathbf{A}_R$ (row echelon form).
Suppose $\mathbf{A}_R$ has r nonzero rows and $n - r$ zero rows.
Consequently, $r < n$.

- if any zero row of $\mathbf{A}_R$ is associated with a nonzero entry in $\mathbf{b}_R$, then the system of equations is **overdetermined** and has no solutions
- if every zero row of $\mathbf{A}_R$ corresponds to a zero entry in $\mathbf{b}_R$, then the system of equations is **underdetermined** and has infinity of solutions

Solving Systems of Linear Equations II

Solving Systems of Linear Equations II
Inverting the coefficient matrix

Inverse I

Consider a real-valued number c, where $c \neq 0$. Then there exists a c^{-1} such that $c \cdot c^{-1} = 1$.

Inverse I

Consider a real-valued number c, where $c \neq 0$. Then there exists a c^{-1} such that $c \cdot c^{-1} = 1$.

There is an equivalent in matrix algebra. Recall that an identity matrix is a matrix whose elements on the main diagonal equal to one and all other elements are equal to zero.

Inverse I

Consider a real-valued number c, where $c \neq 0$. Then there exists a c^{-1} such that $c \cdot c^{-1} = 1$.

There is an equivalent in matrix algebra. Recall that an identity matrix is a matrix whose elements on the main diagonal equal to one and all other elements are equal to zero.

Given a square matrix $\mathbf{A}$, we say that $\mathbf{X}$ is an **inverse** of $\mathbf{A}$ if there exists a matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}$.

Inverse I

Consider a real-valued number c, where $c \neq 0$. Then there exists a c^{-1} such that $c \cdot c^{-1} = 1$.

There is an equivalent in matrix algebra. Recall that an identity matrix is a matrix whose elements on the main diagonal equal to one and all other elements are equal to zero.

Given a square matrix $\mathbf{A}$, we say that $\mathbf{X}$ is an **inverse** of $\mathbf{A}$ if there exists a matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}$.

If the inverse of $\mathbf{A}$ exists, we denote it by $\mathbf{A}^{-1}$.

Inverse I

Consider a real-valued number c, where $c \neq 0$. Then there exists a c^{-1} such that $c \cdot c^{-1} = 1$.

There is an equivalent in matrix algebra. Recall that an identity matrix is a matrix whose elements on the main diagonal equal to one and all other elements are equal to zero.

Given a square matrix $\mathbf{A}$, we say that $\mathbf{X}$ is an **inverse** of $\mathbf{A}$ if there exists a matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}$.

If the inverse of $\mathbf{A}$ exists, we denote it by $\mathbf{A}^{-1}$.

A matrix is invertible if and only if $\det(\mathbf{A}) \neq 0$. $\mathbf{A}$ is said to be **nonsingular** in this case. In the opposite case of $\det(\mathbf{A}) = 0$ we call **A singular**.

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right)$$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right) \qquad \begin{array}{l} \mathrm{I} - 3 \cdot \mathrm{III} \\ \mathrm{II} - \mathrm{I} \\ \mathrm{III} - \mathrm{I} \end{array}$$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right) \quad \begin{array}{l} \text{I} - 3 \cdot \text{III} \\ \text{II} - \text{I} \\ \text{III} - \text{I} \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & 0 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right) \quad \begin{array}{l} \mathrm{I} - 3 \cdot \mathrm{III} \\ \mathrm{II} - \mathrm{I} \\ \mathrm{III} - \mathrm{I} \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & 0 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \quad \mathrm{I} - 3 \cdot \mathrm{II}$$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & 0 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

$\mathrm{I} - 3 \cdot \mathrm{III}$
$\mathrm{II} - \mathrm{I}$
$\mathrm{III} - \mathrm{I}$
$\mathrm{I} - 3 \cdot \mathrm{II}$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right) \quad \begin{array}{l} \mathrm{I} - 3\cdot\mathrm{III} \\ \mathrm{II} - \mathrm{I} \\ \mathrm{III} - \mathrm{I} \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & 0 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \quad \mathrm{I} - 3\cdot\mathrm{II}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \mathrm{II} \leftrightarrow \mathrm{III}$$

Inverse II

We can invert a matrix using the Gauss-Jordan algorithm for systems of linear equations.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 3 & 0 & 0 & 1 \end{array}\right) \quad \begin{array}{l} \text{I} - 3 \cdot \text{III} \\ \text{II} - \text{I} \\ \text{III} - \text{I} \\ \text{I} - 3 \cdot \text{II} \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & 0 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \text{II} \leftrightarrow \text{III}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right)$$

Inverse and Equations

What is the benefit of using an inverse?

Inverse and Equations

What is the benefit of using an inverse?

Reconsider the system of linear equations

$$\begin{aligned} b_1 + 3b_2 + 3b_3 &= -7 \\ b_1 + 3b_2 + 4b_3 &= 2 \\ b_1 + 4b_2 + 2b_3 &= -10 \end{aligned}$$

Inverse and Equations

What is the benefit of using an inverse?

Reconsider the system of linear equations

$$\begin{aligned} b_1 + 3b_2 + 3b_3 &= -7 \\ b_1 + 3b_2 + 4b_3 &= 2 \\ b_1 + 4b_2 + 2b_3 &= -10 \end{aligned}$$

We can express this system using matrix notation

$$\begin{aligned} \begin{pmatrix} 1 & 3 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} &= \begin{pmatrix} -7 \\ 2 \\ -10 \end{pmatrix} \\ \mathbf{Xb} &= \mathbf{y} \end{aligned}$$

Inverse and Equations

What is the benefit of using an inverse?

