MA180-4 Mathematics: Algebra

Professor Dane Flannery

dane.flannery@nuigalway.ie

This course covers ideas and methods from abstract algebra and discrete mathematics that are indispensible in all sciences:

- logic and sets
- permutations and polynomials
- mathematical induction and probability.

Blackboard will be used for posting Algebra learning materials, announcements, etc.

References

- Norman L. Biggs, *Discrete Mathematics*, Oxford University Press.
- Mark V. Lawson, *Algebra & Geometry: An Introduction to University Mathematics*, Taylor & Francis.
- Online! many sources available for the individual topics.

Assessment

- Continuous: six online assignments.
- End-of-semester examination, covering all topics studied during the semester.

Logic

Logic is the study of methods to reason validly: obtaining justified conclusions from assumed premises.

Example. A certain island has two types of inhabitants: knights and knaves (every inhabitant is either a knight or a knave).
Knights *always* tell the truth.
Knaves *always* lie.
You talk to two inhabitants, called A and B.
A says: "Exactly one of us is a knave".
B says: "At least one of us is a knight".
Who is telling the truth?

Solution. To solve this puzzle, we use a *truth table*.

In two columns on the left, list all possible *truth values* (T: true, F: false) of the claims 'X is a knight' (X = A or B).
Then in another two columns, write the corresponding truth values for each row of the statements $\mathcal{S}_\mathsf{A}$:= "Exactly one is a knave"; $\mathcal{S}_\mathsf{B}$:= "At least one is a knight".

A	B	$\mathcal{S}_\mathsf{A}$	$\mathcal{S}_\mathsf{B}$
T	T	F	T
T	F	T	T
F	T	T	T
F	T	F	F

As knights always tell the truth and knaves always lie, a solution is a row where column X matches column $\mathcal{S}_\mathsf{X}$, for X = A **and** X = B.

Row 4 is the unique solution: both are knaves, both are lying. □

The atoms of logic are called *propositions*. A proposition is a statement that has one and only one truth value (T or F).

e.g., $1 + 1 = 2$: T and $2 + 2 = 5$: F (in $\mathbb{R}$).

e.g., 'A is a knight', 'B is a knave' in the puzzle above.

e.g., 'This statement is false' is not a proposition. Can't be assigned T or F (try it!). Cf. Liar's Paradox, Barber's Paradox. Self-contradictions.

e.g., All humans have two eyes. [Premise]
John is a human. [Premise]
Therefore, John has two eyes. [Conclusion]

Connectives

Compound statements are built up recursively from atoms joined together using *connectives*; e.g.,

'and'	$\wedge$	[conjunction]
'or'	$\vee$	[disjunction]
'not'	$\neg$	[negation].

These are three basic Boolean operators: multivariable functions whose values and arguments are Booleans (T or F). $\wedge, \vee$ are binary Boolean operators; $\neg$ is unary.

Each connective is defined by its truth table.

Truth table definition of $\wedge$:

a	b	$a \wedge b$
T	T	T
T	F	F
F	T	F
F	F	F

a, b are propositional variables. So, $a \wedge b$ is true only when a **and** b are both true.

$4 = 2^2$ rows in the table since $\wedge$ is a binary operator.

Table shows that $a \wedge b$ always has the same truth value as $b \wedge a$; i.e., $\wedge$ is *commutative*.

The table for $\wedge$ can be used to evaluate longer expressions $a \wedge (b \wedge c)$, $(a \wedge b) \wedge c$, etc. (*Exercise*: prove that these two expressions always have the same truth value. Need $2^3 = 8$ rows in the truth table.)

Truth table definition of $\vee$:

a	b	$a \vee b$
T	T	T
T	F	T
F	T	T
F	F	F

So $a \vee b$ is true only when either a **or** b is true.

Again, 4 rows in the table since the operator is binary.

Also again, the table shows that the connective is commutative: $a \vee b$ always has the same truth value as $b \vee a$.

And $\vee$ is *associative*, i.e., $a \vee (b \vee c)$ always has the same truth value as $(a \vee b) \vee c$. [*Exercise*: prove it.]

$\neg$, negation, is a unary operator, so has just 2 rows in its truth table:

a	$\neg a$
T	F
F	T

$\neg(\neg a)$ always has same truth value as a: $\neg$ is an *self-inverse* operator.

