Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Sets: Basics and Vocabulary

Video companion

1 Set theory basics

- What is a set?
- Cardinality (size)
- Intersections
- Unions

2 What is a set?

Vocab: A *set* is made up of *elements*.

Example: $A = \{1, 2, -3, 7\}$ and $E = \{\text{apple, monkey, Daniel Egger}\}$

- $2 \in A$: "2 is an element of A"
- $8 \notin A$: "8 is NOT an element of A"

3 Cardinality

Vocab: The *cardinality* (size) of a set is the number of elements in it.

- $|A| = 4$ (there are 4 elements in A, so the cardinality is 4)
- $|E| = 3$ (there are 3 elements in E, so the cardinality is 3)

4 Intersections

The *intersection* is defined as elements that are in both sets.

Symbol $\cap$: "intersects" (and)

Example: $A = \{1, 2, -3, 7\}$ and $B = \{2, -3, 8, 10\}$ and $D = \{5, 10\}$

- $A \cap B = \{2, -3\}$
- $B \cap D = \{10\}$

In general, $A \cap B = \{x : x \in A \textit{ and } x \in B\}$

If there are no elements in common, the answer is the empty set $\emptyset$. The cardinality of the empty set $|\emptyset| = 0$.

- $A \cap D = \emptyset$

5 Unions

The *union* is defined as elements that are in either set.

Symbol $\cup$: "union" (or)

Example: $A = \{1, 2, -3, 7\}$ and $B = \{2, -3, 8, 10\}$ and $D = \{5, 10\}$

- $A \cup B = \{1, 2, -3, 7, 8, 10\}$
- $A \cup D = \{1, 2, -3, 7, 5, 10\}$

In general, $A \cup B = \{x \in A \textit{ or } x \in B\}$.

Sets: Medical Testing Example

Video companion

1 Example using set theory

VBS: "very bad syndrome"

X = set of people in a clinical trial

$S = \{x \in X : x \text{ has VBS}\}$
$H = \{x \in X : x \text{ does not have VBS}\}$

$X = S \cup H$ (you either have VBS or you don't)
$S \cap H = \emptyset$ (no one both has and doesn't have it)

Point of medical testing to figure out whether a person is in S or in H

2 Test

$P = \{x \in X : x \text{ tests positive for VBS}\}$
$N = \{x \in X : x \text{ tests negative for VBS}\}$

$P \cup N = X$ (you either test positive or negative)
$P \cap N = \emptyset$ (no one tests both positive and negative)

In a perfect world, S would equal P—the sick people would always test positive, and H would equal N—the healthy people would always test negative.
...but this is not always the case.

$S \cap P$	$H \cap N$	$S \cap N$	$H \cap P$
true positive	true negative	false negative	false positive
Accurate	Accurate	Inaccurate	Inaccurate

3 Cardinality

$\frac{|S|}{|X|}$ = proportion of people in the study who do genuinely have VBS
$\frac{|H|}{|X|}$ = proportion of people in the study without VBS
$\frac{|S|}{|X|} + \frac{|H|}{|X|} = 1$

$\frac{\lvert S\cap P\rvert}{\lvert S\rvert}$	true positive rate	would like to be close to 1
$\frac{\lvert H\cap P\rvert}{\lvert H\rvert}$	false positive rate	would like to be as small as possible
$\frac{\lvert S\cap N\rvert}{\lvert S\rvert}$	false negative rate	would like to be as small as possible
$\frac{\lvert H\cap N\rvert}{\lvert H\rvert}$	true negative rate	would like to be close to 1

Sets: Venn Diagrams

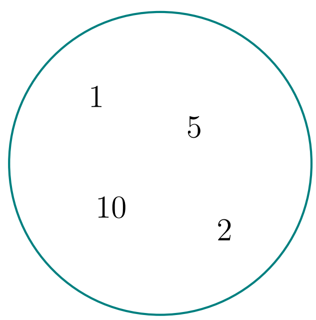
Video companion

1 Visualizing sets

- Venn diagrams
- Inclusion-exclusion formula
- Medical testing example, re-visited

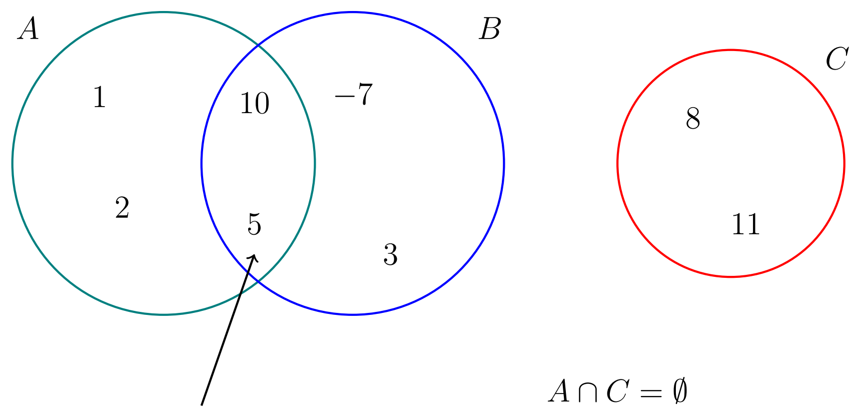
2 Single set

$A = \{1, 5, 10, 2\} \qquad |A| = 4$

$A =$ 

3 Multiple sets

$A = \{1, 5, 10, 2\}$
$B = \{5, -7, 10, 3\}$
$C = \{8, 11\}$

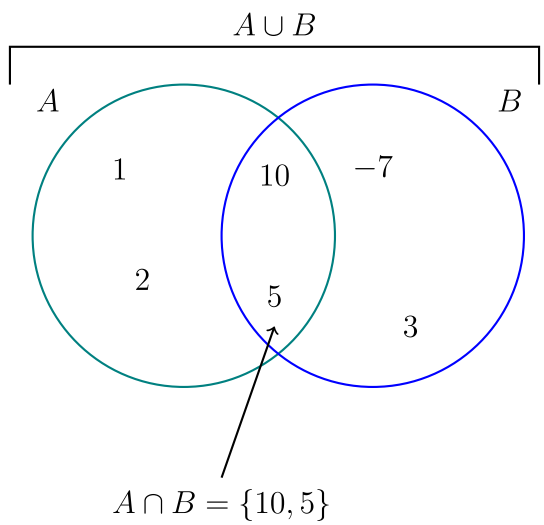


$A \cap B = \{10, 5\}$

$A \cap C = \emptyset$

$B \cap C = \emptyset$

4 Inclusion-exclusion formula



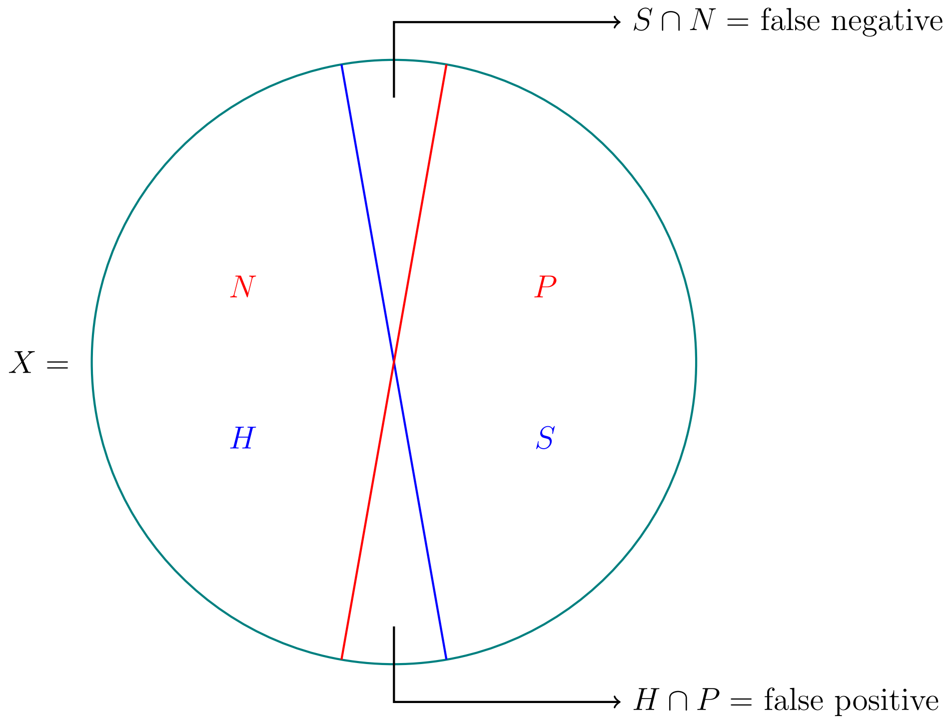
Inclusion-exclusion formula:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Check with this example:

$$6 \stackrel{?}{=} 4 + 4 - 2$$
$$6 = 6 \quad \checkmark$$

5 Medical testing example



$$X = H \cup S \qquad H \cap S = \emptyset$$
$$S = N \cup P \qquad N \cap P = \emptyset$$

Numbers: The Real Number Line

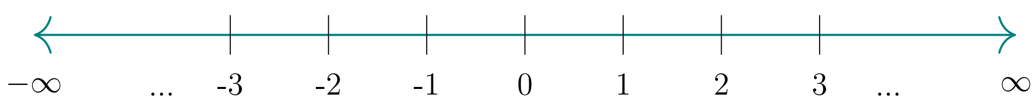
Video companion

1 Introduction

- What is $\mathbb{R}$?
- Positive, negative
- Absolute value

2 Integers and rational numbers

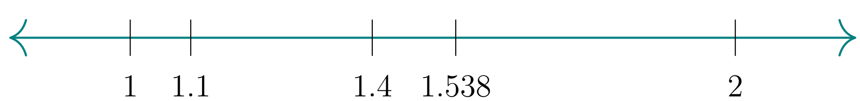
Graph of $\mathbb{R}$, the real numbers:



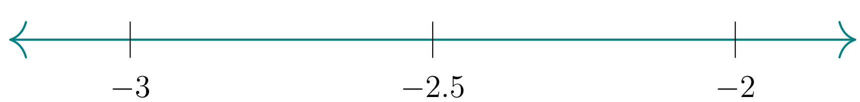
Subset of real numbers, integers:

$\mathbb{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\}$

Segment between 1 and 2:

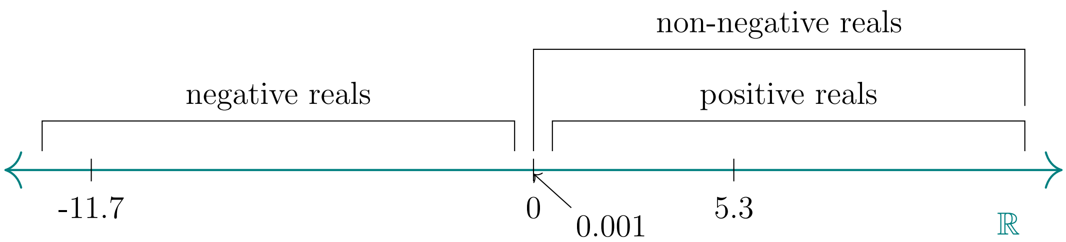


Segment between -3 and -2:

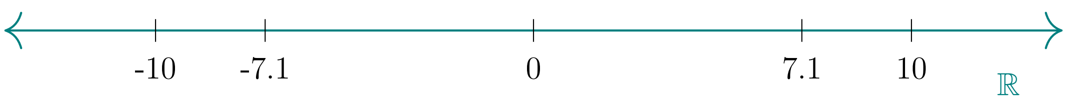


Some real numbers terminate, and some do not.
The number $\pi = 3.14159...$ is *irrational*, i.e. it does not repeat after the decimal point.

3 Sets of real numbers



4 Absolute value



The absolute value of a number x, $|x|$, is the distance from x to 0.

Example:

$$\begin{aligned} |7.1| &= 7.1 \\ |-7.1| &= 7.1 = -(-7.1) \end{aligned}$$

General rule:
For any $x \in \mathbb{R}$,

$$|x| = \begin{cases} x, & \text{if } x \text{ is non-negative} \\ -x, & \text{if } x \text{ is negative} \end{cases}$$

Check:

$$\begin{aligned} |8.7| &= 8.7 \\ |-10| &= -(-10) = 10 \end{aligned}$$

Numbers: Greater-than and Less-than

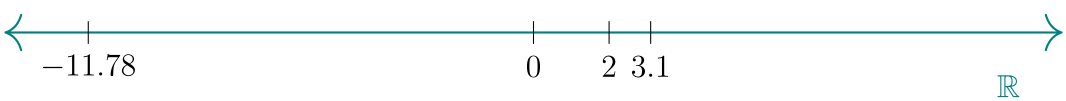
Video companion

1 Inequalities, basic idea

Introduction to symbols:

$a < b$	"a is less than b"
$x > y$	"x is greater than y"
$c \leq d$	"c is less than or equal to d"
$z \geq w$	"z is greater than or equal to w"
$e \ll f$	"e is much, much less than f"

2 Inequality on the real number line



$2 < 3.1$	"2 is to the left of 3.1 on the real number line"
$-11.78 < 3.1$	"−11.78 is to the left of 3.1 on the real number line"

For any $a < b$, a must be to the left of b on the real number line.

$3.1 > 2$	"3.1 is to the right of 2 on the real number line"

In general, a is less than b, if, and only if, b is greater than a:

$$\boxed{a < b \iff b > a}$$

3 Much, much less than

$x \ll y$ "x is much, much less than y"
(Not proper math, but used frequently in data science)

For example, $1 \ll 1,000,000$, which is reasonable but not possible to prove "true"

4 Less than or equal to

$a \leq b$ means $a < b$ or $a = b$

Examples:

Is $2 \leq 3.1$ true?

$$\left.\begin{array}{ll} 2 < 3.1 & \checkmark \\ 2 = 3.1 & \times \end{array}\right] \checkmark$$

Is $2 \leq 2$ true?

$$\left.\begin{array}{ll} 2 < 2 & \times \\ 2 = 2 & \checkmark \end{array}\right] \checkmark$$

Is $2 \leq 0.8$ true?

$$\left.\begin{array}{ll} 2 < 0.8 & \times \\ 2 = 0.8 & \times \end{array}\right] \times$$

Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Numbers: Algebra with Inequalities

Video companion

1 Introduction

- Review algebra with equalities (=)
  - how?
  - why?
- Learn algebra with inequalities ($<, >, \leq, \geq$)
  - what works
  - A BIG WARNING

2 Algebra with equalities

$$\begin{aligned} 4 &= 4 \\ 4+3 &= 4+3 \\ 7 &= 7 \quad \checkmark \end{aligned}$$

Rule:
If $a = b$, then $a + c = b + c$.

Example:

$$\begin{aligned} x+3 &= 10 \\ (x+3)-3 &= 10-3 \\ x &= 7 \end{aligned}$$

Similarly with multiplication,

$$\begin{aligned} 4 &= 4 \\ 2 \cdot 4 &= 2 \cdot 4 \\ 8 &= 8 \quad \checkmark \end{aligned}$$

$$
\begin{aligned}
4 &= 4 \\
(-3) \cdot 4 &= (-3) \cdot 4 \\
-12 &= -12 \quad \checkmark
\end{aligned}
$$

Rule:

If a, b, and c are numbers, and $c \neq 0$, and $a = b$, then $c \cdot a = c \cdot b$.

Example:

$$
\begin{aligned}
-5x &= 15 \\
\left(-\frac{1}{5}\right) \cdot (-5x) &= \left(-\frac{1}{5}\right) \cdot 15 \\
x &= -3
\end{aligned}
$$

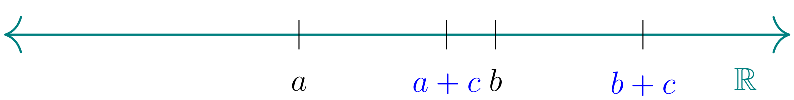
3 Algebra with inequalities

$$
\begin{aligned}
4 &< 7 \\
4 + 2 &\stackrel{?}{<} 7 + 2 \\
6 &\stackrel{?}{<} 9 \quad \checkmark
\end{aligned}
$$

$$
\begin{aligned}
4 &< 7 \\
4 - 1 &\stackrel{?}{<} 7 - 1 \\
3 &\stackrel{?}{<} 6 \quad \checkmark
\end{aligned}
$$

Rule:

If $a < b$, then $a + c < b + c$.