Reconsider the system of linear equations

$$\begin{aligned} b_1 + 3b_2 + 3b_3 &= -7 \\ b_1 + 3b_2 + 4b_3 &= 2 \\ b_1 + 4b_2 + 2b_3 &= -10 \end{aligned}$$

We can express this system using matrix notation

$$\begin{pmatrix} 1 & 3 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} -7 \\ 2 \\ -10 \end{pmatrix}$$

$$\mathbf{Xb} = \mathbf{y}$$

If we multiply both sides of the equation with the inverse of $\mathbf{X}$ from the left, we solve the system for $\mathbf{b}$.

$$\mathbf{b} = \mathbf{X}^{-1}\mathbf{y}$$

Other Ways of Calculating the Inverse

The inverse of every 2×2 matrix $\mathbf{A}$ can be derived the following way.

$$\text{Let } \mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } \det(\mathbf{A}) = ad - bc \neq 0, \text{ then}$$

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Other Ways of Calculating the Inverse

The inverse of every 2×2 matrix $\mathbf{A}$ can be derived the following way.

$$\text{Let } \mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } \det(\mathbf{A}) = ad - bc \neq 0, \text{ then}$$

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

An additional way for finding the inverse of an $n \times n$ matrix $\mathbf{A}$ that does not employ Gaussian elimination uses the so-called adjoint of $\mathbf{A}$ (see Sydsæter/Hammond 2008, 597).

Rules of Matrix Inversion

Rules of Matrix Inversion

Let **A** and **B** be invertible $n \times n$ matrices. Then:

Rules of Matrix Inversion

Rules of Matrix Inversion

Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then:

- $\mathbf{A}^{-1}$ is invertible, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

Rules of Matrix Inversion

Rules of Matrix Inversion

Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then:

- $\mathbf{A}^{-1}$ is invertible, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $\mathbf{AB}$ is invertible, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Rules of Matrix Inversion

Rules of Matrix Inversion

Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then:

- $\mathbf{A}^{-1}$ is invertible, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $\mathbf{AB}$ is invertible, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- The transpose $\mathbf{A}'$ is invertible, and $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$

Rules of Matrix Inversion

Rules of Matrix Inversion

Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then:

- $\mathbf{A}^{-1}$ is invertible, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $\mathbf{AB}$ is invertible, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- The transpose $\mathbf{A}'$ is invertible, and $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
- $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$ whenever $c \neq 0$

Rules of Matrix Inversion

Rules of Matrix Inversion

Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then:

- $\mathbf{A}^{-1}$ is invertible, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $\mathbf{AB}$ is invertible, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- The transpose $\mathbf{A}'$ is invertible, and $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
- $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$ whenever $c \neq 0$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$

Probability Theory

Defintions

- **Experiment**: A probabilistic process that realizes an outcome from a sample space.

Defintions

- **Experiment**: A probabilistic process that realizes an outcome from a sample space.
- **Sample Space**: S (or Ω), a finite set, the collection of all possible outcomes in an experiment

Defintions

- **Experiment**: A probabilistic process that realizes an outcome from a sample space.
- **Sample Space**: S (or Ω), a finite set, the collection of all possible outcomes in an experiment
- **Event**: $A \subseteq S$, a subset from the sample space

Axioms and Definition of Probability

Definition (Probability)

A probability distribution or simply a probability for event A, on a sample space S, is a specification of numbers $Pr(A)$ which satisfy Axioms 1-3 (Kolmogorov probability axioms).

Axioms and Definition of Probability

Definition (Probability)

A probability distribution or simply a probability for event A, on a sample space S, is a specification of numbers $Pr(A)$ which satisfy Axioms 1-3 (Kolmogorov probability axioms).

- Axiom 1 (Non-Negativity):

$$Pr(A_i) \geq 0 \; \forall i$$

Axioms and Definition of Probability

Definition (Probability)

A probability distribution or simply a probability for event A, on a sample space S, is a specification of numbers $Pr(A)$ which satisfy Axioms 1-3 (Kolmogorov probability axioms).

- Axiom 1 (Non-Negativity):

$$Pr(A_i) \geq 0 \; \forall i$$

- Axiom 2 (Normalization):

$$Pr(S) = 1.$$

Axioms and Definition of Probability

Definition (Probability)

A probability distribution or simply a probability for event A, on a sample space S, is a specification of numbers $Pr(A)$ which satisfy Axioms 1-3 (Kolmogorov probability axioms).

- Axiom 1 (Non-Negativity):

$$Pr(A_i) \geq 0 \ \forall i$$

- Axiom 2 (Normalization):

$$Pr(S) = 1.$$

- Axiom 3 (Additivity):

$$Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$$

with all A_i are disjoint.