Another important binary connective is implication, denoted $\rightarrow$.

a	b	$a \rightarrow b$
T	T	T
T	F	F
F	T	T
F	F	T

N.B. T $\rightarrow$ F should be false, as indicated: if premises in a true implication are true then the conclusion should true.

Recall the definition of the implication connective $\rightarrow$ ("if...then") :

a	b	$a \rightarrow b$
T	T	T
T	F	F
F	T	T
F	F	T

If the premise a fails to be satisfied (last two rows of the table) then the conclusion can be true or false without invalidating the argument; the compound statement is *vacuously* true if the premises are false.

e.g., interpret "If Earth is 2000 years old then I am President" as true;

also interpret "If Earth is 2000 years old then I am not President" as true (premise is false in both cases).

e.g.,

If it rains then I am carrying an umbrella

is T on clear days, and on rainy days when I have an umbrella; F otherwise (i.e., the single case that it is a rainy day and I don't have any umbrella).

Just remember

- T can never imply F
- F can imply anything.

(Cf. definition of empty set $\emptyset$ in set theory.)

Truth table also shows that $a \to b$ is **not** the same as $b \to a$.

The biconditional connective $\leftrightarrow$ (iff: "if and only if") is defined by the truth table

a	b	$a \leftrightarrow b$
T	T	T
T	F	F
F	T	F
F	F	T

This is really $(a \to b) \wedge (b \to a)$; compare the truth tables (exercise).

Iff statements are common in mathematics; e.g.,

> *a positive integer is prime iff it is greater than* 1 *and its only positive integer divisors are itself and* 1.

and, e.g.,

> *a positive integer* n *is odd iff* n^2 *is odd.*

Tautologies and contradictions

A *tautology* is a statement that is *always true*, regardless of the truth values of constituent propositional variables.

A proposition is a *contradiction* if its truth value is F for all possible combinations of the truth values of its propositional variables.

We can determine whether a given statement is a tautology, a contradiction, or neither, by truth table.

e.g., $a \vee \neg a, \quad a \to (b \to a), \quad \neg(\neg a) \to a$ are all tautologies.

e.g., $a \wedge a$ is not a tautology, nor a contradiction: it is T if a is T, and F if a is F.

e.g., $p \wedge \neg p$, $a \wedge \mathrm{F}$ are both contradictions.

Example. $(\neg a \to \neg b) \to (b \to a)$ is a tautology.

Use the truth table definitions of $\neg$ and $\to$ to construct truth table for this compound statement.

a	b	$\neg a$	$\neg b$	$\neg a \to \neg b$	$b \to a$	$(\neg a \to \neg b) \to (b \to a)$
T	T	F	F	T	T	T
T	F	F	T	T	T	T
F	T	T	F	F	F	T
F	F	T	T	T	T	T

Final column (all T) means that the statement is a tautology.

Indeed, $(\neg a \to \neg b) \leftrightarrow (b \to a)$ is a tautology (check: the truth table above only needs modifying in a new final column, using definition of $\leftrightarrow$).

Two statements are *(logically) equivalent* if they have the same truth value for each assignment of truth values to constituent propositions; otherwise, they are *inequivalent*.

If p and q are equivalent then we write $p \equiv q$; if p and q are inequivalent then we write $p \not\equiv q$.

e.g., if p is a tautology then $p \equiv \mathsf{T}$.

e.g., if p is a contradiction then $p \equiv \mathsf{F}$.

e.g., $a \to b \not\equiv b \to a$.

e.g., commutativity of $\wedge$, $\vee$: $a \wedge b \equiv b \wedge a$, $a \vee b \equiv b \vee a$.

e.g., associativity of $\wedge$, $\vee$: $(a \wedge b) \wedge c \equiv a \wedge (b \wedge c)$,
$(a \vee b) \vee c \equiv a \vee (b \vee c)$.

Example. We can decide logical equivalence by truth table.

The following shows that $\boxed{a \to b \equiv \neg a \vee b}$

a	b	$a \to b$	$\neg a$	$\neg a \vee b$
T	T	T	F	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

e.g.,

If you don't attend lectures then you will fail.

is equivalent to

Either you attend lectures, or you will fail.

Some other important equivalences are below. Again, all can be proved by truth table.