Example:

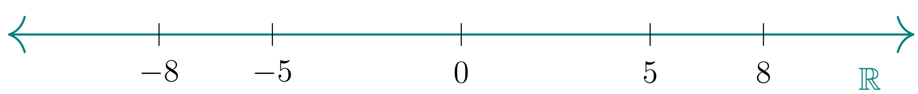
$$\begin{aligned} x + 3 &< 10 \\ (x+3) - 3 &< 10 - 3 \\ x &< 7 \end{aligned}$$

$x \in (-\infty, 7)$

Test cases with multiplication:

$$\begin{aligned} 5 &< 8 \\ 3 \cdot 5 &\overset{?}{<} 3 \cdot 8 \\ 15 &\overset{?}{<} 40 \quad \checkmark \end{aligned}$$

$$\begin{aligned} 5 &< 8 \\ (-1) \cdot 5 &\overset{?}{<} (-1) \cdot 8 \\ -5 &\overset{?}{<} -8 \quad \times \\ -5 &> -8 \quad ! \end{aligned}$$



Rule:

Suppose $a < b$.

If $c > 0$, then $a \cdot c < b \cdot c$.

If $c < 0$, then $a \cdot c > b \cdot c$.

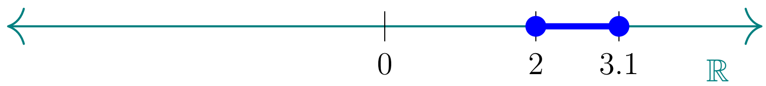
Example:

$$\begin{aligned} -2x &< 10 \\ \left(-\frac{1}{2}\right) \cdot (-2x) &> \left(-\frac{1}{2}\right) \cdot 10 \\ x &> -5 \end{aligned}$$

Numbers: Intervals and Interval Notation

Video companion

1 Closed intervals

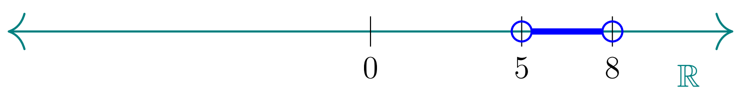


Real number line is an infinite set. There are also infinite subsets.

$[2, 3.1] = \{x \in \mathbb{R} : 2 \leq x \leq 3.1\}$

$$\begin{aligned} 2.3 &\in [2, 3.1] \qquad \text{because } 2 \leq 2.3 \leq 3.1 \\ 3 &\in [2, 3.1] \\ 3.1 &\in [2, 3.1] \\ 1 &\notin [2, 3.1] \qquad \text{because } 2 \not\leq 1 \leq 3.1 \end{aligned}$$

2 Open intervals

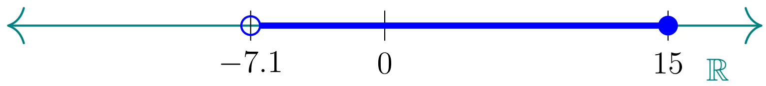


$(5, 8) = \{x \in \mathbb{R} : 5 < x < 8\}$

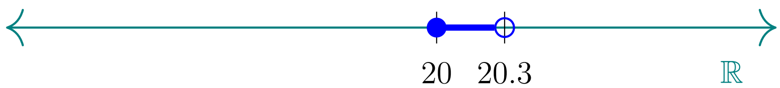
$$\begin{aligned} 5.5 &\in (5, 8) \qquad \text{because } 5 < 5.5 < 8 \\ 5.0001 &\in (5, 8) \\ 5 &\notin (5, 8) \qquad \text{because } 5 \not< 5 < 8 \end{aligned}$$

The intervals $[5, 8]$ and $(5, 8)$ differ at exactly two numbers: 5 and 8.

3 Half-open intervals



$(-7.1, 15] = \{x \in \mathbb{R} : -7.1 < x \leq 15\}$

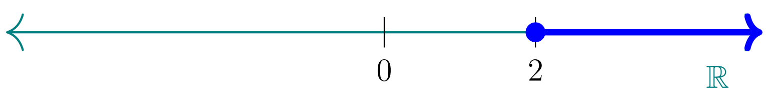


$[20, 20.3) = \{x \in \mathbb{R} : 20 \leq x < 20.3\}$

4 Recap vocabulary

- Closed intervals $[2, 3.1]$
- Open intervals $(5, 8)$
- Half-open intervals $(2, 3]$, $[20, 20.3)$

5 Rays



$[2, \infty) = \{x \in \mathbb{R} : x \geq 2\}$

Another example:
$(-\infty, 7.1) = \{x \in \mathbb{R} : x < 7.1\}$

6 What does an "answer" mean?

Solving an equality gives you a number:

$$\begin{aligned} x + 5 &= 10 \\ x &= 5 \end{aligned}$$

Solving an inequality give you an interval:

$$\begin{aligned} 1 \leq x + 5 &< 10 \\ -4 \leq x &< 5 \\ x &\in [-4, 5) \end{aligned}$$

Sigma Notation: Introduction to Summation

Video companion

1 Sigma notation (Σ)

Examples that will be seen in this video:

$$\sum_{i=1}^{4} i^2 = 30$$

$$\sum_{i=1}^{5} (2i+3) = 45$$

$$\sum_{j=3}^{7} \frac{j}{2} = \frac{25}{2}$$

2 First example

Example:

$$\begin{aligned}\sum_{i=1}^{4} i^2 &= 1^2 + 2^2 + 3^2 + 4^2 \\ &= 30\end{aligned}$$

$i = 1$ on bottom tells us to *start* with $i = 1$.
4 on top tells us to *finish* with $i = 4$.
Implicitly know that you increment by 1.

For each number i that you count,

$$\begin{aligned}i = 1 &: i^2 = 1^2 \\ i = 2 &: i^2 = 2^2 \\ i = 3 &: i^2 = 3^2 \\ i = 4 &: i^2 = 4^2\end{aligned}$$

then the Σ tells you to *sum* the results.

3 Second example

Example:

$$\begin{aligned}\sum_{i=1}^{5}(2i+3) &= (2(1)+3)+(2(2)+3)+(2(3)+3)+(2(4)+3)+(2(5)+3) \\ &= 45\end{aligned}$$

Work for problem:

$$\begin{aligned}&i=1: 2i+3 = 2(1)+3 \\ &i=2: 2i+3 = 2(2)+3 \\ &i=3: 2i+3 = 2(3)+3 \\ &i=4: 2i+3 = 2(4)+3 \\ &i=5: 2i+3 = 2(5)+3\end{aligned}$$

4 Third example

Example:

$$\begin{aligned}&\sum_{j=3}^{7}\frac{j}{2} = \frac{3}{2}+\frac{4}{2}+\frac{5}{2}+\frac{6}{2}+\frac{7}{2} = \frac{25}{2} \\ &\sum_{r=3}^{7}\frac{r}{2} = \frac{25}{2}\end{aligned}$$

j and r are "dummy indices," symbols for counters.

$$\sum_{☺=3}^{7}\frac{☺}{2} = \frac{25}{2}$$

Common choices for indices:
i, j, k, l, r, m, n

Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Sigma Notation: Simplification Rules

Video companion

1 Distributive property

Examples:

$$\sum_{i=1}^{4} i^2 = 30$$

$$\begin{aligned}\sum_{i=1}^{4} 3i^2 &= 3(1)^2 + 3(2)^2 + 3(3)^2 + 3(4)^2\\ &= 3[1^2 + 2^2 + 3^2 + 4^2]\\ &= 3\left[\sum_{i=1}^{4} i^2\right]\end{aligned}$$

$$\sum_{r=4}^{25} 18r^3 = 18\left[\sum_{r=4}^{25} r^3\right]$$

This is due to the *distributive property*:

$$a(b+c) = ab + ac$$

In other words, constants inside the summed expression can be pulled outside.

2 Commutative property

$$\begin{aligned}\sum_{i=1}^{4}(i^2+2i) &= \big(1^2+2(1)\big)+\big(2^2+2(2)\big)+\big(3^2+2(3)\big)+\big(4^2+2(4)\big)\\ &= \big(1^2+2^2+3^2+4^2\big)+\left(2(1)+2(2)+2(3)+2(4)\right)\\ &= \left(\sum_{i=1}^{4} i^2\right)+\left(\sum_{i=1}^{4} 2i\right)\end{aligned}$$

This is due to the *commutative property*:

$$a+b=b+a$$

In other words, we can add the terms in any order.

3 Summation of constants

Examples:

$$\begin{aligned}\sum_{k=1}^{10} 5 &= 5+5+5+5+5+5+5+5+5+5\\ &= 10\cdot 5\\ &= 50\end{aligned}$$

$$\begin{aligned}\sum_{r=1}^{7} 8 &= 8+8+8+8+8+8+8\\ &= 7\cdot 8\\ &= 56\end{aligned}$$

When summing constants, you can multiply the constant by the number of indices you count.

Sigma Notation: Mean and Variance

Video companion

1 Introduction

Important equations for this video:

$$
\begin{aligned}
X &= \{x_1, ..., x_n\} \\
\mu_x &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
{\sigma_x}^2 &= \frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \mu_x)^2\right]
\end{aligned}
$$

The symbol μ_x is the "mean of x," and ${\sigma_x}^2$ is the "variance of x." The standard deviation is denoted σ_x.

2 Mean

Example:

$$
\begin{aligned}
Z &= \{1, 5, 12\} \\
|Z| &= 3 \\
\mu_z &= \frac{1+5+12}{3} = \frac{18}{3} = 6
\end{aligned}
$$

The mean μ_z is also denoted $\mu(z)$ or simply μ.

Symbolic example:

$$
\begin{aligned}
Y &= \{y_1, y_2, y_3, y_4\} \\
\mu_y &= \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \\
&= \frac{1}{4}\left(\sum_{i=1}^{4} y_i\right)
\end{aligned}
$$

In general, suppose you have a set

$$X = \{x_1, x_2, ..., x_n\},$$

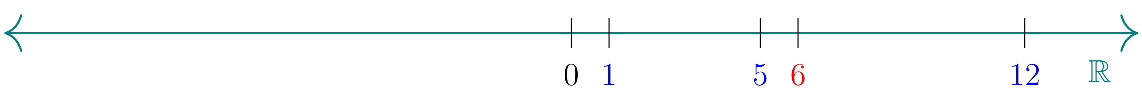
then the mean of X is

$$\mu_x = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right).$$

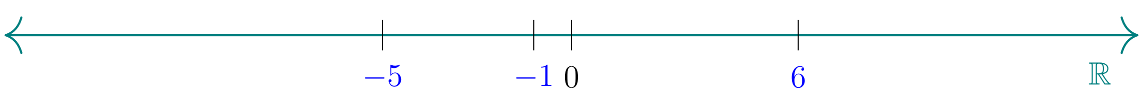
The variable i is a counter. The variable n is a number, which tells you when to stop counting.

3 Mean centering

$$Z = \{1, 5, 12\}$$
$$\mu_z = 6$$



$$\begin{aligned} Z' &= \{1-6, 5-6, 12-6\} \\ &= \{-5, -1, 6\} \end{aligned}$$
$$\mu_{z'} = 0$$

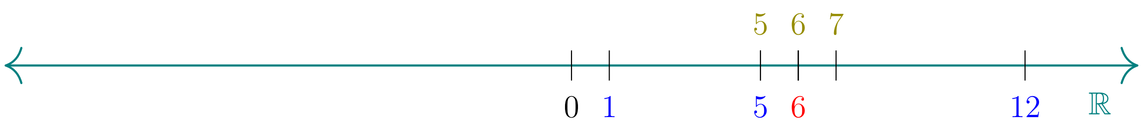


Mean centering data produces a new data set, which has the same relationships, but the mean is zero.

4 Variance

$Z = \{1, 5, 12\}$
$\mu_z = 6$

$W = \{5, 6, 7\}$
$\mu_w = 6$



Set Z (blue) is more "spread out" than set W (olive).

If $X = \{x_1, ..., x_n\}$, the variance of X is

$$\sigma_x{}^2 = \frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \mu_x)^2\right].$$

The standard deviation is given by

$$\sigma_x = \sqrt{\sigma_x{}^2}.$$

Z and W have the same mean, but Z is more spread out, so σ_z should be greater than σ_w.

$$\begin{aligned}
\sigma_w^2 &= \frac{1}{3}\left[\sum_{i=1}^{3}(w_i - \mu_w)^2\right] \\
&= \frac{1}{3}\left[(5-6)^2 + (6-6)^2 + (7-6)^2\right] \\
&= \frac{1}{3}\left[(-1)^2 + 0^2 + 1^2\right] \\
&= \frac{2}{3} \\
\sigma_w &= \sqrt{\frac{2}{3}}
\end{aligned}$$

$$
\begin{aligned}
\sigma_z^2 &= \frac{1}{3}\left[\sum_{i=1}^{3}(z_i - \mu_z)^2\right] \\
&= \frac{1}{3}\left[(1-6)^2 + (5-6)^2 + (12-6)^2\right] \\
&= \frac{1}{3}\left[(-5)^2 + (-1)^2 + 6^2\right] \\
&= \frac{62}{3} \\
\sigma_z &= \sqrt{\frac{62}{3}}
\end{aligned}
$$

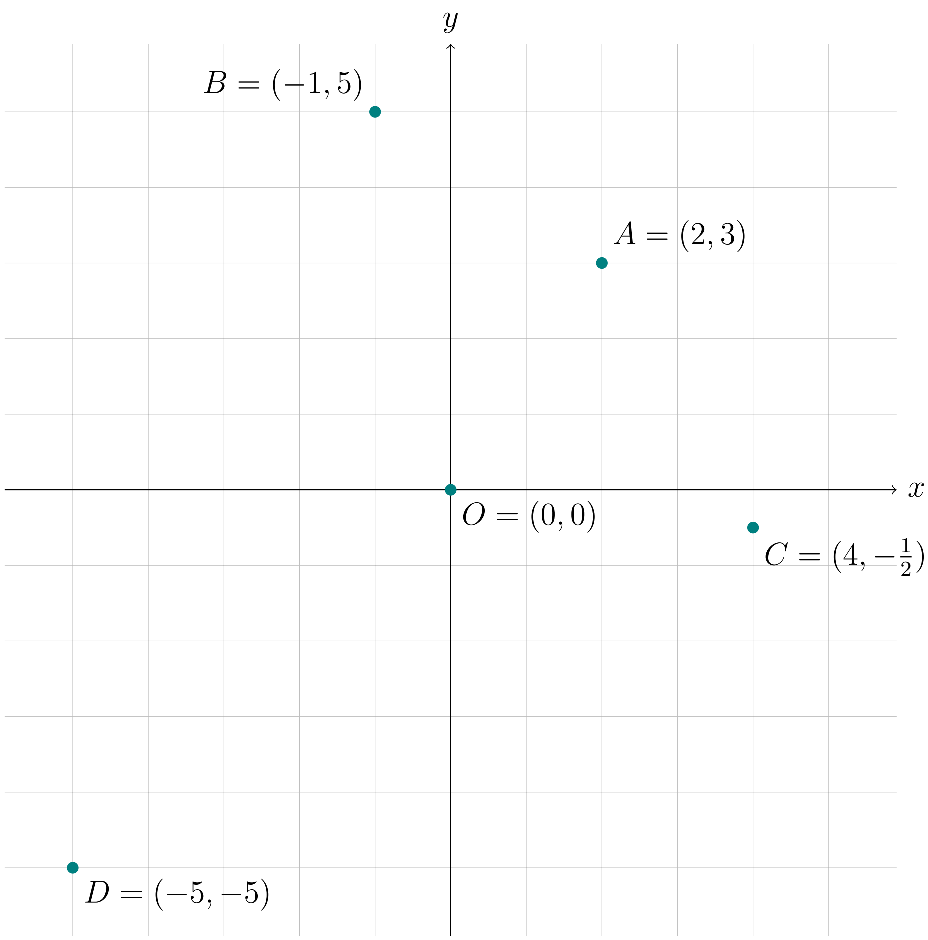
$\sigma_z^2 \gg \sigma_w^2$, so Z is much more spread out than W.