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$
- Event $A \subseteq S$

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$
- Event $A \subseteq S$
- Let $|A| = k$

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$
- Event $A \subseteq S$
- Let $|A| = k$

$$Pr(A) = k/n$$

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$
- Event $A \subseteq S$
- Let $|A| = k$

$$Pr(A) = k/n$$

- to determine n and k it is often useful to consider counting rules

Classical Probability of an Event

- Simple Sample Space: $|S| = n$ with $S = \{s_1, ..., s_n\}$
- Event $A \subseteq S$
- Let $|A| = k$

$$Pr(A) = k/n$$

- to determine n and k it is often useful to consider counting rules
- note: classical probability $\neq$ empirical probability $\neq$ subjective probability

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$
- $A \subset B \implies Pr(A) \leq Pr(B)$

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$
- $A \subset B \implies Pr(A) \leq Pr(B)$
- $Pr(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n Pr(A_i)$ with all A_i are disjoint

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$
- $A \subset B \implies Pr(A) \leq Pr(B)$
- $Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i)$ with all A_i are disjoint
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$, where $P(A \cap B)$ is the joint probability of A and B

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$
- $A \subset B \implies Pr(A) \leq Pr(B)$
- $Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i)$ with all A_i are disjoint
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$, where $P(A \cap B)$ is the joint probability of A and B

Basic Theorems

Let $A, B \subseteq S$:

- $Pr(\emptyset) = 0$
- $Pr(A^c) = 1 - Pr(A)$ where A^c is the complement set to A
- $0 \leq Pr(A) \leq 1$
- $A \subset B \implies Pr(A) \leq Pr(B)$
- $Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i)$ with all A_i are disjoint
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$, where $P(A \cap B)$ is the joint probability of A and B

Note: $Pr(A \cap B)$ is also denoted $Pr(AB)$ or $P(A, B)$

Probability Theory

Combinatorics

Permutation and Combination

	with replacement	without replacement
Permutation (considering sequence)	n^k	$\binom{n}{k}k! = \frac{n!}{(n-k)!}$
Combination (disregarding sequence)	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Binomial Coefficient

- "n choose k"

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \; \forall \; 0 < k \leq n$$

Binomial Coefficient

- "n choose k"

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \;\forall\; 0 < k \leq n$$

- Example: How many ways can a voter select three candidates from a field of seven?

Binomial Coefficient

- "n choose k"

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \; \forall \; 0 < k \leq n$$

- Example: How many ways can a voter select three candidates from a field of seven?

$$\binom{7}{3} =$$

Binomial Coefficient

- "n choose k"

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \; \forall \; 0 < k \leq n$$

- Example: How many ways can a voter select three candidates from a field of seven?

$$\binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7!}{3! \times 4!} = \frac{7 \times 6 \times 5}{3 \times 2} = 35.$$

Examples I

$$k = 2,\ S = \{A, B, C\} \implies n = 3$$

Examples I

$k = 2,\ S = \{A, B, C\} \implies n = 3$

	with replacement	without replacement
Permutation (considering sequence)	$\|\{AB, BA, BB, AC, CA$ $AA, BC, CB, CC\}\| = 9$	
Combination (disregarding sequence)		

Examples I

$k = 2,\ S = \{A, B, C\} \implies n = 3$

	with replacement	without replacement
Permutation (considering sequence)	$\|\{AB, BA, BB, AC, CA$ $AA, BC, CB, CC\}\| = 9$	$\|\{AB, BA, AC, CA$ $BC, CB\}\| = 6$
Combination (disregarding sequence)		

Examples I

$k = 2,\ S = \{A, B, C\} \implies n = 3$

	with replacement	without replacement
Permutation (considering sequence)	$\lvert\{AB, BA, BB, AC, CA$ $AA, BC, CB, CC\}\rvert = 9$	$\lvert\{AB, BA, AC, CA$ $BC, CB\}\rvert = 6$
Combination (disregarding sequence)	$\lvert\{AB, AC, BC, AA, BB$ $CC\}\rvert = 6$	

Examples I

$k = 2,\ S = \{A, B, C\} \implies n = 3$

	with replacement	without replacement
Permutation (considering sequence)	$\lvert\{AB, BA, BB, AC, CA$ $AA, BC, CB, CC\}\rvert = 9$	$\lvert\{AB, BA, AC, CA$ $BC, CB\}\rvert = 6$
Combination (disregarding sequence)	$\lvert\{AB, AC, BC, AA, BB$ $CC\}\rvert = 6$	$\lvert\{AB, AC$ $BC\}\rvert = 3$

Examples II

Let there be 4 train passengers waiting for tickets. How many sequences are there to sell them their train tickets?

Examples II

Let there be 4 train passengers waiting for tickets. How many sequences are there to sell them their train tickets?
$k = n = 4 \implies \binom{n}{k} k! = 24$

Probability Theory

Conditional Probability

Definition

Definition (Conditional Probability)

Let A,B be two events with probability larger than zero. The conditional probability of A given B is:
$p(A|B) = p(A \cap B)/p(B)$

Definition

Definition (Conditional Probability)

Let A,B be two events with probability larger than zero. The conditional probability of A given B is:
$p(A|B) = p(A \cap B)/p(B)$

Interpretation: Given that B occurred, what is the probability for A?

Corollaries

- Multiplication Rule:

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:
  - $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:
  - $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$
- Law of Total Probability

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:
  - $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$
- Law of Total Probability
  - Let $A_1, ..., A_k$ be disjoint events and $\bigcup_{i=1}^{k} A_i = S$. For any event B in S and as long as $p(A_j) > 0 \forall j$: $p(B) = \sum_{i=1}^{k} p(A_i)p(B|A_i)$.