Distributivity of $\wedge$ over $\vee$, and vice versa:

- $a \wedge (b \vee c) \equiv (a \wedge b) \vee (a \wedge c)$
- $a \vee (b \wedge c) \equiv (a \vee b) \wedge (a \vee c)$.

De Morgan's Laws:

- $\neg(a \vee b) \equiv \neg a \wedge \neg b$
- $\neg(a \wedge b) \equiv \neg a \vee \neg b$.

Strictly speaking, negation does not distribute over $\wedge$, $\vee$: it flips each connective to the other.

Note that the second De Morgan law follows from the first (and vice versa). Assuming the first law $\neg(a \vee b) \equiv \neg a \wedge \neg b$ and applying $\neg$ to both sides:

$$\neg\neg(a \vee b) \equiv \neg(\neg a \wedge \neg b).$$

Then use that $\neg\neg$ is the identity:

$$a \vee b \equiv \neg(\neg a \wedge \neg b).$$

Replacing a by $\neg a$ and b by $\neg b$:

$$\neg a \vee \neg b \equiv \neg(\neg\neg a \wedge \neg\neg b).$$

Finally, using $\neg\neg = \mathrm{id}$ again:

$$\neg a \vee \neg b \equiv \neg(a \wedge b),$$

which is the second De Morgan law.

Variations of $p \to q$

- $q \to p$ is the *converse* of $p \to q$.
- $\neg p \to \neg q$ is the *inverse* of $p \to q$.
- $\neg q \to \neg p$ is the *contrapositive* of $p \to q$.

Note:

1. $p \to q \equiv \neg q \to \neg p$.
2. converse $\equiv$ inverse (follows from 1. by swapping q, p).
3. An implication is not equivalent to its converse; hence is not equivalent to its inverse.

e.g. (1., equivalence of implication with its contrapositive):
"I'm sad when it rains" $\equiv$ "If I'm not sad then it's not raining."

Example (proof that an implication $\equiv$ its contrapositive). Recall $a \to b \equiv \neg a \vee b$. Label this equivalence (α). Then

$$\begin{aligned} \neg q \to \neg p &\equiv \neg\neg q \vee \neg p && (\alpha) \\ &\equiv q \vee \neg p && \neg\neg = \text{ id} \\ &\equiv \neg p \vee q && \vee \text{ commutative} \\ &\equiv p \to q && (\alpha). \end{aligned}$$

Example. Proof that $p \to q \not\equiv \neg p \to \neg q$:

p	q	$\neg p$	$\neg q$	$p \to q$	$\neg p \to \neg q$
T	T	F	F	T	T
T	F	F	T	F	T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Predicates

We begin with a little set theory. A *set* is a collection of *elements* (no other structure assumed).

Let S be a set. If x is an element S then we write $x \in S$. If x is not an element of S then we write $x \notin S$.

Each set is specified entirely by its elements. Thus, two sets A and B are equal, denoted $A = B$, if and only if $a \in A$ implies $a \in B$ and $b \in B$ implies $b \in A$.

Standard notation expresses the elements that define the set in some explicit way, usually between braces. E.g., $\{1, \ldots, 99\}$; $\{x, y, z\}$; $\{a \in \mathbb{Z} \mid a \text{ is divisible by } 2\}$, $\emptyset$. (Note that some sources use ':' in place of '|' in the definition of a set.)

A set is *finite* if it has just a finite number of elements. E.g., $\{0, 1\}$ is finite; the set $\mathbb{R}$ of real numbers is not.

A *predicate* $P(x)$ is a statement involving a variable x, that becomes a proposition (i.e., has truth value T or F) when x is replaced by a value (in the domain of P). So a predicate is a special kind of function.

Example. Let $E(n)$ be the predicate "n is even", where $n \in \mathbb{Z}$. The statement is either T or F, depending on n; e.g., $E(14) \equiv$ T; $E(13) \equiv$ F.

Predicates can be combined by connectives. E.g., if $P(n) =$ "n is prime", then $E(n) \wedge P(n) \equiv$ T for just one n, namely 2.

Predicates can have more than one variable. E.g., $L(x, y) = x < y$ for $x, y \in \mathbb{R}$.

Quantifying predicates

Predicates can also be turned into propositions by *quantification*.

Let $P(x)$ be a predicate and S be a set.