Cartesian Plane: Plotting Points

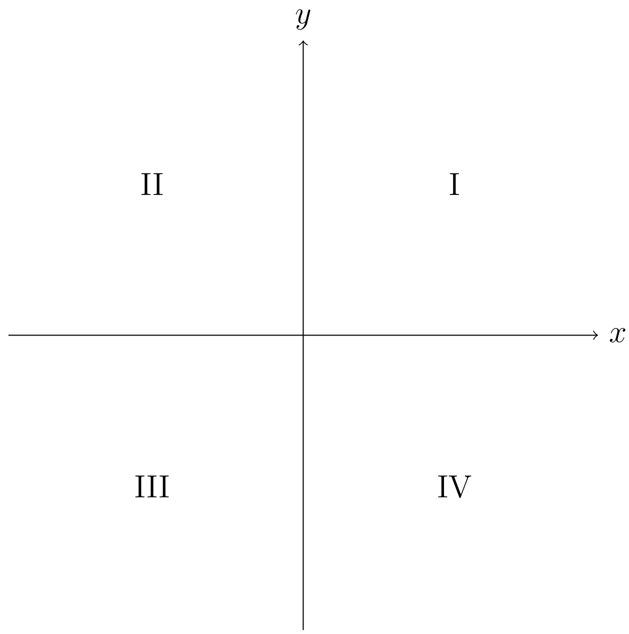
Video companion

1 Introduction

Cartesian plane denoted $\mathbb{R}^2$



2 Axes and quadrants



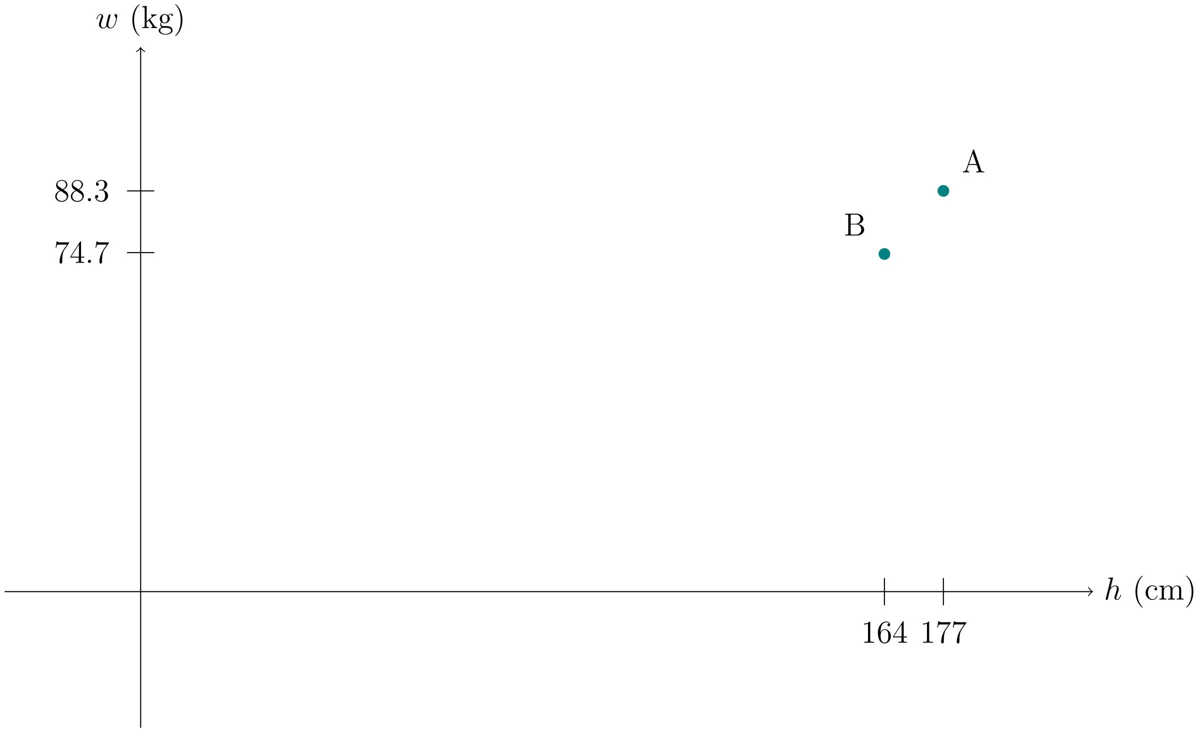
$$x\text{-axis} = \left\{(x,y) \in \mathbb{R}^2 : y = 0\right\}$$
$$y\text{-axis} = \left\{(x,y) \in \mathbb{R}^2 : x = 0\right\}$$

$$\text{first quadrant} = \left\{(x,y) \in \mathbb{R}^2 : x > 0, y > 0\right\}$$
$$\text{second quadrant} = \left\{(x,y) \in \mathbb{R}^2 : x < 0, y > 0\right\}$$
$$\text{third quadrant} = \left\{(x,y) \in \mathbb{R}^2 : \qquad\qquad\qquad \right\}$$
$$\text{fourth quadrant} = \left\{(x,y) \in \mathbb{R}^2 : \qquad\qquad\qquad \right\}$$

3 Real-world example

Table of height and weight:

	h (cm)	w (kg)
A	177	88.3
B	164	74.7



Cartesian Plane: Distance Formula

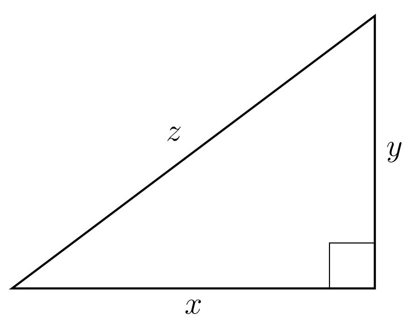
Video companion

1 Introduction

In this video:

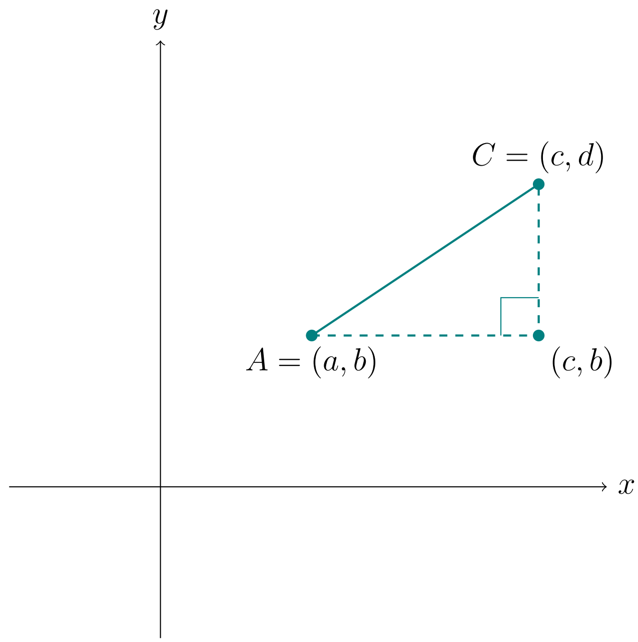
- The distance formula
- Nearest neighbors
- Clustering

2 Pythagorean theorem



$$z^2 = x^2 + y^2$$
$$z = \sqrt{x^2 + y^2}$$

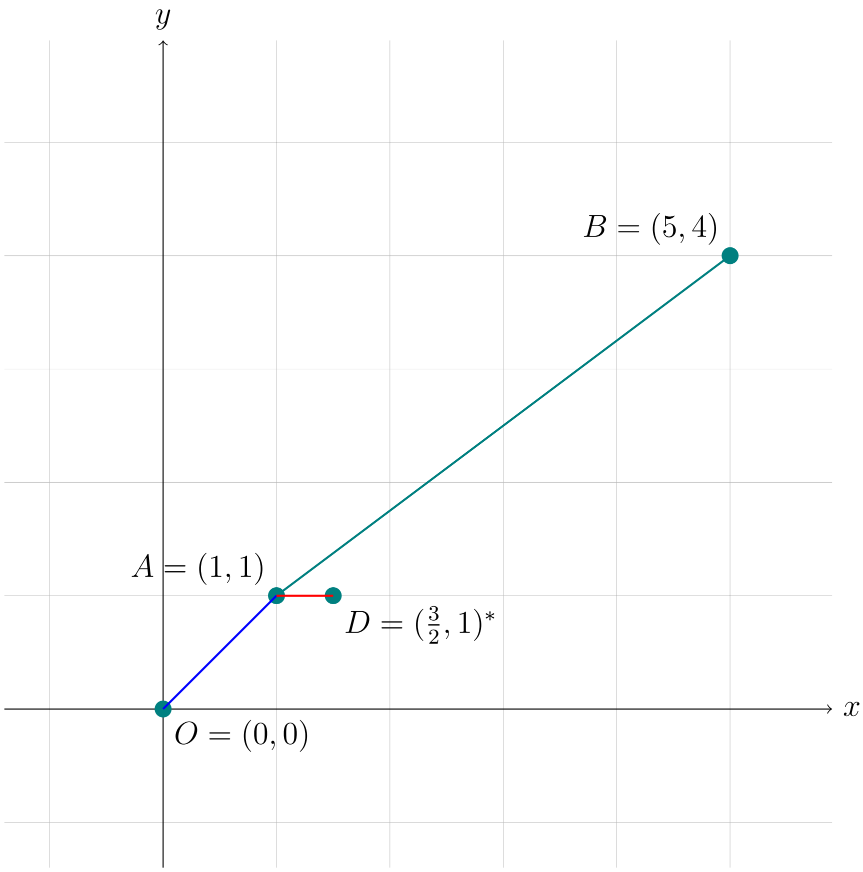
3 Graph of distance formula



Distance formula:

$$\text{dist}(A, C) = \sqrt{(c-a)^2 + (d-b)^2}$$

4 Example and nearest neighbors



$$\begin{aligned}\text{dist}(A, B) &= \sqrt{(5-1)^2 + (4-1)^2} \\ &= 5\end{aligned}$$

$$\begin{aligned}\text{dist}(A, O) &= \sqrt{(1-0)^2 + (1-0)^2} \\ &= \sqrt{2} \approx 1.4\end{aligned}$$

$$\begin{aligned}\text{dist}(A, D) &= \sqrt{\left(\frac{3}{2} - 1\right)^2 + (1-1)^2} \\ &= \frac{1}{2}\end{aligned}$$

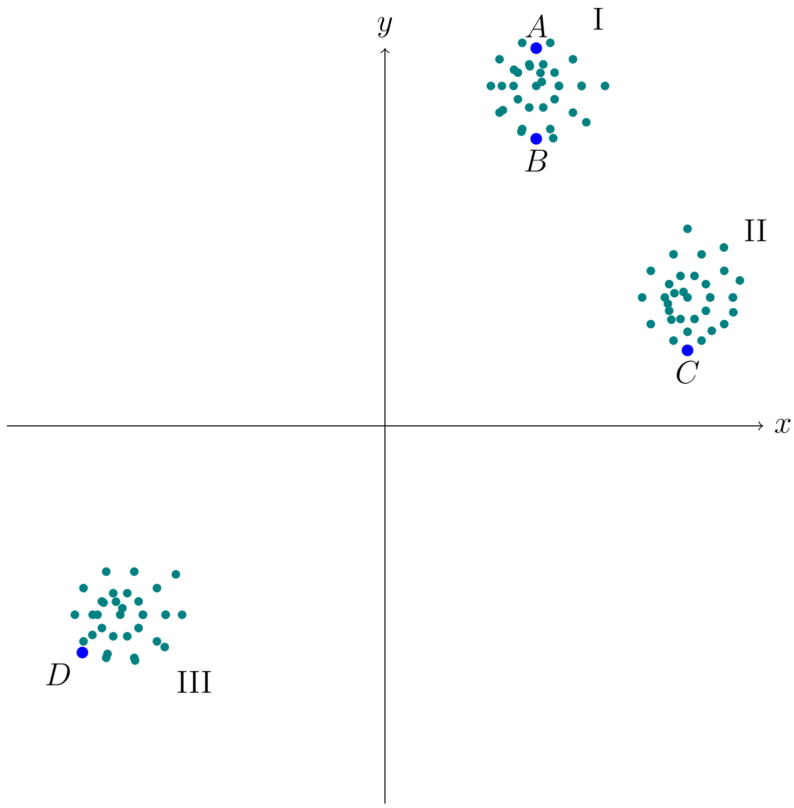
*Note that the x and y values of point D are reversed in the video, but it does not matter in calculating the distance from A.

Consider set S:

$$S = \{O, B, D\}$$

The *nearest neighbor* of A in S is D.
The second nearest neighbor of A in S is O.
The third nearest neighbor of A in S is B.

5 Clustering



Three clusters: I, II, and III

If A and B are in cluster I,
and C is in cluster II,
and D is in cluster III,

Then $\text{dist}(A, B) << \text{dist}(A, C)$,
$<< \text{dist}(A, D)$

Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Cartesian Plane: Point-Slope Formula for Lines

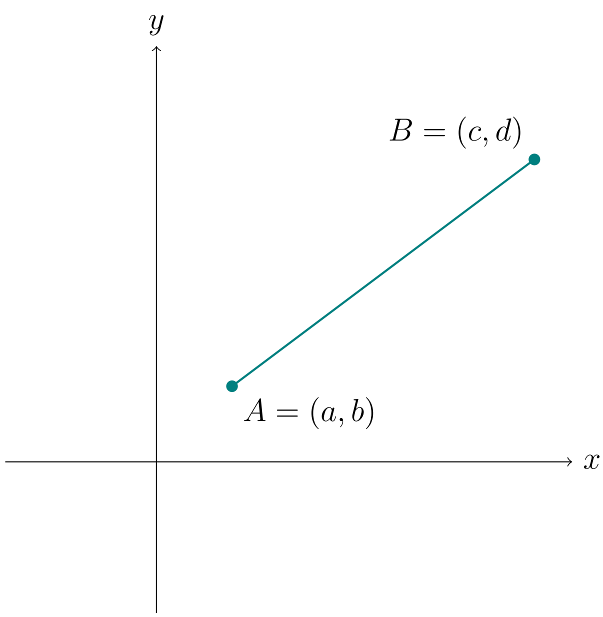
Video companion

1 Introduction

In this video: Demystify formulas for equations of lines

$$
\begin{array}{ll}
y - y_0 = m(x - x_0) & \text{Point-slope form} \\
y = mx + b & \text{Slope-intercept form}
\end{array}
$$

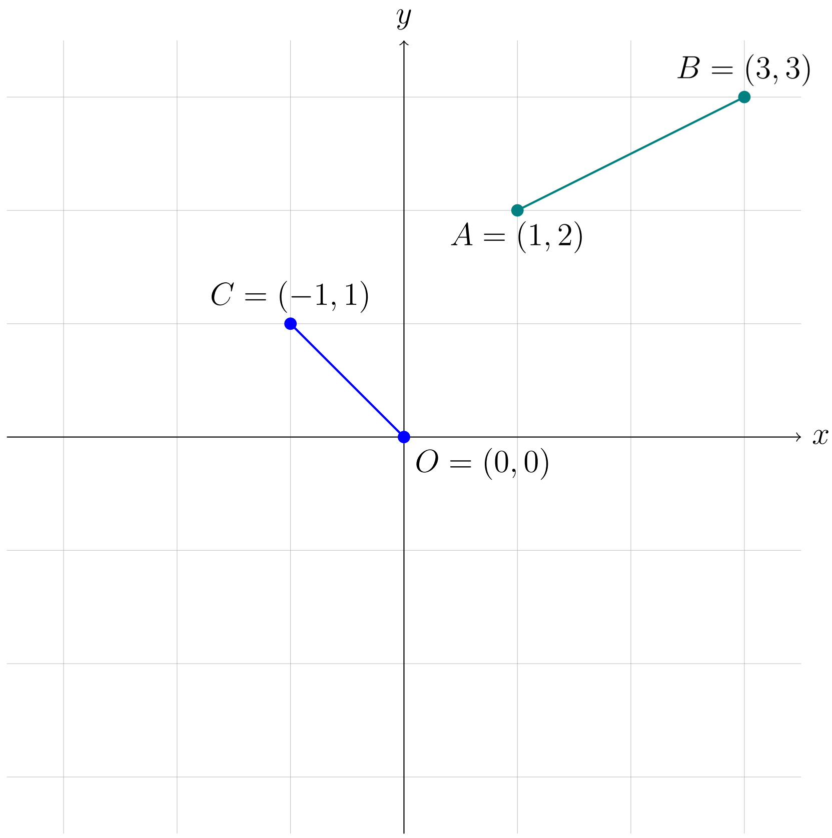
2 Slope of a line segment



Slope of $\overrightarrow{AB}$:

$$m = \frac{d-b}{c-a} = \frac{\text{“rise”}}{\text{“run”}}$$

3 Examples



Slope of $\overrightarrow{AB}$:

$$m = \frac{3-2}{3-1} = \frac{1}{2}$$

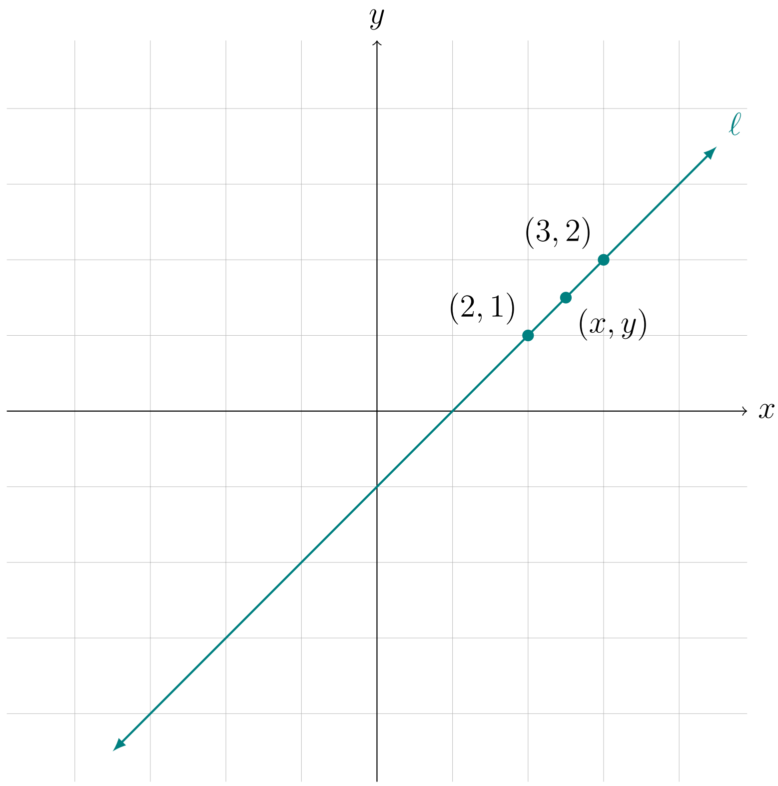
$m = \frac{1}{2}$ is a positive slope.