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:
  - $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$
- Law of Total Probability
  - Let $A_1, ..., A_k$ be disjoint events and $\bigcup_{i=1}^{k} A_i = S$. For any event B in S and as long as $p(A_j) > 0 \forall j$: $p(B) = \sum_{i=1}^{k} p(A_i)p(B|A_i)$.
- Bayes' Theorem

Corollaries

- Multiplication Rule:
  - $p(A \cap B) = p(A|B)p(B)$
- General Product Rule:
  - $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$
- Law of Total Probability
  - Let $A_1, ..., A_k$ be disjoint events and $\bigcup_{i=1}^{k} A_i = S$. For any event B in S and as long as $p(A_j) > 0 \forall j$: $p(B) = \sum_{i=1}^{k} p(A_i)p(B|A_i)$.
- Bayes' Theorem
  - $p(A|B) = \frac{p(B|A)p(A)}{p(B)}$

Example I

Suppose you role a dice, but you can't observe the outcome. What is the probability that you get a 6? Does this probability change when you have been told that the outcome was an even number?

Example I

Suppose you role a dice, but you can't observe the outcome. What is the probability that you get a 6? Does this probability change when you have been told that the outcome was an even number?

$$P(\text{roll a 6}) \quad =$$

Example I

Suppose you role a dice, but you can't observe the outcome. What is the probability that you get a 6? Does this probability change when you have been told that the outcome was an even number?

$$P(\text{roll a 6}) = 1/6$$

Example I

Suppose you role a dice, but you can't observe the outcome. What is the probability that you get a 6? Does this probability change when you have been told that the outcome was an even number?

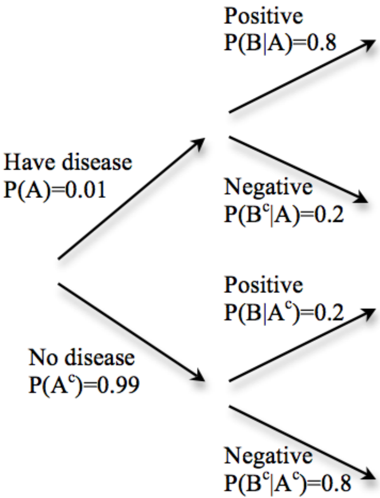
$$\begin{aligned} P(\text{roll a 6}) &= 1/6 \\ P(\text{roll a 6}|\text{even}) &= \end{aligned}$$

Example I

Suppose you role a dice, but you can't observe the outcome. What is the probability that you get a 6? Does this probability change when you have been told that the outcome was an even number?

$$
\begin{aligned}
P(\text{roll a } 6) &= 1/6 \\
P(\text{roll a } 6|\text{even}) &= (1/6)/(1/2) = 1/3
\end{aligned}
$$

Example II


Positive
P(B|A)=0.8
Have disease
P(A)=0.01
Negative
P(B^c|A)=0.2
Positive
P(B|A^c)=0.2
No disease
P(A^c)=0.99
Negative
P(B^c|A^c)=0.8

Bayes' Theorem

- first appeared in an essay by Thomas Bayes, 1763

Bayes' Theorem

- first appeared in an essay by Thomas Bayes, 1763
- post-mortem published by Richard Price

Bayes' Theorem

- first appeared in an essay by Thomas Bayes, 1763
- post-mortem published by Richard Price
- Laplace (1774,1781) provided (independently) most of the relevant analysis

Bayes' Theorem

- first appeared in an essay by Thomas Bayes, 1763
- post-mortem published by Richard Price
- Laplace (1774,1781) provided (independently) most of the relevant analysis
- foundation of Bayesian Statistics, formal modeling of learning, philosophy of scientific progress, ...

Bayes' Theorem

Bayes' Theorem

$$p(A \cap B) = p(A|B)p(B)$$

Bayes' Theorem

$$
\begin{aligned}
p(A \cap B) &= p(A|B)p(B) \\
p(A \cap B) &= p(B|A)p(A)
\end{aligned}
$$

Bayes' Theorem

$$\begin{aligned} p(A \cap B) &= p(A|B)p(B) \\ p(A \cap B) &= p(B|A)p(A) \\ p(A|B)p(B) &= p(B|A)p(A) \end{aligned}$$

Bayes' Theorem

$$
\begin{aligned}
p(A \cap B) &= p(A|B)p(B) \\
p(A \cap B) &= p(B|A)p(A) \\
p(A|B)p(B) &= p(B|A)p(A) \\
p(A|B) &= \frac{p(B|A)p(A)}{p(B)}
\end{aligned}
$$

Applied Bayes: Learning Example I

Example: Is a particular coin fair?

Applied Bayes: Learning Example I

Example: Is a particular coin fair?

- H_1, the event that a head is obtained after tossing

Applied Bayes: Learning Example I

Example: Is a particular coin fair?

- H_1, the event that a head is obtained after tossing
- hypothesis F, the coin is fair; hypothesis $\neg F$, the coin is not fair (has two heads)

Applied Bayes: Learning Example I

Example: Is a particular coin fair?

- H_1, the event that a head is obtained after tossing
- hypothesis F, the coin is fair; hypothesis $\neg F$, the coin is not fair (has two heads)
- suppose you have no reason to belief more in either of the two hypothesis a-priori

Applied Bayes: Learning Example I

Example: Is a particular coin fair?