UNIVERSAL QUANTIFICATION: $\forall x \in S, P(x)$ is the proposition "*for all* elements x of S, the proposition $P(x)$ has value T".

EXISTENTIAL QUANTIFICATION: $\exists x \in S, P(x)$ is the proposition "*for some* element x of S (*there exists* an x), the proposition $P(x)$ has value T".

Note: if $S = \{x_1, x_2, \ldots, x_n\}$ then

- $\exists x \in S, P(x)$ is equivalent to $P(x_1) \vee P(x_2) \vee \cdots \vee P(x_n)$,
- $\forall x \in S, P(x)$ is equivalent to $P(x_1) \wedge P(x_2) \wedge \cdots \wedge P(x_n)$.

Negation of quantifiers/duality: a universal quantifier negates to existential quantifier, and vice versa. We can think of this phenomenon as De Morgan laws for quantifiers:

$$\neg(\forall x \in S, P(x)) \equiv \exists x \in S, \neg P(x).$$

$$\neg(\exists x \in S, P(x)) \equiv \forall x \in S, \neg P(x).$$

Example. With obvious interpretations, "Not all Martians are green" is $\neg(\forall x \in M, G(x)) \equiv \exists x \in M, \neg G(x)$, i.e., there is a Martian who is not green.

Example. "It isn't true that some Martians are green" is $\neg(\exists x \in M, G(x)) \equiv \forall x \in M, \neg G(x)$, i.e., no Martian is green.

Example (proof of De Morgan for quantifiers when S is finite).

Say $S = \{x_1, \ldots, x_n\}$. Then

$$\begin{aligned} \neg(\forall x \in S, P(x)) &\equiv \neg(P(x_1) \wedge \cdots \wedge P(x_n)) \\ &\equiv \neg P(x_1) \vee \cdots \vee \neg P(x_n) \\ &\equiv \exists x \in S, \neg P(x). \end{aligned}$$

Where we used De Morgan for $\wedge, \vee$ in the second line.

Validity of arguments

A (logical) *argument* is a list of statements, ending in a conclusion.

More formally, an argument is a list $p_1, p_2, \ldots, p_n, \therefore c$ where the $p_1, \ldots, p_n$ are *premises* and c is the *conclusion*.

An argument is *valid* if the conclusion follows necessarily from the premises.

The validity of an argument depends only on its logical form, not on the content.

The argument $p_1, p_2, \ldots, p_n, \therefore c$ is valid iff the proposition $(p_1 \wedge p_2 \wedge \cdots \wedge p_n) \rightarrow c$ is a tautology.

A method to test argument validity

1. Identify the premises and the conclusion.
2. Construct a truth table showing the truth values of all premises and the conclusion.
3. A *critical row* of the truth table is a row of the truth table in which all the premises are true. Identify the critical rows and check them as follows.
   - If the **conclusion is true** in every critical row then the argument structure is **valid**. ($\forall$)
   - If there is a critical row in which the **conclusion is false**, then the argument is **invalid**. ($\exists$)

This method is justified simply from the definition of $\rightarrow$ and $\wedge$: the only way $a \rightarrow b$ fails to be T is when $a \equiv$ T and $b \equiv$ F; the only way $x \wedge y$ is T is when both x and y are T.

Example. Premises: $p_1 = (p \to q \vee \neg r)$, $p_2 = (q \to p \wedge r)$.
Conclusion: $c = (p \to r)$.

Check validity of $p_1, p_2, \therefore c$ by the above method:

p	q	r	$\neg r$	$q \vee \neg r$	$p \wedge r$	p_1	p_2	c
T	T	T	F	T	T	T	T	T
T	T	F	T	T	F	T	F	
T	F	T	F	F	T	F	T	
T	F	F	T	T	F	T	T	F
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Fourth row is a critical row in which c is F; thus, the argument is invalid. (*Exercise.* Complete the truth table. Are there any other critical rows with $c \equiv$ F?)

Example. (Modus ponens): $p \to q, p, \therefore q$.

e.g.: If Socrates is human then he is mortal. Socrates is human. Therefore, Socrates is mortal.

Proof by truth table:

p	q	$p \to q$	p	q
T	T	T	T	T
T	F	F	T	
F	T	T	F	
F	F	T	F	

Or, use the equivalence $p \to q \equiv \neg p \vee q$ proved earlier.