Slope of $\overrightarrow{CO}$:

$$m = \frac{0-1}{0-(-1)} = -1$$

$m = -1$ is a negative slope.

4 Equation of a line



For a point (x, y) to be on the line, the line segment from $(2, 1)$ to (x, y) need to have a slope of 1.

$$1 = \frac{y-1}{x-2}$$
$$y - 1 = 1(x - 2)$$

The line is defined by this formula:

$$\ell = \left\{(x, y) \in \mathbb{R}^2 : y - 1 = 1(x - 2)\right\}$$

Check that $(3, 2)$ is on the line:

$$(3, 2) \in \ell\,?$$
$$2 - 1 \stackrel{?}{=} 1(3 - 2)$$
$$1 \stackrel{?}{=} 1 \quad \checkmark$$

Check if $(5,1)$ is on the line:

$$\begin{aligned}(5,1) &\in \ell\,?\\ 1-1 &\stackrel{?}{=} 1(5-2)\\ 0 &\stackrel{?}{=} 3 \quad \times\end{aligned}$$

5 Point-slope formula

If a line ℓ has slope m, *and* if (x_0, y_0) is *any* point on ℓ, then ℓ has the equation

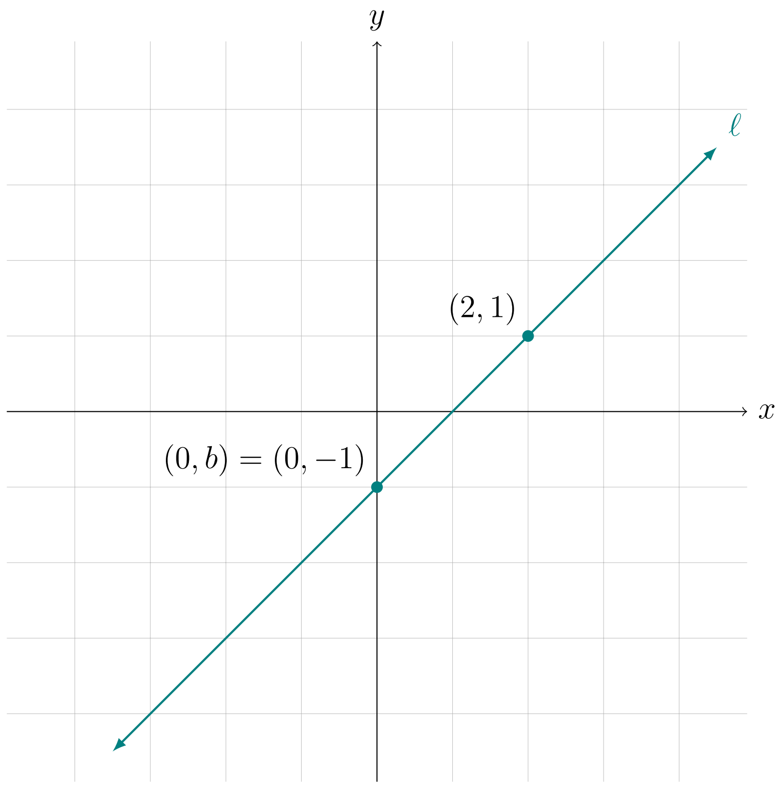
$$\boxed{y - y_0 = m(x - x_0).}$$

Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Cartesian Plane: Slope-Intercept Formula for Lines

Video companion

1 Derivation using point-slope form



From last video, the equation of a line in point-slope form that passes through $(2, 1)$ and has slope $m = 1$ is

$$y - 1 = 1(x - 2).$$

The y-intercept is at point $(0, b)$. To find b, we substitute that point into the definition of the line:

$(0, b) \in \ell$, so

$$b - 1 = 1(0 - 2)$$
$$b = -1$$

Using the y-intercept in the equation for the line in point-slope form:

$$\begin{aligned} y - (-1) &= 1(x - 0) \\ y + 1 &= x \\ y &= 1x - 1 \end{aligned}$$

2 Slope-intercept form

If ℓ has slope m, and ℓ hits the y-axis at $(0, b)$, then

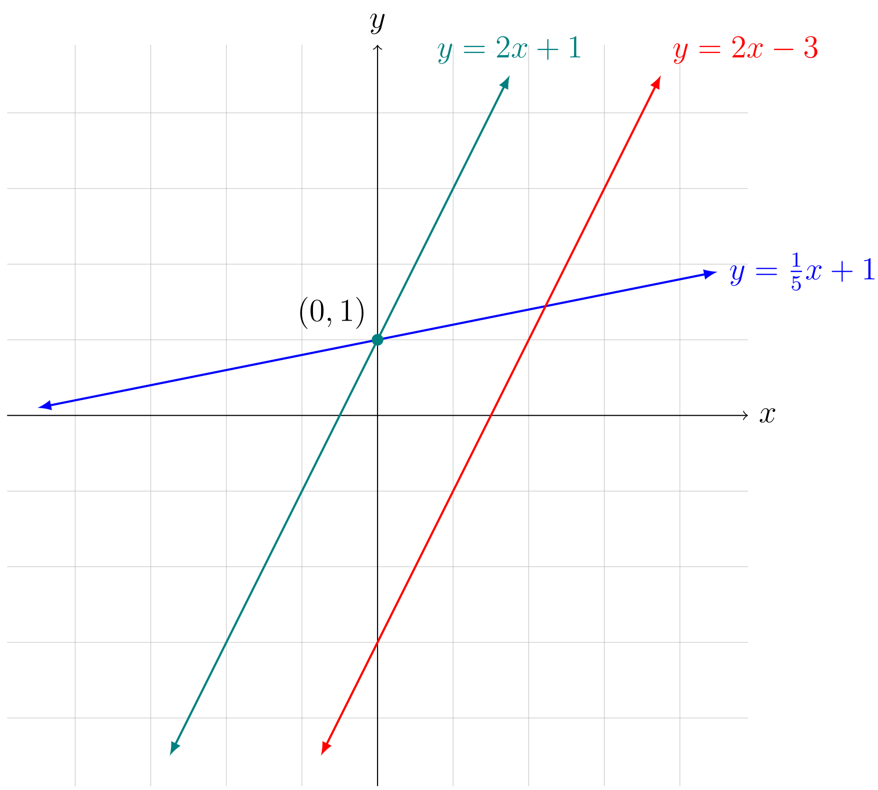
$$\boxed{y = mx + b}$$

is an equation for ℓ, where m is the slope and b is the y-intercept.

3 Drawing lines

Draw line with equation

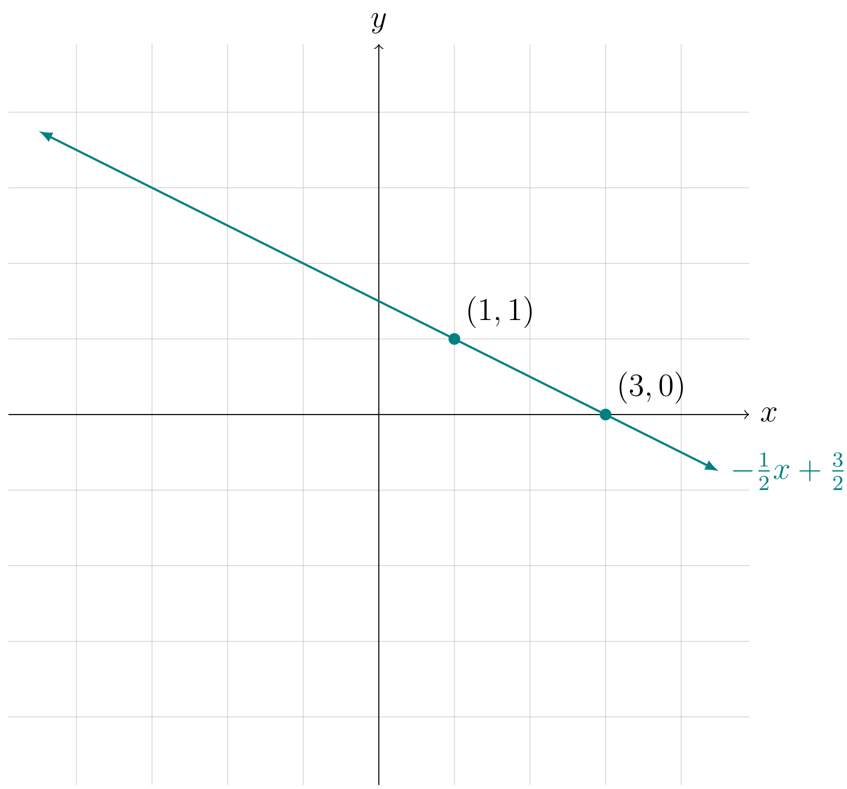
$$y = 2x + 1$$
$$y = \frac{1}{5}x + 1$$
$$y = 2x - 3$$



The slope tells you how to angle the line, and the y-intercept tells you where to anchor it on the y-axis.

4 Example

Problem: Line ℓ has points $(1,1)$ and $(3,0)$ on it. Find an equation for ℓ.



Find the slope:

$$m = \frac{0-1}{3-1} = -\frac{1}{2}$$

Some possible equations for the line in point-slope form:

$$y - 1 = -\frac{1}{2}(x-1)$$
$$y - 0 = -\frac{1}{2}(x-3)$$

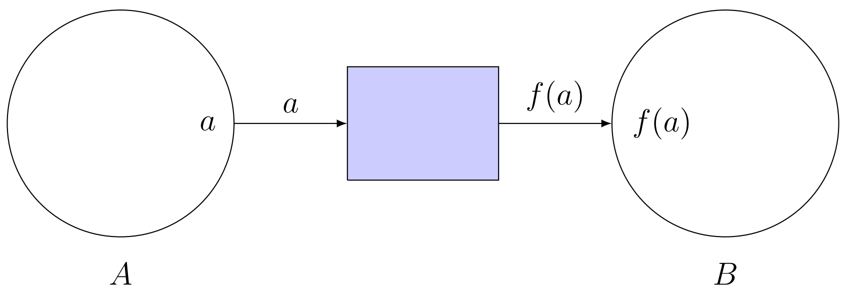
An equation for the line in slope-intercept form:

$$y = -\frac{1}{2}x + \frac{3}{2}$$

Functions: Mapping from Sets to Sets

Video companion

1 Function as a machine



A function $f : A \to B$ is a rule/formula/machine that transforms each element $a \in A$ into $f(a) \in B$.

$$\begin{aligned} a &: \text{input} \\ f(a) &: \text{output} \end{aligned}$$

2 Examples

Abstract example:

$$\begin{aligned} A &= \{1, 2, 10\} \qquad B = \{\text{apple}, \text{DE}, \text{monkey}\} \\ f &: A \to B \\ f(1) &= \text{apple} \\ f(2) &= \text{apple} \\ f(10) &= \text{monkey} \end{aligned}$$

Study participants test positive or negative:

$$\begin{aligned} X &= \{\text{all people in VBS study}\} \qquad Y = \{+, -\} \\ \text{Test} &: X \to Y \\ \text{Test(person)} &= + \\ \text{Test(person)} &= - \end{aligned}$$

Profit by year:

$$\begin{aligned} Y &= \{...2010, 2011, 2012, ...\} \qquad \mathbb{R} \\ \text{Profit} &: Y \to \mathbb{R} \\ \text{Profit(year)} &= \text{profit/loss in year} \\ \text{Profit}(2011) &= 1,007 \\ \text{Profit}(2012) &= -10,000 \end{aligned}$$

3 Supervised learning

Given: some examples of inputs $a \in A$ and outputs $f(a) \in B$
Mission: figure out $f : A \to B$

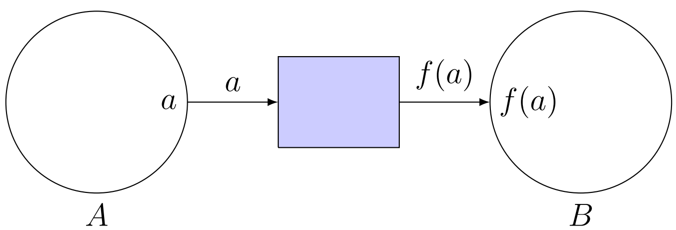
Functions: Graphing in the Cartesian Plane

Video companion

1 Introduction

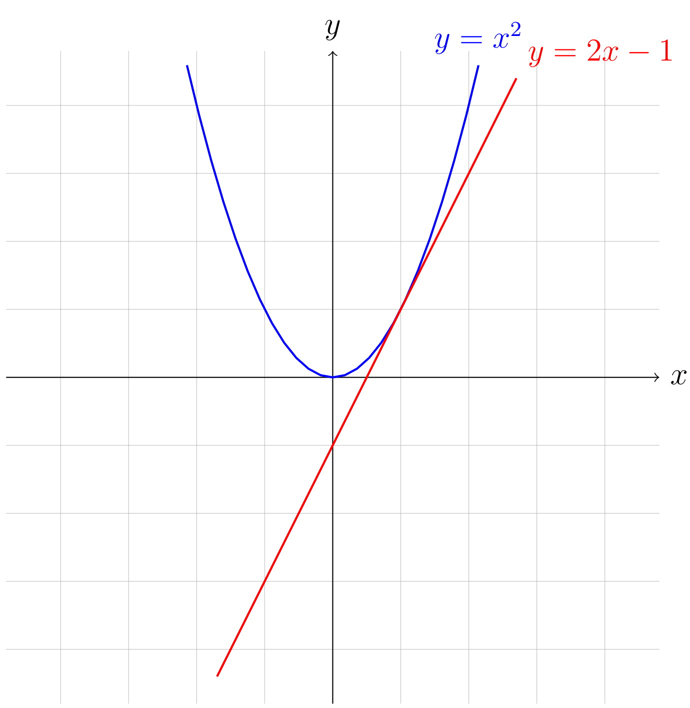
Last time: abstract depiction of a function as a machine

$f : A \to B$



This video: graphs of functions

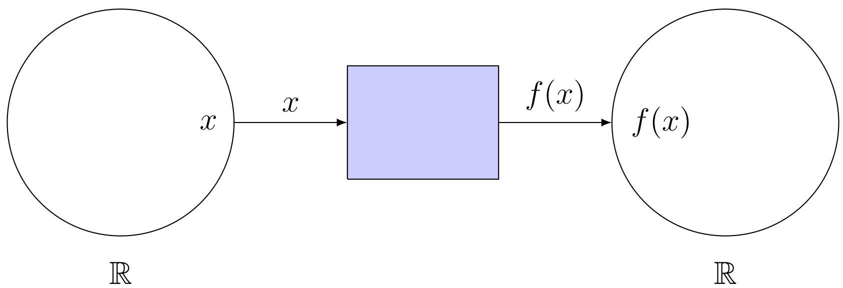
$f : \mathbb{R} \to \mathbb{R}$



You will learn the difference between a *graph* of a function and the function itself.