- H_1, the event that a head is obtained after tossing
- hypothesis F, the coin is fair; hypothesis $\neg F$, the coin is not fair (has two heads)
- suppose you have no reason to belief more in either of the two hypothesis a-priori
- What is the probability of hypothesis F and $\neg F$ after you tossed the coin and you saw a head?

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$
- the probability for $p(H_1)$ is given by the law of total probability

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$
- the probability for $p(H_1)$ is given by the law of total probability
- **posterior** probability is given by Bayes Theorem:

(1)

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$
- the probability for $p(H_1)$ is given by the law of total probability
- **posterior** probability is given by Bayes Theorem:

$$p(F|H_1) = \frac{p(F)p(H_1|F)}{p(F)p(H_1|F)+p(\neg F)p(H_1|\neg F)} \quad (1)$$

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$
- the probability for $p(H_1)$ is given by the law of total probability
- **posterior** probability is given by Bayes Theorem:

$$\begin{aligned} p(F|H_1) &= \frac{p(F)p(H_1|F)}{p(F)p(H_1|F)+p(\neg F)p(H_1|\neg F)} \\ &= \frac{(0.5)(0.5)}{(0.5)(0.5)+(0.5)(1)} \end{aligned} \tag{1}$$

Applied Bayes: Learning Example II

- **prior** probability about the fairness is $p(F) = p(\neg F) = 0.5$
- if the coin is fair, $p(H_1|F) = 0.5$, if it's unfair $p(H_1|\neg F) = 1$
- the probability for $p(H_1)$ is given by the law of total probability
- **posterior** probability is given by Bayes Theorem:

$$\begin{aligned} p(F|H_1) &= \frac{p(F)p(H_1|F)}{p(F)p(H_1|F)+p(\neg F)p(H_1|\neg F)} \\ &= \frac{(0.5)(0.5)}{(0.5)(0.5)+(0.5)(1)} \\ &= 1/3 \end{aligned} \tag{1}$$

Applied Bayes: Learning Example III

- What is the posterior probability to see head when you toss again (event H_2)?

Applied Bayes: Learning Example III

- What is the posterior probability to see head when you toss again (event H_2)?
- now, the prior probability is: $p(F) = 1/3$, $p(\neg F) = 2/3$

Applied Bayes: Learning Example III

- What is the posterior probability to see head when you toss again (event H_2)?
- now, the prior probability is: $p(F) = 1/3$, $p(\neg F) = 2/3$

$$p(F|H_2) = \frac{p(F)p(H_2|F)}{p(F)p(H_2|F)+p(\neg F)p(H_2|\neg F)}$$

Applied Bayes: Learning Example III

- What is the posterior probability to see head when you toss again (event H_2)?
- now, the prior probability is: $p(F) = 1/3$, $p(\neg F) = 2/3$

$$\begin{aligned} p(F|H_2) &= \frac{p(F)p(H_2|F)}{p(F)p(H_2|F)+p(\neg F)p(H_2|\neg F)} \\ &= \frac{(1/3)(0.5)}{(1/3)(0.5)+(2/3)(1)} \\ &= 1/5 \end{aligned}$$

- for three heads in a row $p(F|H_3) = 1/9$ ...

Applied Bayes: Learning Example III

- What is the posterior probability to see head when you toss again (event H_2)?
- now, the prior probability is: $p(F) = 1/3$, $p(\neg F) = 2/3$

$$
\begin{aligned}
p(F|H_2) &= \frac{p(F)p(H_2|F)}{p(F)p(H_2|F)+p(\neg F)p(H_2|\neg F)} \\
&= \frac{(1/3)(0.5)}{(1/3)(0.5)+(2/3)(1)} \\
&= 1/5
\end{aligned}
$$

- for three heads in a row $p(F|H_3) = 1/9$ ...
- this process is called **Bayesian Updating**

Applied Bayes: Statistical Models

- let θ denote a parameter and y the data

Applied Bayes: Statistical Models

- let θ denote a parameter and y the data
- from Bayes' Theorem:

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &= \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}} \\ &\propto \text{likelihood} \times \text{prior} \end{aligned}$$

Applied Bayes: Statistical Models

- let θ denote a parameter and y the data
- from Bayes' Theorem:

$$
\begin{aligned}
p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\
&= \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}} \\
&\propto \text{likelihood} \times \text{prior}
\end{aligned}
$$

- solution to the general problem of inference

Applied Bayes: Statistical Models

- let θ denote a parameter and y the data
- from Bayes' Theorem:

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &= \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}} \\ &\propto \text{likelihood} \times \text{prior} \end{aligned}$$

- solution to the general problem of inference
- learning about the probability (distribution) of a parameter given the data

Applied Bayes: Statistical Models

- let θ denote a parameter and y the data
- from Bayes' Theorem:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$
$$= \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$
$$\propto \text{likelihood} \times \text{prior}$$

- solution to the general problem of inference
- learning about the probability (distribution) of a parameter given the data
- impossible from a frequentist point of view

Probability Theory

Probability Distributions

Random Variable I

Definition (Random Variable)

Let Ω be the sample space for an experiment. A real-valued function that is defined on Ω is called a **random variable**. The set of values the variable might take is the **distribution** of the random variable.