2 Map real line to real line

$f : \mathbb{R} \to \mathbb{R}$



A function is a formula, a rule for how to operate the machine.

$$\begin{aligned} f(x) &= 2x - 1 \\ f(1) &= 2(1) - 1 = 1 \\ f(0) &= 2(0) - 1 = -1 \\ f(5.1) &= 2(5.1) - 1 = 9.2 \end{aligned}$$

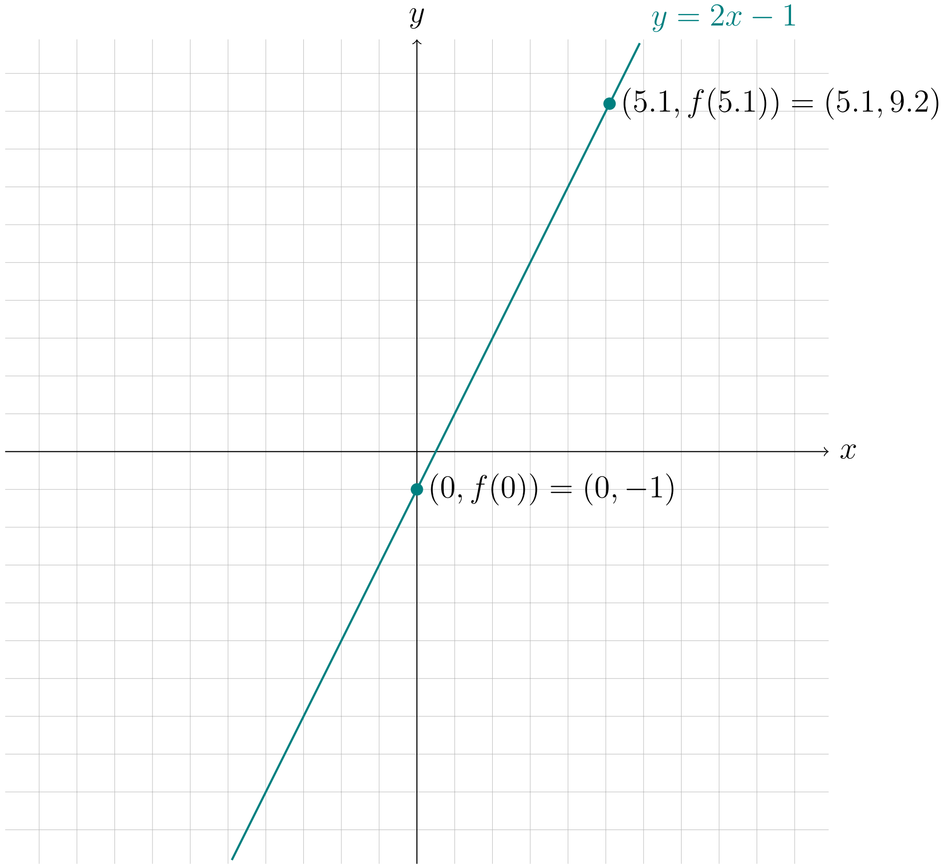
More complicated formulas, like absolute value:

$$\begin{aligned} g(x) &= |x| \\ &= \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases} \end{aligned}$$

Both f and g are functions, with a formula for how to compute the result.

3 What is a graph?

Graph of the function $f : \mathbb{R} \to \mathbb{R}$

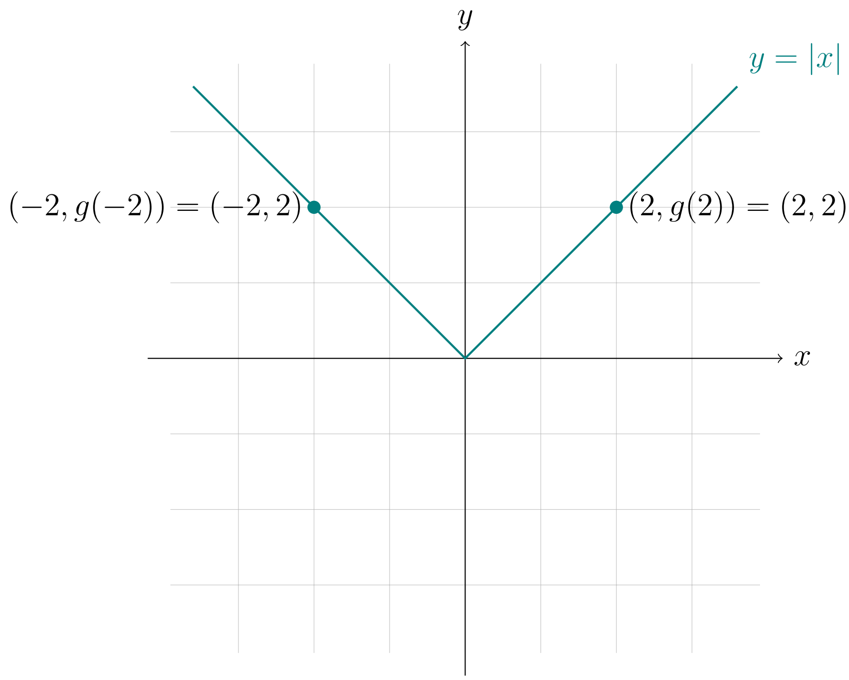


If g is a function $: \mathbb{R} \to \mathbb{R}$, the graph of $g = \{(x, y) \in \mathbb{R}^2 : y = g(x)\}$

4 Examples

Absolute value function

$$
\begin{aligned}
g(x) &= |x| \\
&= \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}
\end{aligned}
$$



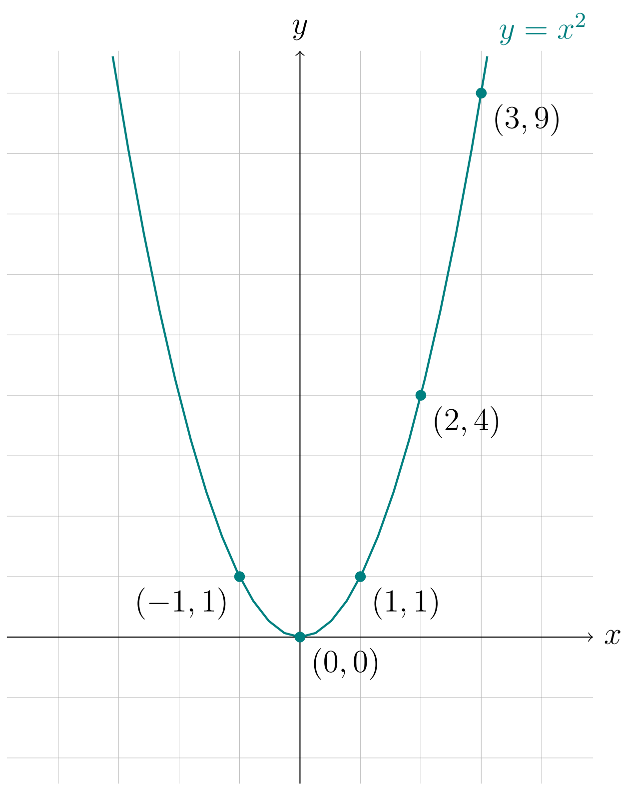
Quadratic function

$$h(x) = x^2$$

Graph a function by testing input and output pairs, see a pattern, and try to draw a curve through it. This is similar to *querying* in supervised learning.

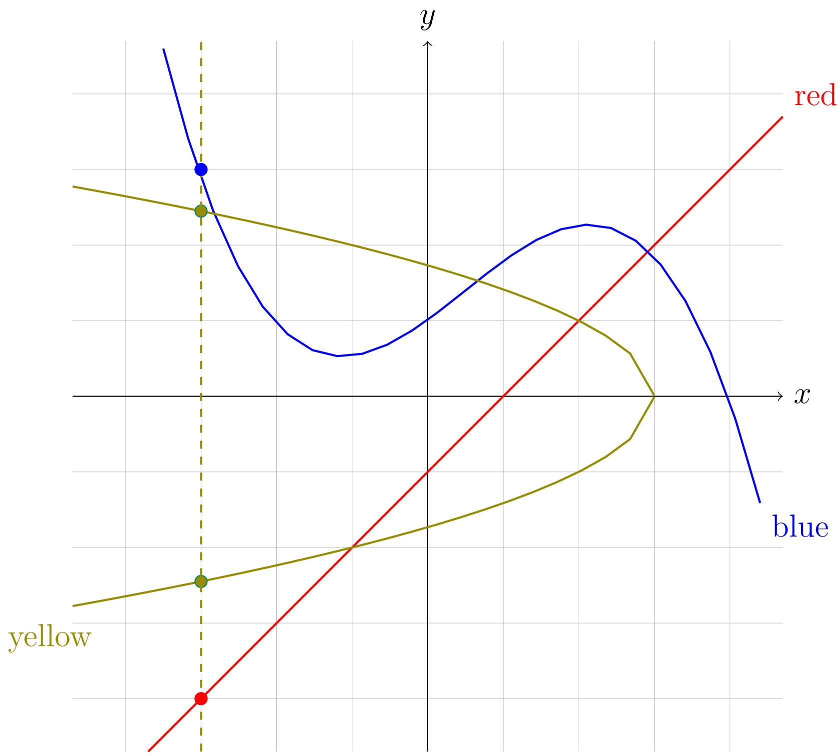
Table of values:

x	$h(x)$
0	$0^2 = 0$
1	$1^2 = 1$
2	$2^2 = 4$
3	$3^2 = 9$
-1	$(-1)^2 = 1$



$h(x) = x^2$ is a *quadratic* function.

5 Vertical line test

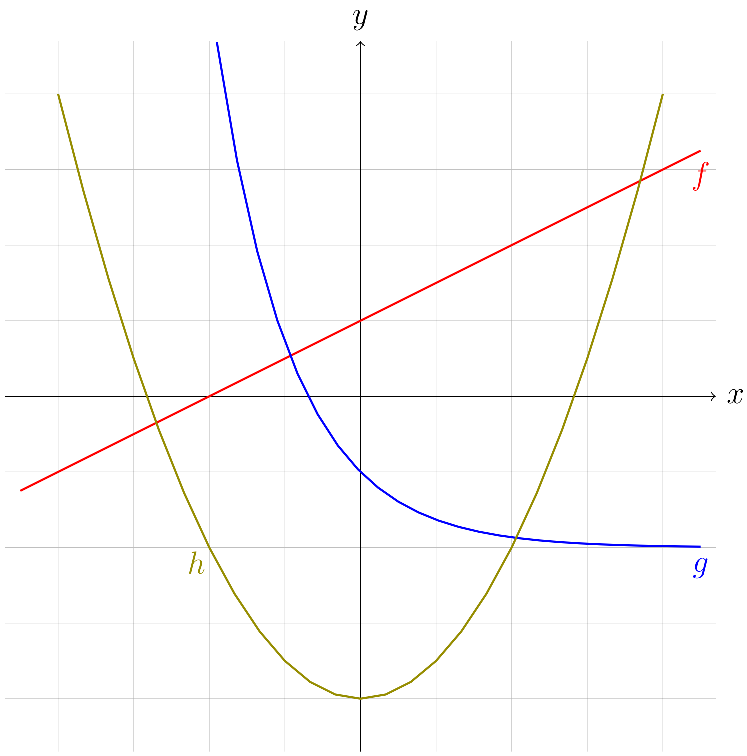


Red and blue could be graphs of functions. Yellow could not be the graph of a function because it violates the *vertical line test*, which states that *any vertical line intersects the graph of a function once.*

Functions: Increasing and Decreasing Functions

Video companion

1 Introduction



- f is *strictly increasing*
- g is *strictly decreasing*
- h is neither

Let $f : \mathbb{R} \to \mathbb{R}$,
f is strictly increasing if whenever $a < b$, we have $f(a) < f(b)$.
f is strictly decreasing if whenever $a < b$, we have $f(a) > f(b)$.

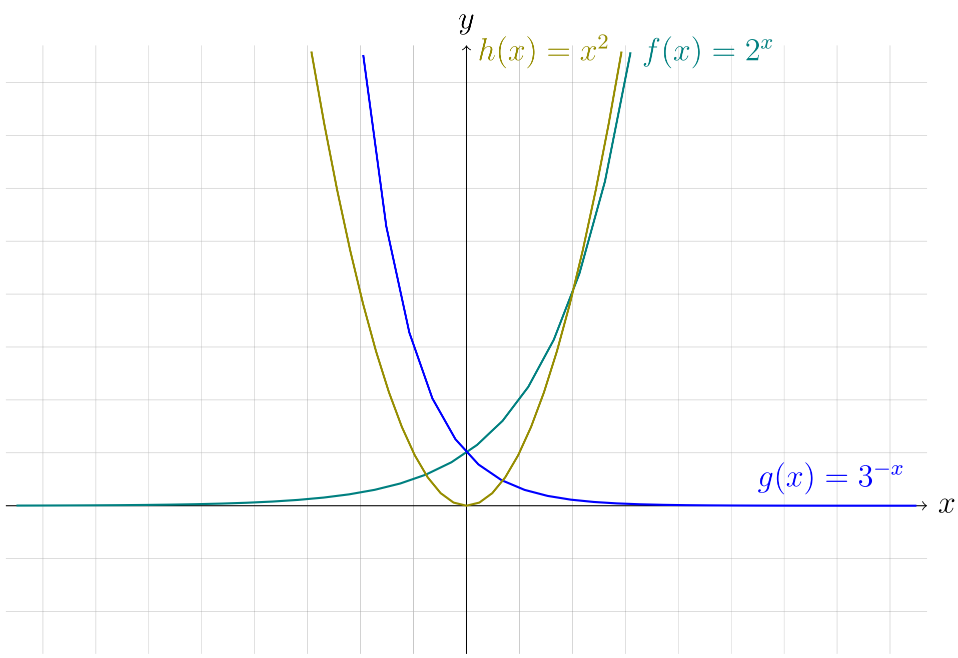
2 Examples

$f(x) = 2^x \quad$ (exponential function)
$g(x) = 3^{-x}$
$h(x) = x^2$

x	$f(x)$
0	$2^0 = 1$
1	$2^1 = 2$
2	$2^2 = 4$
3	$2^3 = 8$
-1	$2^{-1} = \frac{1}{2}$

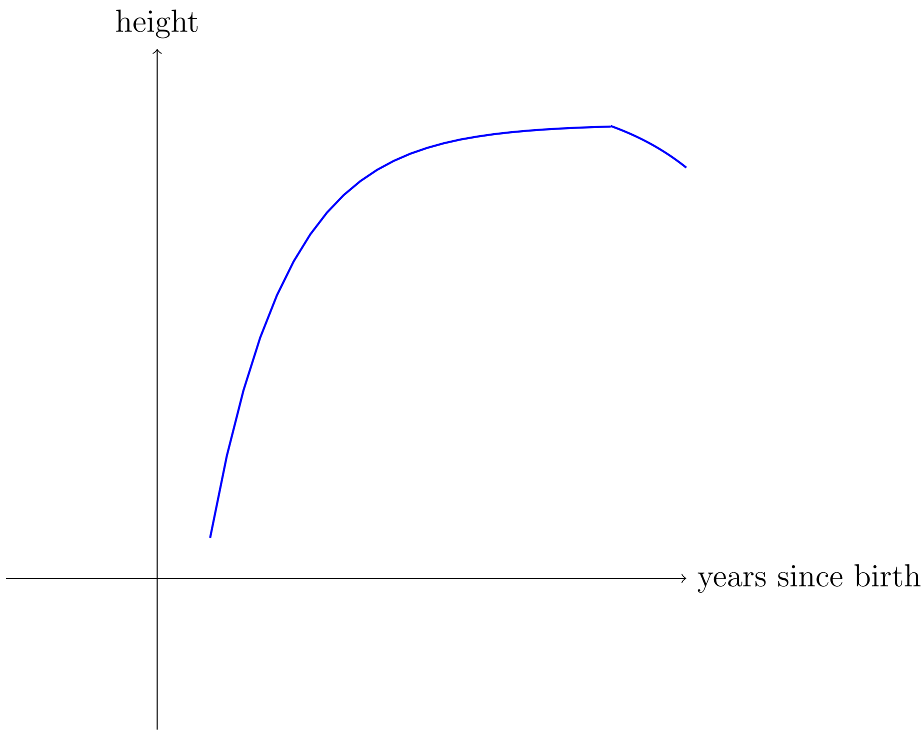
x	$g(x)$
0	$3^0 = 1$
1	$3^{-1} = \frac{1}{3}$
2	$3^{-2} = \frac{1}{9}$
3	$3^{-3} = \frac{1}{27}$
-1	$3^1 = 3$