Random Variable II

Definition (Discrete Random Variable)

We say that a random variable X is a **discrete random variable** or that it has a **discrete distribution**, if X can take only a finite number k of different values or, at most, an infinite sequence of different values.

Random Variable II

Definition (Discrete Random Variable)

We say that a random variable X is a **discrete random variable** or that it has a **discrete distribution**, if X can take only a finite number k of different values or, at most, an infinite sequence of different values.

Definition (Continuous Random Variable)

We say that a random variable X is a **continuous random variable** or that it has a **continuous distribution**, if X can take an uncountably infinite number of possible values.

Random Variable II

Definition (Discrete Random Variable)

We say that a random variable X is a **discrete random variable** or that it has a **discrete distribution**, if X can take only a finite number k of different values or, at most, an infinite sequence of different values.

Definition (Continuous Random Variable)

We say that a random variable X is a **continuous random variable** or that it has a **continuous distribution**, if X can take an uncountably infinite number of possible values.

Note, that a random variable is usually denoted with a capital letter, while its realizations are denoted with lowercase letters.

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.
- Random variable: e.g. number of heads,
  $X =$ Number of Heads

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.
- Random variable: e.g. number of heads,
  $X = \text{Number of Heads}$

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.
- Random variable: e.g. number of heads,
  $X =$ Number of Heads

Consider the sequence $q = HHTTTHTTTH$, then $X(q) = 4$.

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.
- Random variable: e.g. number of heads,
  $X = \text{Number of Heads}$

Consider the sequence $q = HHTTTHTTTH$, then $X(q) = 4$.
Define another random variable as $Y = 10 - X$, the number of tails.

Random Variable - Examples: Coin Toss

- Experiment: toss the coin 10 times.
- Sample space: all possible sequences of of 10 heads and/or tails.
- Random variable: e.g. number of heads, $X =$ Number of Heads

Consider the sequence $q = HHTTTHTTTH$, then $X(q) = 4$. Define another random variable as $Y = 10 - X$, the number of tails. Then, $Y(q) = 6$.

Probability Mass Function

Definition (Probability Mass Function, p.m.f.)

For a discrete random variable X the **probability mass function** of X is defined as a function $f(\cdot)$ such that for every real number x,

$$f(x) = \Pr(X = x) = Pr(s \in \Omega : X(s) = x)$$

Probability Mass Function

Definition (Probability Mass Function, p.m.f.)

For a discrete random variable X the **probability mass function** of X is defined as a function $f(\cdot)$ such that for every real number x,

$$f(x) = \Pr(X = x) = Pr(s \in \Omega : X(s) = x)$$

Probability Mass Function

Definition (Probability Mass Function, p.m.f.)

For a discrete random variable X the **probability mass function** of X is defined as a function $f(\cdot)$ such that for every real number x,

$$f(x) = \Pr(X = x) = Pr(s \in \Omega : X(s) = x)$$

Remarks:

- if $x \notin \Omega \implies f(x) = 0$

Probability Mass Function

Definition (Probability Mass Function, p.m.f.)

For a discrete random variable X the **probability mass function** of X is defined as a function $f(\cdot)$ such that for every real number x,

$$f(x) = \Pr(X = x) = Pr(s \in \Omega : X(s) = x)$$

Remarks:

- if $x \notin \Omega \implies f(x) = 0$
- if the sequence $x_1, x_2, \ldots$ includes all the possible values of X, then $\sum_{i=1}^{\infty} f(x_i) = 1$.

Probability Mass Function

Definition (Probability Mass Function, p.m.f.)

For a discrete random variable X the **probability mass function** of X is defined as a function $f(\cdot)$ such that for every real number x,

$$f(x) = \Pr(X = x) = Pr(s \in \Omega : X(s) = x)$$

Remarks:

- if $x \notin \Omega \implies f(x) = 0$
- if the sequence $x_1, x_2, \ldots$ includes all the possible values of X, then $\sum_{i=1}^{\infty} f(x_i) = 1$.
- $Pr(C \subset \Omega) = \sum_{x_i \in C} f(x_i)$

Discrete Distributions

- Bernoulli: a single coin toss

Discrete Distributions

- Bernoulli: a single coin toss
- Binomial: 'successes' of multiple coin tosses

Discrete Distributions

- Bernoulli: a single coin toss
- Binomial: 'successes' of multiple coin tosses
- Poisson: counts

Discrete Distributions

- Bernoulli: a single coin toss
- Binomial: 'successes' of multiple coin tosses
- Poisson: counts
- ...

Continuous Distributions

- Normal distribution

Continuous Distributions

- Normal distribution
- Beta distribution

Continuous Distributions

- Normal distribution
- Beta distribution
- Gamma distribution

Continuous Distributions

- Normal distribution
- Beta distribution
- Gamma distribution
- χ^2 distribution

Continuous Distributions

- Normal distribution
- Beta distribution
- Gamma distribution
- χ^2 distribution
- t distribtion

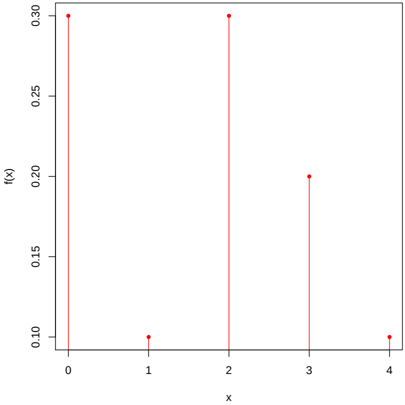
Continuous Distributions

- Normal distribution
- Beta distribution
- Gamma distribution
- χ^2 distribution
- t distribtion
- ...