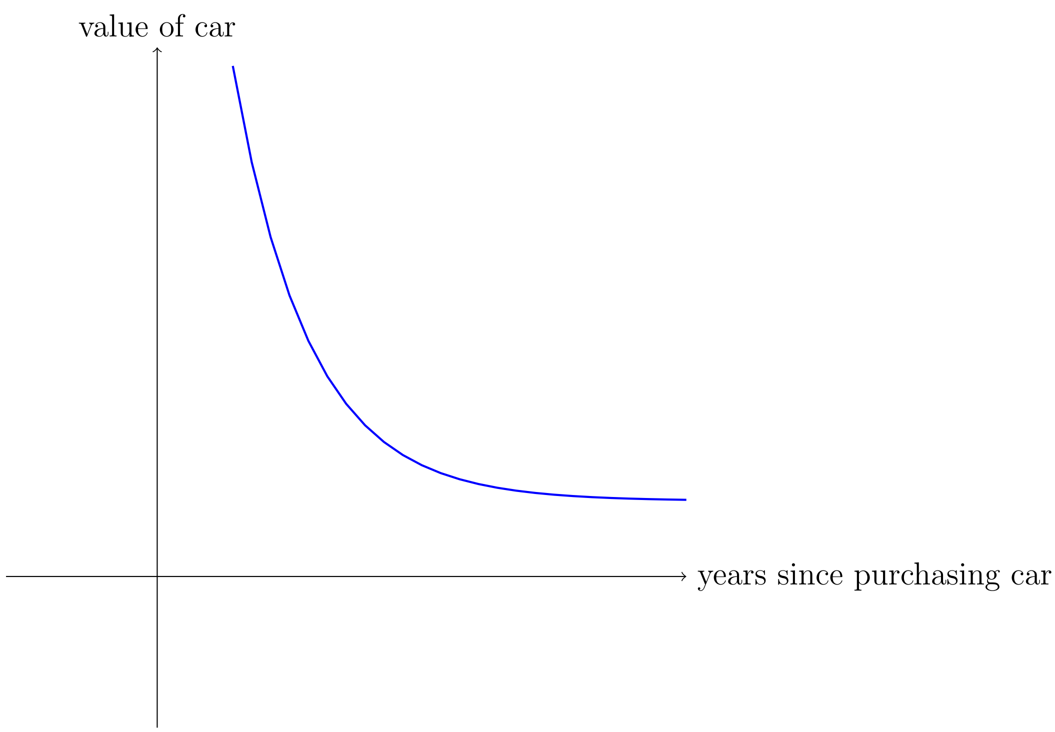
x	$h(x)$
0	$0^2 = 0$
1	$1^2 = 1$
2	$2^2 = 4$
3	$3^2 = 9$
-1	$(-1)^2 = 1$



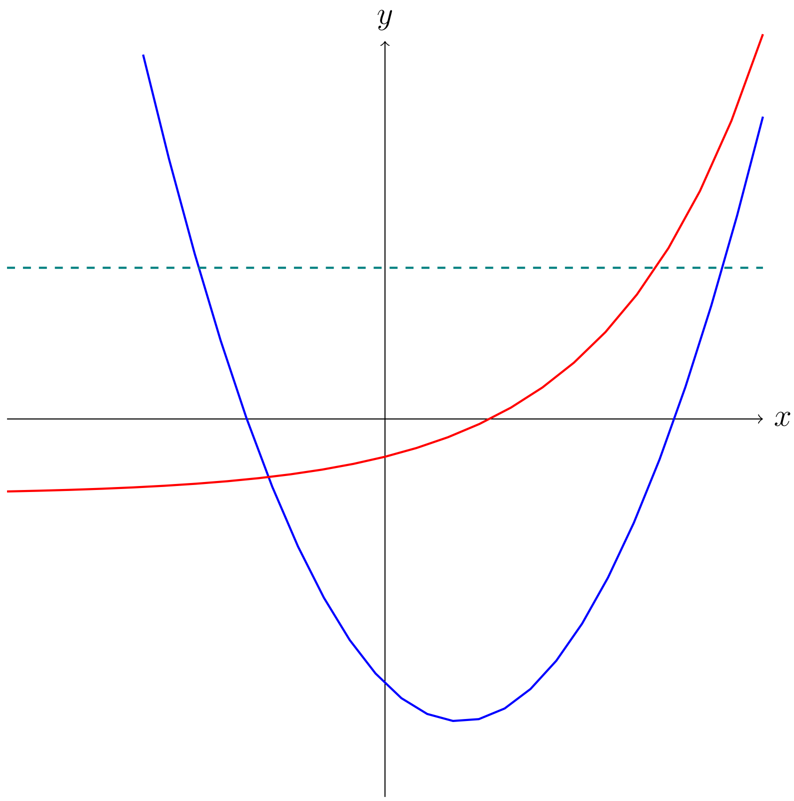
- f is strictly increasing
- g is strictly decreasing
- h is neither
  - h is strictly increasing on $[0, \infty)$
  - h is strictly decreasing on $(-\infty, 0]$

3 "Real-world" examples





4 Horizontal line test



A function is strictly increasing or strictly decreasing if a horizontal line crosses it only once.

Functions: Composition and Inverse

Video companion

1 Introduction

- Composing two functions
  - Basic identity
  - A warning
- Inverse functions
  - Basic identity
  - A neat picture
  - A warning

2 Composing functions

Definition: Given functions f and g, $(g \circ f)(x) = g(f(x))$, and $(f \circ g)(x) = f(g(x))$

Example:

$$f(x) = x^2$$
$$g(x) = x + 5$$

$$\begin{aligned}
(g \circ f)(x) &= g(f(x)) \\
&= g(x^2) = x^2 + 5 \\
g(f(2)) &= g(2^2) = 2^2 + 5 = 9 \\
(f \circ g)(x) &= f(g(x)) \\
&= f(x+5) = (x+5)^2 \neq x^2 + 5
\end{aligned}$$

3 Inverse functions

Example:

$$f(x) = 2x$$
$$g(x) = \frac{1}{2}x$$

$$(g \circ f)(x) = g(f(x)) = g(2x) = \frac{1}{2}(2x) = x$$

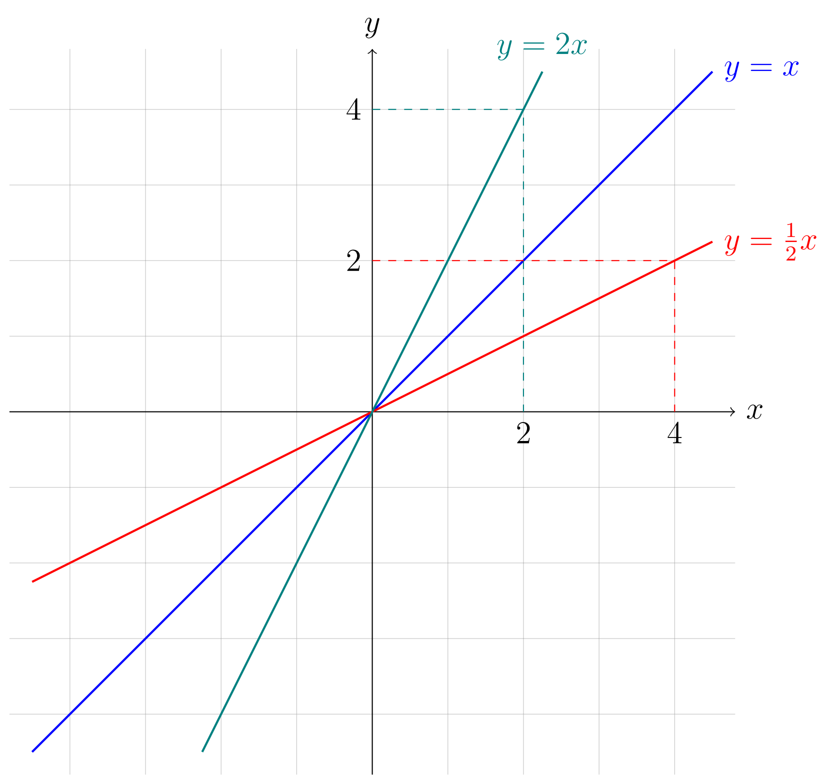
Notice: true for all x

$$(g \circ f)(3) = g(f(3)) = g(2 * 3) = \frac{1}{2}(2 * 3) = 3$$
$$(g \circ f)(\pi) = g(f(\pi)) = g(2 * \pi) = \frac{1}{2}(2 * \pi) = \pi$$

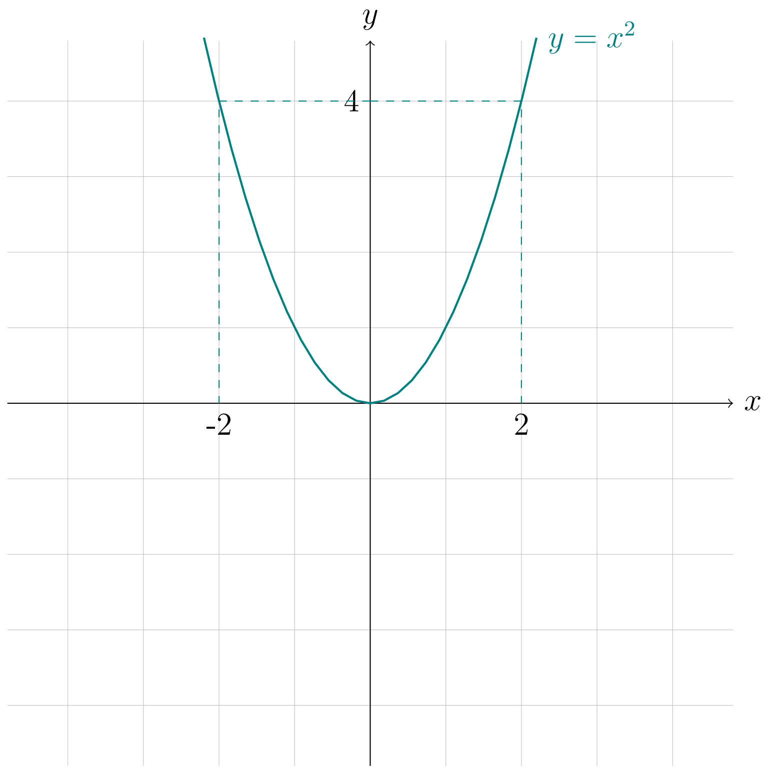
f and g are *inverses* of each other, i.e. f undoes what g does.

$$g = f^{-1}$$

4 Graphical depiction



Warning: not every function $f : \mathbb{R} \to \mathbb{R}$ has an inverse.

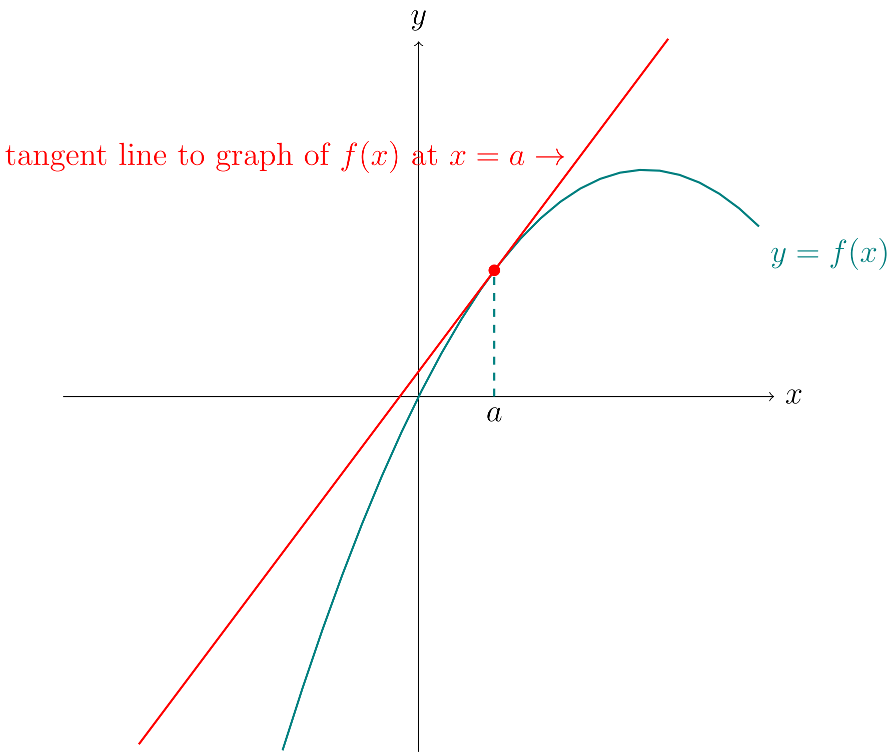


Warning: if the graph of f fails the horizontal line test, then f has no inverse. The only invertible functions are those that are either strictly increasing or strictly decreasing.

Tangent Lines: The Slope of a Graph at a Point

Video companion

1 Introduction



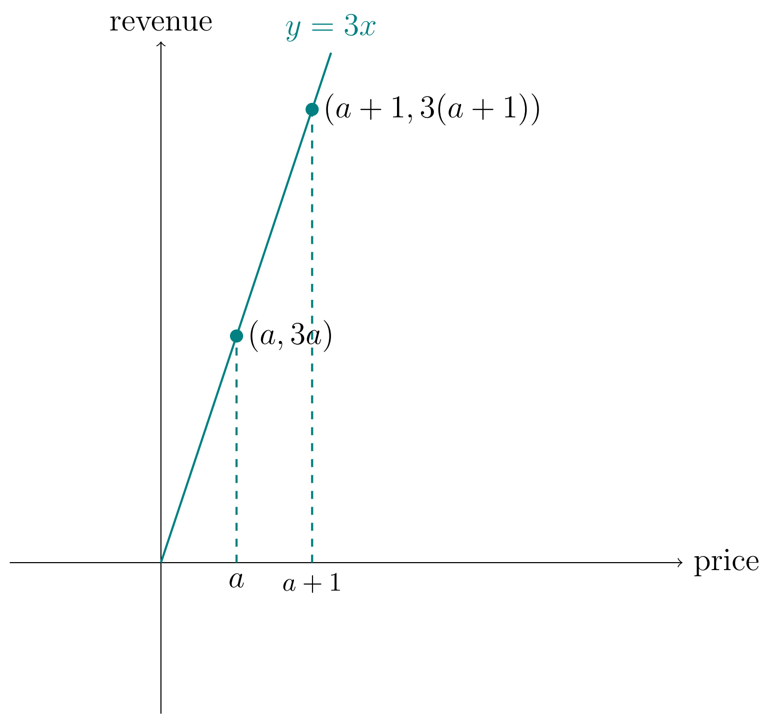
Question: How fast is $f(x)$ changing at $x = a$?

The slope of the *tangent line* gives the instantaneous rate of change. This is also called the *derivative* of the function at that point, or $f'(a)$.

Limit to find slope at $x = a$:

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

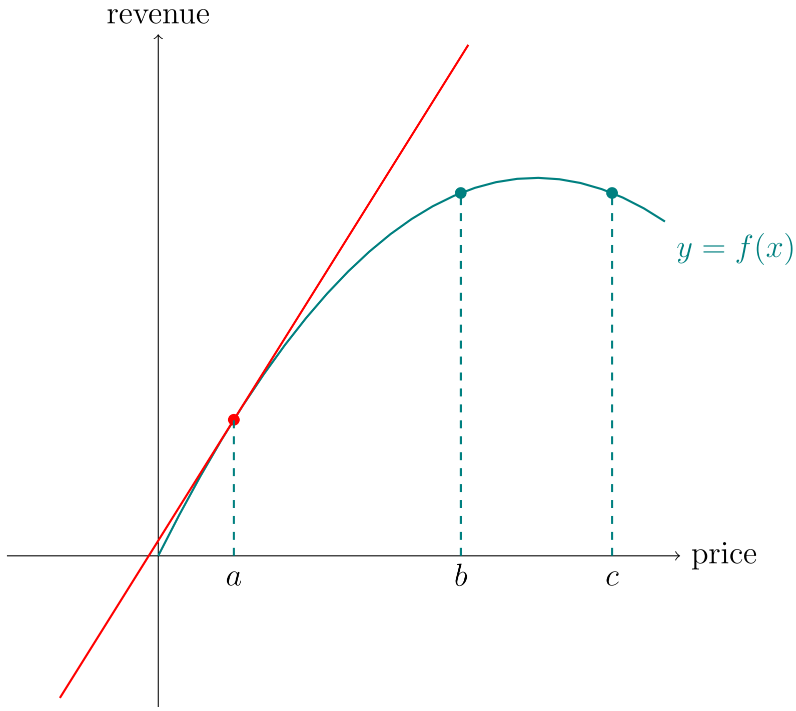
2 Simple example



Slope:

$$\frac{3(a+1) - 3a}{a + 1 - a} = 3$$

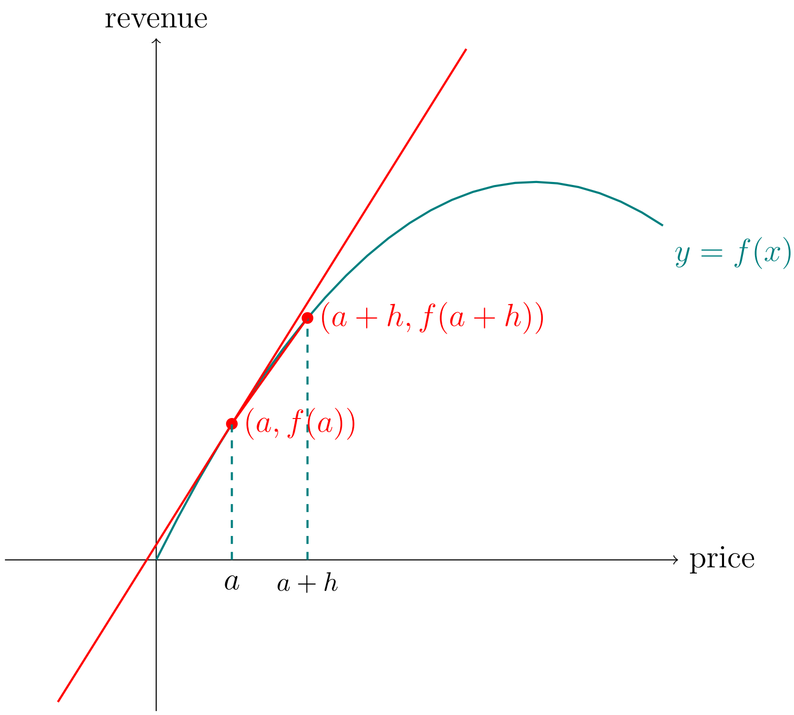
3 More realistic example



What is the instantaneous rate of change of revenue at a price point? It depends on the slope of the tangent line, which changes depending on the price point.

The answer is the slope, or derivative of the function at the price value $x = a$: $f'(a)$.

Related question: What is the slope of a line segment through a and another point on the line?



Slope of line from a to $a+h$:

$$\frac{f(a+h) - f(a)}{h}$$

Slope of tangent line at $x = a$:

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

This is calculus.

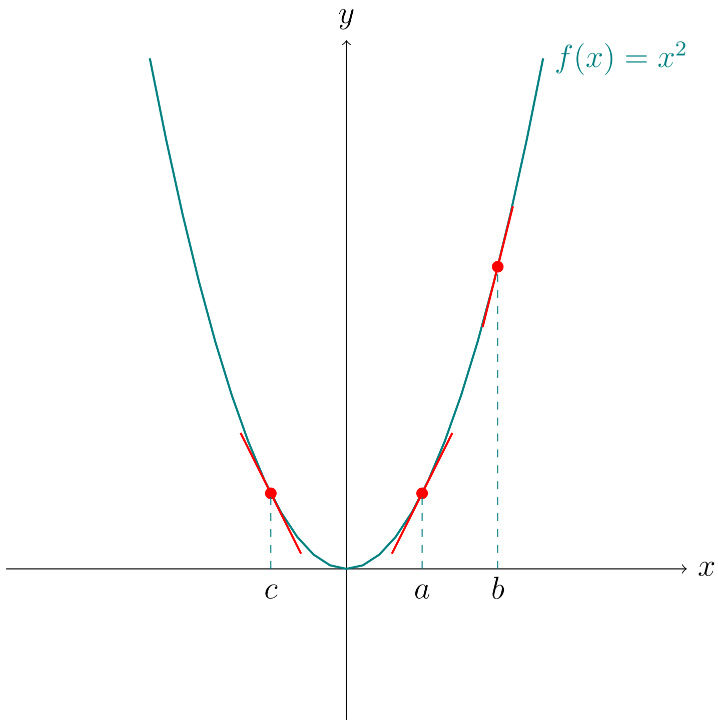
Tangent Lines: The Derivative Function

Video companion

1 Introduction

Derivative formula:

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



- Slope is positive at a: $f'(a) > 0$
- Slope is positive at b and greater than at a: $f'(b) > f'(a)$
- Slope is negative at c: $f'(c) < 0$

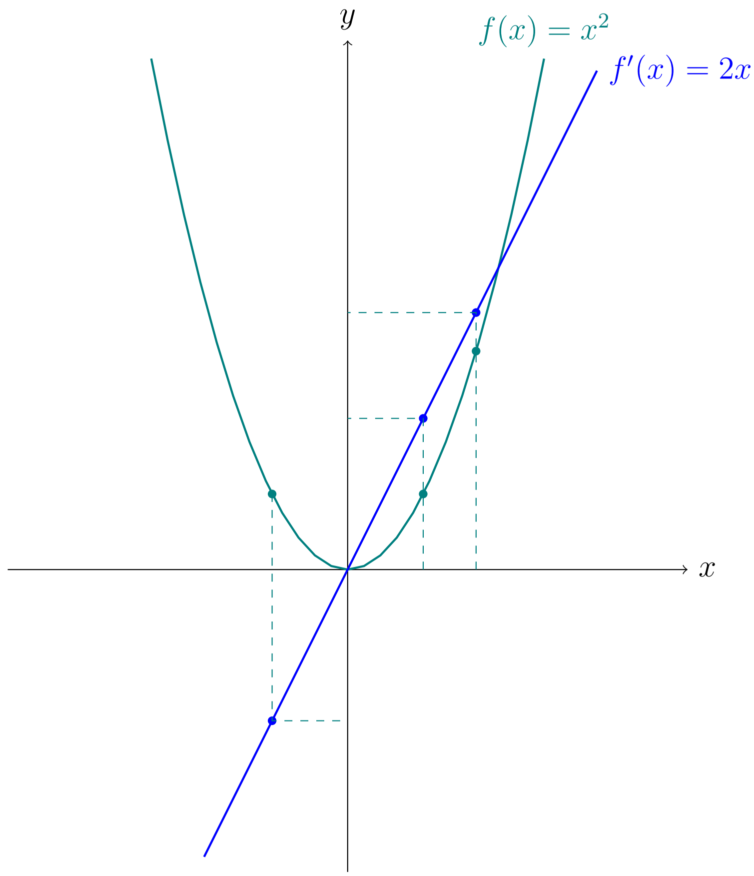
2 Calculate derivative

$$
\begin{aligned}
f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\
&= \lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} \\
&= \lim_{h \to 0} \frac{a^2 + 2ah + h^2 - a^2}{h} \\
&= \lim_{h \to 0} \frac{h(2a+h)}{h} \\
&= \lim_{h \to 0} (2a+h) \\
&= 2a
\end{aligned}
$$

$$
\begin{aligned}
f'(a) &= 2a \\
f'(b) &= 2b \\
f'(c) &= 2c
\end{aligned}
$$

Can verify $2a > 0$, $2b > 2a$, and $2c < 0$

3 Graph of derivative function



Next video: Finding where derivative is zero (where the tangent line to the function is horizontal) is important for optimization problems.

Fast Growth, Slow Growth: Using Integer Exponents

Video companion

1 Positive Integer Exponents

$$\begin{aligned} 9 &= 3 \cdot 3 &&= 3^2 \\ 27 &= 3 \cdot 3 \cdot 3 &&= 3^3 \\ 81 &= 3 \cdot 3 \cdot 3 \cdot 3 &&= 3^4 \end{aligned}$$

Exponents count how many times factors repeat in a number. 3^4 is pronounced "three to the fourth power" or "three to the fourth."

Example

$$248,832 = 12 \cdot 12 \cdot 12 \cdot 12 \cdot 12 = 12^5$$

A note on pronunciation $4 \cdot 4 = 4^2$ can be pronounced "four to the second"—but also "four squared." Similarly, $4 \cdot 4 \cdot 4 = 4^3$ can be pronounced "four to the third"—but also "four cubed."

2 Zero as an Exponent

$$1^0 = 1 \qquad (2\pi)^0 = 1$$

$$2^0 = 1 \qquad \left(\frac{1}{x^3}\right)^0 = 1$$

$$3^0 = 1$$

Definition By the definition of exponents, any number, except for zero, raised to the zeroth power is one. Note that 0^0 is undefined.

3 Negative Integer Exponents

$$2^{-1} = \frac{1}{2^1} = \frac{1}{2} \qquad \frac{1}{2^{-1}} = 2^1 = 2$$
$$2^{-2} = \frac{1}{2^2} = \frac{1}{4} \qquad \frac{1}{2^{-2}} = 2^2 = 4$$
$$2^{-3} = \frac{1}{2^3} = \frac{1}{8} \qquad \frac{1}{2^{-3}} = 2^3 = 8$$

General rule

$$x^{-n} = \frac{1}{x^n}$$
$$\frac{1}{x^{-n}} = x^n$$

4 Scientific Notation

Mass of the Earth (kg)

$$5,972,000,000,000,000,000,000,000$$
$$= 5.972 \times 10^{24}$$

Mass of an electron (kg)

$$0.0000000000000000000000000000009109$$
$$= 9.109 \times 10^{-31}$$

Keep the significant digits, and there is always one digit to the left of the decimal.

Fast Growth, Slow Growth: Simplification Rules for Algebra Using Exponents

Video companion

1 Exponent simplification rules

Five rules for simplifying algebraic expressions with exponents

1. Multiplication rule
   $x^n x^m = x^{(n+m)}$
2. Power to a power
   $(x^n)^m = x^{nm}$
3. Product to a power
   $(xy)^n = x^n y^n$
4. Fraction to a power
   $\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$
5. Division and negative powers
   $\frac{x^n}{x^m} = x^{(n-m)}$

2 Examples

Simple examples

$$\left(7^3\right)\left(7^7\right) = 7^{(3+7)} = 7^{10}$$

$$\left(4^3\right)^5 = 4^{(3\cdot5)} = 4^{15}$$

$$(8\cdot9)^7 = \left(8^7\right)\left(9^7\right) = 1.00306\times10^{13}$$

$$\left(\frac{2}{7}\right)^3 = \frac{2^3}{7^3} = 0.023323615$$

$$\frac{10^5}{10^3} = 10^{(5-3)} = 10^2 = 100$$

Complex examples

$$\frac{x^3y^4z^5}{x^3y^5z^2} = \frac{x^3}{x^3}\frac{y^4}{y^5}\frac{z^5}{z^2} = x^{(3-3)}y^{(4-5)}z^{(5-2)} = y^{-1}z^3 = \frac{z^3}{y}$$

$$\left[\frac{(xy)^2}{x^{-3}y^2}\right]^{-1} = \left[\frac{x^2y^2}{x^{-3}y^2}\right]^{-1} = \left[x^{(2--3)}y^{(2-2)}\right]^{-1} = \left[x^5\right]^{-1} = x^{-5} = \frac{1}{x^5}$$

3 Fractional exponents

In general

$$x^{\frac{a}{b}} = \sqrt[b]{x^a}$$

Examples

$$8^{\frac{2}{3}} = \left[\sqrt[3]{8}\right]^2$$
$$= \left[\sqrt[3]{2 \cdot 2 \cdot 2}\right]^2 = 2^2 = 4$$

or

$$= \sqrt[3]{8^2}$$
$$= \sqrt[3]{64} = \sqrt[3]{4 \cdot 4 \cdot 4} = 4$$

$$125^{\frac{4}{3}} = \left[\sqrt[3]{125}\right]^4$$
$$= \left[\sqrt[3]{5 \cdot 5 \cdot 5}\right]^4 = 5^4 = 625$$

Fast Growth, Slow Growth: How Logarithms and Exponents Are Related

Video companion

1 Introduction

Logarithm means "raised to what power?"

If the question is "what power of two is $2 \cdot 2 \cdot 2 = 8$?" then the answer is the logarithm to the base two of eight, which is $\log_2(8) = 3$.

Two general forms

$b^x = N$	"exponential form"
$x = \log_b(N)$	"logarithmic form"

Examples

If $b = 2$, $x = 3$, and $N = 8$:

$$2^3 = 8$$
$$3 = \log_2(8)$$

If $b = 2$, $x = 4$, and $N = 16$:

$$2^4 = 16$$
$$4 = \log_2(16)$$

2 Logs of one

Recall that raising any number to the power of zero is one, $b^0 = 1$. Therefore, the log, to any base, of one is zero.

$\log_2(1) = 0$	$2^0 = 1$
$\log_{10}(1) = 0$	$10^0 = 1$
$\log_{20}(1) = 0$	$20^0 = 1$

3 General rules

1. Product rule
   $\log(xy) = \log(x) + \log(y)$
2. Quotient rule
   $\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$
3. Power and root rule
   $\log(x^n) = n\log(x)$

Examples

$$\begin{aligned}
\log_b(35) &= \log_b(5) + \log_b(7) \\
&= \log_b(70) - \log_b(2) \\
\log_2\left(\frac{16}{4}\right) &= \log_2(16) - \log_2(4) = 4 - 2 = 2 \\
\log_2(1000)^{\frac{1}{3}} &= \frac{1}{3}\log_2(1000) \\
\log_{10}(7)^5 &= 5\log_{10} 7 \\
\log_b(x)^{-1} &= -\log_b(x)
\end{aligned}$$

$$\begin{aligned}
\log_b x^2 y^{-3} &= \log_b x^2 + \log_b y^{-3} \\
&= 2\log_b x - 3\log_b y \\
\log_b \frac{x^2}{y^{-\frac{1}{2}}} &= \log_b x^2 - \log_b y^{-\frac{1}{2}} \\
&= 2\log_b x + \frac{1}{2}\log_b y
\end{aligned}$$

4 Problem-solving technique

Problem-solving technique: Treat both sides of an equation as though they were exponents.

$x = y$ $\qquad\qquad$ $z^x = z^y$

Example

$$\begin{aligned}
\log_2\left(\frac{39x}{(x-5)}\right) &= 4 \\
2^{\log_2\left(\frac{39x}{(x-5)}\right)} &= 2^4 \\
\frac{39x}{(x-5)} &= 16 \\
39x &= 16x - 80 \\
23x &= -80 \\
x &= -\frac{80}{23}
\end{aligned}$$

Fast Growth, Slow Growth: The Change of Base Formula

Video companion

1 Introduction

Generally use base 10, base 2, and natural log (base e) in data science.

$$\log_2(12) = 3.585 \qquad \log_2(7) = 2.807$$
$$\log_{10}(12) = 1.079 \qquad \log_{10}(7) = 0.8451$$

$$2^{3.585} = 12 \qquad 2^{2.807} = 7$$
$$10^{1.079} = 12 \qquad 10^{0.8451} = 7$$

The change of base formula: "Old" base is x, "new" base is a,

$$\log_a(b) = \frac{\log_x(b)}{\log_x(a)}$$

Examples
Want to convert $\log_{10}(12)$ to base $a = 2$:

$$\log_2(12) = \frac{\log_{10}(12)}{\log_{10}(2)} = \frac{1.079}{0.30103} = 3.585$$

Want to convert $\log_2(7)$ to base $a = 10$:

$$\log_{10}(7) = \frac{\log_2(7)}{\log_2(10)} = \frac{2.8073}{3.3219} = 0.8540$$

Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Fast Growth, Slow Growth:
The Rate of Growth of Continuous Processes

Video companion

1 Introduction

"Exponential rate of growth" can be a *discrete* exponential rate of growth or a *continuous* exponential rate of growth

Discrete rate of growth

$\$1.00(1+r)^t$

How much money would grow in discrete intervals of time t

If $r = 100\%$/year and $t = 1$, then would have \$2.00 after one year,
After 2 years, would have \$4.00
After 3 years, would have \$8.00...

2 Continuous exponential growth

Euler's constant e

100% interest per year (discrete)
50% interest for 6 months, then interest on interest for another 6 months.

Interval	Factor	Repeats		Result
1 year	$1+1$	1	$(2)^1 =$	2
6 months	1.5	2	$(1.5)^2 =$	2.25
3 months	1.25	4	$(1.25)^4 =$	2.44

As time intervals decrease, does result increase in an unlimited way?

No...

Interval	Factor	Repeats		Result
1 month	1.08	12	$(1.08)^{12} =$	2.613
1 week	1.019	52	$(1.019)^{52} =$	2.693
1 day	1.002739	365	$(1.002739)^{365} =$	2.7146
1 hour	1.000114	8760	$(1.000114)^{8760} =$	2.71813
1 minute	1.0000019	525,600	$(1.0000019)^{525,600} =$	2.71828
1 second	1.0000000317	31,536,000	$(1.0000000317)^{31,536,000} =$	2.71828

$e = 2.71828$, Euler's constant

Problem A baby elephant weighing 200 kg grows at a continuously compounded rate of 5%/year. How much does it weigh in 3 years?

$$(200\text{ kg})e^{(0.05)(3)} = 232.4\text{ kg}$$

3 Continuous rate of return

"Log to the base e of x" is given by the symbol $\ln(x)$, where ln stands for *natural logarithm.*

Problem Rabbit population increases in mass at a rate of 200% per year. Population starts at 10 kg. If they increase at a continuously compounded rate, how many years is it until they weigh as much as the Earth (5.972×10^{24} kg)?

$$\begin{aligned} 5.972 \times 10^{24}\text{ kg} &= (10\text{ kg})e^{2t} \\ 5.972 \times 10^{23} &= e^{2t} \\ \ln(5.972 \times 10^{23}) &= \ln(e^{2t}) = 2t \\ \frac{\ln(5.972 \times 10^{23})}{2} &= t = 27.37\text{ years} \end{aligned}$$

Basic Probability Definitions: Probability Definitions and Notation

Video companion

1 Introduction

Definition
probability—the degree of belief in the truth or falsity of a statement

Range of uncertainty from 0 to 1
Certain statement is true: probability 1
Certain statement is false: probability 0

Example Statement x: "It is raining."