Example I

A p.m.f. defined as:

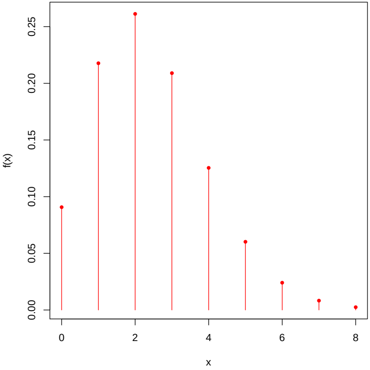
$$f(x) = \begin{cases} 0.3 & if \quad x = 0 \\ 0.1 & if \quad x = 1 \\ 0.3 & if \quad x = 2 \\ 0.2 & if \quad x = 3 \\ 0.1 & if \quad x = 4 \end{cases}$$



Example II

Let $\lambda \in \mathbb{R}_{>0}$ (intensity), the Poisson p.m.f. is defined as

$$f(x;\lambda) = \begin{cases} \frac{\lambda^x exp(-\lambda)}{x!} & \forall x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$



Comments

- p.m.f. (as in c.d.f. / p.d.f.) have parameters which determine the "shape" of the distribution, e.g. the Poisson p.m.f. has one parameter (λ)

Comments

- p.m.f. (as in c.d.f. / p.d.f.) have parameters which determine the "shape" of the distribution, e.g. the Poisson p.m.f. has one parameter (λ)
- parameters can be included in the function definition, e.g. $f(x; \lambda)$

Comments

- p.m.f. (as in c.d.f. / p.d.f.) have parameters which determine the "shape" of the distribution, e.g. the Poisson p.m.f. has one parameter (λ)
- parameters can be included in the function definition, e.g. $f(x; \lambda)$
- another notation for the Poisson p.m.f. is $X \sim \text{Pois}(\lambda)$ (similar notations exists for common other distributions)

Comments

- p.m.f. (as in c.d.f. / p.d.f.) have parameters which determine the "shape" of the distribution, e.g. the Poisson p.m.f. has one parameter (λ)
- parameters can be included in the function definition, e.g. $f(x; \lambda)$
- another notation for the Poisson p.m.f. is $X \sim \mathrm{Pois}(\lambda)$ (similar notations exists for common other distributions)
- some authors use $f(X = x)$ instead of $f(x)$ only.

Cumulative Distribution Function

Definition (Cumulative Distribution Function, c.d.f.)

The **cumulative distribution function** $F(\cdot)$ of a discrete or continuous random variable X is the function

$$F(x) = \Pr(X \leq x), \text{ for } -\infty < x < \infty$$

Cumulative Distribution Function

Definition (Cumulative Distribution Function, c.d.f.)

The **cumulative distribution function** $F(\cdot)$ of a discrete or continuous random variable X is the function

$$F(x) = \Pr(X \leq x), \text{ for } -\infty < x < \infty$$

Properties:

- $F(x)$ is nondecreasing as x increases; i.e., if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.

Cumulative Distribution Function

Definition (Cumulative Distribution Function, c.d.f.)

The **cumulative distribution function** $F(\cdot)$ of a discrete or continuous random variable X is the function

$$F(x) = \Pr(X \leq x), \text{ for } -\infty < x < \infty$$

Properties:

- $F(x)$ is nondecreasing as x increases; i.e., if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Cumulative Distribution Function

Definition (Cumulative Distribution Function, c.d.f.)

The **cumulative distribution function** $F(\cdot)$ of a discrete or continuous random variable X is the function

$$F(x) = \Pr(X \leq x), \text{ for } -\infty < x < \infty$$

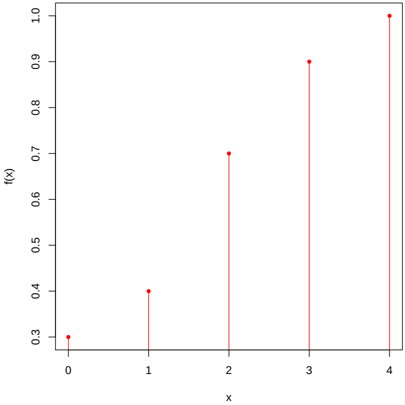
Properties:

- $F(x)$ is nondecreasing as x increases; i.e., if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
- c.d.f. is always continuous from the right, i.e. $F(x) = F(x^+)$ at every point x.

Example I

A c.d.f. defined as:

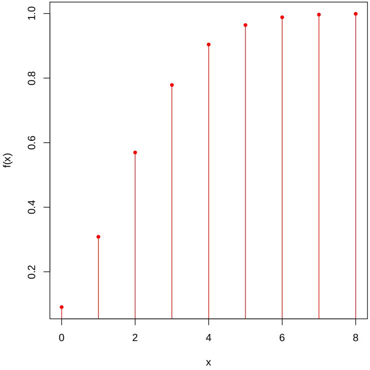
$$F(x) = \begin{cases} 0.3 & \textit{if} \quad x = 0 \\ 0.4 & \textit{if} \quad x = 1 \\ 0.7 & \textit{if} \quad x = 2 \\ 0.9 & \textit{if} \quad x = 3 \\ 1.0 & \textit{if} \quad x = 4 \end{cases}$$



Example II

Let $\lambda \in \mathbb{R}_{>0}$ (intensity), the Poisson c.d.f. is defined as

$$F(x) = exp(-\lambda) \sum_{i=0}^{|k|} \frac{\lambda^i}{i!}, \forall k \leq 0$$



Determining Probabilities from the c.d.f.

Let $F(x^-) = \lim_{y \to x} F(y) \forall y < x$ and
$F(x^+) = \lim_{y \to x} F(y) \forall y > x$.

For any value:

- x, $\Pr(X > x) = 1 - F(x)$

Determining Probabilities from the c.d.f.

Let $F(x^-) = \lim_{y \to x} F(y) \forall y < x$ and
$F(x^+) = \lim_{y \to x} F(y) \forall y > x$.

For any value:

- x, $\Pr(X > x) = 1 - F(x)$
- x_1 and x_2, such that $x_1 < x_2$,
  $\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

Determining Probabilities from the c.d.f.

Let $F(x^-) = \lim_{y \to x} F(y) \forall y < x$ and $F(x^+) = \lim_{y \to x} F(y) \forall y > x$.

For any value:

- x, $\Pr(X > x) = 1 - F(x)$
- x_1 and x_2, such that $x_1 < x_2$, $\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- x, $\Pr(X < x) = F(x^-)$

Determining Probabilities from the c.d.f.

Let $F(x^-) = \lim_{y \to x} F(y) \forall y < x$ and
$F(x^+) = \lim_{y \to x} F(y) \forall y > x$.

For any value:

- x, $\Pr(X > x) = 1 - F(x)$
- x_1 and x_2, such that $x_1 < x_2$,
  $\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- x, $\Pr(X < x) = F(x^-)$
- x, $\Pr(X = x) = F(x) - F(x^-)$

Probability Density Function, p.d.f.

Definition (Probability Density Function)

Let x be a continuous random variable. A p.d.f. is a nonnegative function $f(\cdot)$, defined on the real line, such that:

$$f(x) = F(x)'$$

Probability Density Function, p.d.f.

Definition (Probability Density Function)

Let x be a continuous random variable. A p.d.f. is a nonnegative function $f(\cdot)$, defined on the real line, such that:

$$f(x) = F(x)'$$

Remarks:

- $f(x) \geq 0, \forall x$

Probability Density Function, p.d.f.

Definition (Probability Density Function)

Let x be a continuous random variable. A p.d.f. is a nonnegative function $f(\cdot)$, defined on the real line, such that:

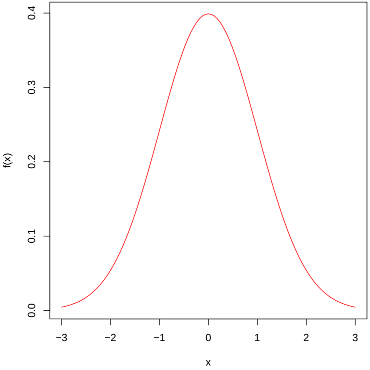
$$f(x) = F(x)'$$

Remarks:

- $f(x) \geq 0, \forall x$
- $\int_a^b f(x)dx = 1$ where a, b are the bounds of the suppport for x

Example I

The p.d.f. of a normal (or Gaussian) distribution is defined as $f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$ where $\mu \in \mathbb{R}$ (mean) and $\sigma^2 \in \mathbb{R}_{>0}$ (variance). For the standard normal (picture) $\mu = 0$ and $\sigma^2 = 1$.



Probability Theory

Properties of Distributions

Expectation I

Definition (Expectation)

Let X be a discrete random variable with a p.m.f. $f(\cdot)$. The **expectation** (also: expected value, mean) of X, denoted $E(X)$ is a scalar defined as $E(X) = \sum_x xf(x)$. Similarly, if X is a continuous random variable, the **expectation** is a scalar defined as $E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx$.

Variance

Definition (Variance)

Let X be a random variable with mean $\mu = E(X)$. The variance of X denoted by $Var(x)$ is defined as: $Var(x) = E\big((X - \mu)^2\big)$.

Properties:

- $Var(aX + b) = a^2 Var(X)$

Remark: For some distributions, the variance is infinite (e.g. Pareto with $\alpha = 0.5$).

Variance

Definition (Variance)

Let X be a random variable with mean $\mu = E(X)$. The variance of X denoted by $Var(x)$ is defined as: $Var(x) = E\big((X-\mu)^2\big)$.

Properties:

- $Var(aX + b) = a^2 Var(X)$
- $Var(X) = E(X^2) - (E(X))^2$

Remark: For some distributions, the variance is infinite (e.g. Pareto with $\alpha = 0.5$).

Variance

Definition (Variance)

Let X be a random variable with mean $\mu = E(X)$. The variance of X denoted by $Var(x)$ is defined as: $Var(x) = E\big((X-\mu)^2\big)$.

Properties:

- $Var(aX + b) = a^2 Var(X)$
- $Var(X) = E(X^2) - (E(X))^2$
- $Var(X + Y) = Var(X) + Var(Y)$ iff (X, Y) are independent

Remark: For some distributions, the variance is infinite (e.g. Pareto with $\alpha = 0.5$).