2 Notation

$P(x)$	probability of x
$\sim x$	negation of statement x

Law of excluded middle

$$P(x) + P(\sim x) = 1$$

Probability of a statement and the probability of the negation of a statement must sum to 1.

If $P(x) = 1$, then $P(\sim x) = 0$, and vice versa.

In general, all outcomes of a probability distribution must sum to 1.

Definitions
probability distribution—collection of statements that are *exclusive* and *exhaustive*
exclusive—given complete information, no more than one of the statements can be true

exhaustive—given complete information, at least one of the statements must be true

A distribution X consisting of n statements would be denoted

$$X = \{x_1, x_2, x_3, ..., x_n\}.$$

The probability of each statement must sum to 1, which is denoted.

$$P(x_1) + P(x_2) + P(x_3) + ... + P(x_n) = 1.$$

3 Principle of indifference

For the i-th outcome x_i in a distribution with n possible outcomes,

$$P(x_i) = \frac{1}{n}.$$

Example: Drawing an ace of spades from a well-shuffled deck of 52 cards. The probability of drawing the ace of spades is $\frac{1}{52}$.

General statement
When there is no basis to choose some outcomes as more likely than others,

$$P(\text{event}) = \frac{\text{number of outcomes as defined in event}}{\text{total number of possible outcomes in universe}}.$$

Example: Event is drawing a queen, which has four outcomes in the event. The total number of outcomes is 52, so the probability of drawing a queen is $\frac{4}{52} = \frac{1}{13}$.

Example: Event is rolling an even number on a six-sided die, which has three outcomes in the event. The total number of outcomes is 6, so the probability of rolling an even is $\frac{3}{6} = \frac{1}{2}$.

Basic Probability Definitions: Joint Probabilities

Video companion

1 Introduction

Definition
joint probability—probability that two separate events with separate probability distributions are both true

$P(A \text{ and } B)$ is written $P(A, B)$, and read "the joint probability of A and B" or "the probability that A is true and B is true."

2 Order of joint probabilities

For probability distributions X and Y:

$$X = \{x_1, x_2, x_3, ..., x_n\}$$
$$Y = \{y_1, y_2, y_3, ..., y_n\}$$

Ordering does not matter in joint probabilities, for either the probability distributions or the individual events.

$$P(X, Y) = P(Y, X)$$
$$P(x_1, y_1) = P(y_1, x_1)$$

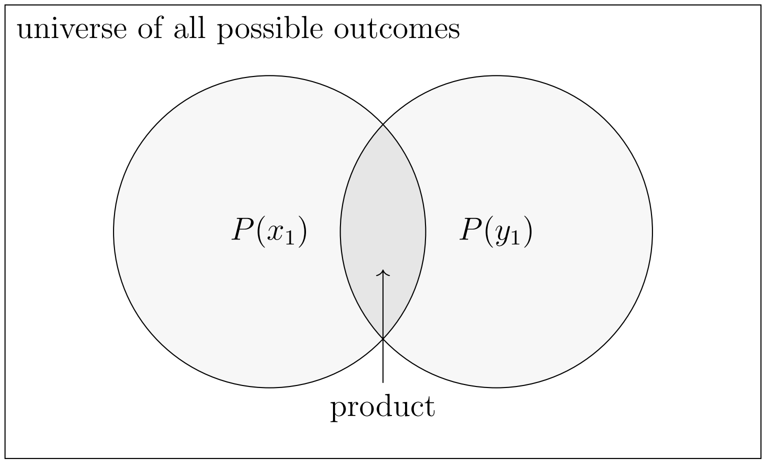
3 Independence

Definition
independence—knowing the outcome of one event does not change the probability of the other

The probability of two independent events:

$$\underbrace{P(x_1, y_1)}_{\text{"joint distribution"}} = \underbrace{P(x_1)P(y_1)}_{\text{"product distribution"}}$$

Venn diagrams—show the intersection and union of sets

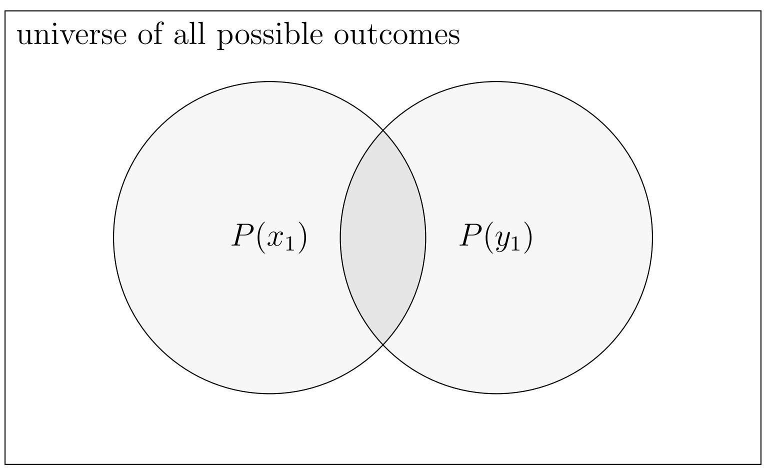


4 OR probability

Probability that either of two events occurs:

$$P(x_1 \text{ or } y_1) = P(x_1) + P(y_1) - P(x_1, y_1)$$

Venn diagram:



Problem Solving Methods:
Permutations and Combinations

Video companion

1 Introduction

Topic: Probability of events occurring in an order or the probability of a group of events occurring

Definitions
permutation—order matters, e.g. placing five people in five different positions: 120 ways
combination—order does not matter, e.g. forming a five-person team from five people: 1 way

2 Replacement

Sampling *with replacement* (independent), e.g. drawing a card and putting it back in the deck

Sampling *without replacement*, e.g. drawing a card from a deck and not putting it back

With the options permutation, combination, with replacement, and without replacement, we have most of the probability situations that are likely to arise in a basic probability course.

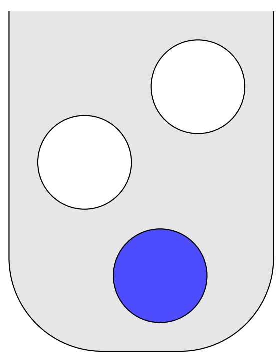
Problem Solving Methods: Using Factorial and "M Choose N"

Video companion

1 Introduction

Urn—a container you cannot see into

Example Drawing a marble from an urn containing two white and one blue marble. Can draw with or without replacement. Drawing with replacement means events are independent.



With replacement:

Draw	Probability
1 white	2/3
1 blue	1/3
2 white (in a row)	(2/3)(2/3) = 4/9

Without replacement:

Draw	Probability
1 white	2/3
1 blue	1/3
2 white (in a row)	(2/3)(1/2) = 1/3

2 Factorial

A factorial is the operation where we take a number and multiply it by each integer that is 1 less until we get down to 1.

Notation: 5! is read "five factorial." The operation means:

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120.$$

Factorial quotients:

$$\frac{7!}{5!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 7 \cdot 6 = 42$$

Convention:

$$0! = 1$$

3 "m choose n"

Draw n items from a group of m items without replacement.

Example: How many unique committees of five people from a group of ten people?
In this example, "10 choose 5," $m = 10$ and $n = 5$. The notation is given by:

$$\begin{aligned}\binom{10}{5} &= \frac{10!}{5! \cdot 5!} \\ &= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2 \cdot 3 \cdot 2 \cdot 7 \cdot 3 = 252\end{aligned}$$

General formula

$$\binom{m}{n} = \frac{m!}{(m-n)! \cdot n!}$$

Paul Bendich and Daniel Egger
Duke University

Problem Solving Methods: The Sum Rule, Conditional Probability, and the Product Rule

Video companion

1 Marginal probabilities and the sum rule

Often know the joint probabilities, but don't know individual probabilities.

Table of known joint probabilities:

(X,Y)		X		
		x_1	x_2	x_3
	y_1	$P(x_1,y_1)$ 0.01	$P(x_2,y_1)$ 0.02	$P(x_3,y_1)$ 0.03
Y	y_2	$P(x_1,y_2)$ 0.10	$P(x_2,y_2)$ 0.20	$P(x_3,y_2)$ 0.49
	y_3	$P(x_1,y_3)$ 0.04	$P(x_2,y_3)$ 0.05	$P(x_3,y_3)$ 0.06

Can refer to $P(x_1)$ as the "marginal probability of x_1" because it is in the margins of the matrix.

Sum rule: The marginal probability is equal to the sum of the joint probabilities.

For x_1, this means:

$$\begin{aligned} P(x_1) &= P(x_1,y_1) + P(x_1,y_2) + P(x_1,y_3) \\ &= 0.01 + 0.10 + 0.04 = 0.15 \end{aligned}$$

and for y_2:

$$\begin{aligned} P(y_2) &= P(x_1,y_2) + P(x_2,y_2) + P(x_3,y_2) \\ &= 0.10 + 0.20 + 0.49 = 0.79 \end{aligned}$$

Sum rule for binary probability distribution:

$$P(A) = P(A, B) + P(A, \sim B)$$

Sum rule for series of n probabilities:

$$P(A) = P(A, B_1) + P(A, B_2) + ... + P(A, B_n)$$

2 Conditional probability

Definition:
conditional probability—the probability that a statement is true given that some other statement is true with certainty.

Symbol $P(A \mid B)$ means the "probability of A given that B is true with certainty."

Example: What is the probability of rolling a 3 on a 6-sided die, given that the roll is odd? Of the outcomes when the roll is odd (3), one is the relevant outcome, so the probability is 1/3.

Example: What is the probability of rolling an odd, given that the roll is a 3? Of the outcomes when the roll is a 3 (1), one is the relevant outcome, so the probability is 1.

Formula for conditional probability:

$$P(A \mid B) = \frac{\text{(relevant outcomes)}}{\text{(total outcomes remaining in universe, when } B \text{ is true)}}$$

3 Product rule

Want to relate concepts of joint probability, marginal probability, and conditional probability.

Product rule:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

Conditional probability of A given that B is true is equal to the joint probability that A and B are true, divided by the probability that B is true.

Old definition of independence:

$$P(A, B) = P(A)P(B)$$

Dividing by $P(B)$ gives

$$P(A) = \frac{P(A, B)}{P(B)}$$

Using the product rule gives another definition of independence.

New definition of independence:

$$P(A \mid B) = P(A)$$

Intuitively, this means knowing that B is true tells us nothing about the probabilities of A. The outcome B has no effect on A; therefore, they are independent.

Conversely, if $P(A \mid B) \neq P(A)$, then the events are *dependent*.

Bayes' Theorem, Part 1

Video companion

1 Derivation

Starting from the product rule,

$$P(A \mid B) = \frac{P(A,B)}{P(B)},$$

and multiplying by the probability of B $P(B)$, gives

$$P(A \mid B)P(B) = P(A,B).$$

Substituting the equivalent $P(B,A)$ for $P(A,B)$,

$$P(A \mid B)P(B) = P(B,A),$$

and using the product rule $P(B \mid A) = P(B,A)/P(A)$, gives

$$P(A \mid B)P(B) = P(B \mid A)P(A),$$

which when rearranged is Bayes' theorem.

Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

2 Inverse probability

An inverse probability problem is one where the answer is in the form of the probability that a certain process with a certain probability parameter is being used to generate the observed data.

The symbol B is used to represent the observed data. The symbol A_i is used to represent a possible process with probability parameter θ_i.

Example: Urn 1 has 20% white marbles, and urn 2 has 10% white marbles. We observe three white marbles in a row drawn with replacement. What is the probability that we are observing urn 1? Urn 2?

What we know:

Process		P(white marble)
Urn 1	A_1	20%
Urn 2	A_2	10%

where "white marble" is the parameter.

In forward probability we are interested in the probability of an event given a known process.

In this problem, we know the outcome and want to know how probable it is that each process was involved.

Written in terms of conditional probability,

$$P(\text{process parameter} \mid \text{observed data}) = \frac{P(\text{observed data} \mid \text{process parameter}_i)P(\text{process parameter}_i)}{P(\text{data} \mid \text{process}_1)P(\text{process}_1) + P(\text{data} \mid \text{process}_2)P(\text{process}_2) + ... + P(\text{data} \mid \text{process}_n)P(\text{process}_n)}$$

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + ... + P(B \mid A_n)P(A_n)}$$

Example arithmetic:

First, solve for likelihoods:

Urn 1: P(3 white marbles in a row | 20% white) = $(0.2)(0.2)(0.2) = 8/1000$
Urn 2: P(3 white marbles in a row | 10% white) = $(0.1)(0.1)(0.1) = 1/1000$

Using principle of indifference,

$$P(A_1) = 0.5$$
$$P(A_2) = 0.5$$

because we are neutral before observing any data. These are called the "prior probability."

$$\begin{aligned} P(A_1 \mid B) &= \frac{P(B \mid A_1)P(A_1)}{P(B, A_1) + P(B, A_2)} \\ &= \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} \\ &= \frac{(8/1000)(1/2)}{(8/1000)(1/2) + (1/1000)(1/2)} = 8/9 \end{aligned}$$

Probability that we observed urn 1: 8/9
Probability that we observed urn 2: 1/9

Bayes' Theorem: Updating with New Data

Video companion

1 Updating probabilities

Bayes' theorem allows for updating probabilities based on new data.

From previous solution:

Process		$P(\text{white marble})$
Urn 1	A_1	20%
Urn 2	A_2	10%

$P(\text{Urn 1} \mid \text{3 white marbles in a row}) = 8/9$
$P(\text{Urn 2} \mid \text{3 white marbles in a row}) = 1/9$

New information: We draw a fourth marble that is also white.

$P(A_1)$ and $P(A_2)$ become our *new prior probabilities* or *new priors*.

$$\begin{aligned}&P(\text{urn 1} \mid \text{3 white marbles in a row, } \textit{and} \text{ a 4th})\\ &= \frac{P(\text{white} \mid \text{urn 1})P(\text{urn 1})}{P(\text{white} \mid \text{urn 1})P(\text{urn 1}) + P(\text{white} \mid \text{urn 2})P(\text{urn 2})}\\ &= \frac{(0.2)(8/9)}{(0.2)(8/9) + (0.1)(1/9)}\end{aligned}$$

$P(\text{urn 1}) = 94.12\%$ $\uparrow$
$P(\text{urn 2}) = 5.88\%$ $\downarrow$

2 Technical vocabulary

Technical vocabulary of Bayesian inverse probability:

$$\underbrace{P(\theta_i \mid D)}_{\text{posterior probability}} = \frac{\overbrace{P(D \mid \theta_i)}^{\text{likelihood}} \quad \overbrace{P(\theta_i)}^{\text{prior probability}}}{\underbrace{P(D)}_{\text{marginal probability}}}$$

posterior probability—probability after new data is observed
prior probability—probability before any data is observed or before new data is observed
likelihood—standard forward probability of data given parameters
marginal probability—probability of the data

The Binomial Theorem and Bayes' Theorem

Video companion

1 Introduction

Binomial theorem used when there are two possible outcomes—a success or a non-success, for example, flipping a coin—heads are a success, binary outcome.

Not limited to fair coins, where the probability of success is 0.5. Probability can be any value > 0 and < 1.

2 Binomial theorem

Probability of s successes in n trials, when probability of 1 success is p:

$$\binom{n}{s} p^s (1-p)^{n-s}$$

where n is the number of independent trials (with replacement), s is the number of successes, and p is the probability of one success

Example: 72 heads out of 100 coin tosses of a fair coin

$$n = 100$$
$$s = 72$$
$$p = 0.5$$

$$\binom{100}{72} (0.5)^{72} (1-0.5)^{100-72}$$
$$= \binom{100}{72} (0.5)^{72} (0.5)^{28} = 3.94 \times 10^{-6}$$

3 With Bayes' theorem

Question: Is it more likely a fair coin ($p = 0.5$) heads or a bent coin ($p = 0.55$) heads?

$$
\begin{aligned}
&P(\text{fair coin} \mid 72 \text{ heads}/100) \\
&= \frac{P(72 \text{ heads}/100 \mid \text{fair coin})P(\text{fair coin})}{P(72 \text{ heads}/100 \mid \text{fair coin})P(\text{fair coin}) + P(72 \text{ heads}/100 \mid \text{bent coin})P(\text{bent coin})} \\
&= \frac{(3.94 \times 10^{-6})(1/2)}{(3.94 \times 10^{-6})(1/2) + (1.972 \times 10^{-4})(1/2)} \\
&= 1.96\%
\end{aligned}
$$

(assuming it is equally likely that the coin is fair or bent)

Therefore, there is $< 2\%$ probability that the coin is fair and $> 98\%$ probability that the coin is bent.

Bayes' theorem together with binomial theorem can tell us the probability of a process given data that we have observed.