Example. (Modus tollens): $p \to q, \neg q, \therefore \neg p$.

e.g.: If pigs can fly then they have wings. Pigs don't have wings. Therefore, pigs cannot fly.

Proof by truth table:

p	q	$p \to q$	$\neg q$	$\neg p$
T	T	T	F	
T	F	F	T	
F	T	T	F	
F	F	T	T	T

Or, use the equivalence $p \to q \equiv \neg p \vee q$.

Example. (Transitivity of implication): $p \to q, q \to r, \therefore p \to r$.

Proof by truth table:

p	q	r	$p \to q$	$q \to r$	$p \to r$
T	T	T	T	T	T
T	T	F		F	
T	F	T	F		
T	F	F	F		
F	T	T	T	T	T
F	T	F		F	
F	F	T	T	T	T
F	F	F	T	T	T

Common logical fallacies

The *converse fallacy*: $p \to q, q, \therefore p$.

e.g.:

If Socrates is human then he is mortal.

Socrates is mortal.

$\therefore$ Socrates is human.

This is **wrong**, as, e.g., the truth table method shows:

p	q	$p \to q$	q	p
T	T	T	T	T
T	F	F		
F	T	T	T	F

The *inverse fallacy*: $p \to q, \neg p, \therefore \neg q$.

e.g.:

If pigs can fly then they have wings.

Pigs cannot fly.

$\therefore$ Pigs do not have wings.

Again, **wrong**:

p	q	$p \to q$	$\neg p$	$\neg q$
T	T	T	F	
T	F	F		
F	T	T	T	F

Example. Recall the knights and knaves puzzle.

Knights always tell the truth; knaves always lie.
A says: "Exactly one of us is a knave".
B says: "At least one of us is a knight".
Each of A, B is either a knight or a knave.

We can now formalize the argument to solve this puzzle.

Let a, b respectively be the propositions 'A is a knight', 'B is a knight'. Premises:

1. $a \to \neg b$ (from A's statement)
2. $\neg a \to \neg b$ (from A's statement)
3. $b \to a \vee b$ (from B's statement)
4. $\neg b \to \neg a \wedge \neg b$ (from B's statement).

From $a \vee \neg a$, 1. $(a \to \neg b)$, 2. $(\neg a \to \neg b)$, and modus ponens, we conclude $\neg b$. So B is definitely a knave.

Next, combining $\neg b$ with 4. $(\neg b \to \neg a \wedge \neg b)$, we conclude (by modus ponens) $\neg a \wedge \neg b$.

Hence A and B are both knaves.

Set theory

Recall our understanding of a set as being completely determined by the elements that it contains.

A set B is a *subset* of a set A if each element of B is also an element of A.

If B is a subset of A then we write $B \subseteq A$.

Thus, $B \subseteq A$ if $b \in A$ for all $b \in B$.

Also note that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. Equality of sets is often proved by proving both these containments.

All sets are assumed to be subsets of a universal set, or universe U.

A set is *finite* if it has just finitely many elements. The *size* (or *cardinality*) of a finite set is the number of its elements.

Let $A, B \subseteq U$.

- The *union* of A and B is the set
  $A \cup B := \{x \in U \mid x \in A \text{ or } x \in B\}$ (cf. $\vee$ in logic).
- The *intersection* of A and B is the set
  $A \cap B := \{x \in U \mid x \in A \text{ and } x \in B\}$ (cf. $\wedge$ in logic).
- The *set difference* of A and B is the set
  $A \backslash B := \{x \in U \mid x \in A \text{ and } x \notin B\}$.
- The *complement* of A (in U) is the set
  $A' := \{x \in U \mid x \notin A\}$ (cf. $\neg$ in logic).

Note: $A \cap A' = \emptyset$, $A \cup A' = U$, $\emptyset' = U$, $U' = \emptyset$.

The various set theory operations can be combined to produce identities.

Example. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ for all $A, B, C \subseteq U$.

To prove this claim, let $x \in A \cap (B \cup C)$.
By definition of $\cap$ & $\cup$, $x \in A$ and $x \in B$ or $x \in C$.
If $x \in B$ then $x \in A \cap B$, so $x \in (A \cap B) \cup (A \cap C)$.
Similarly, if $x \in C$ then $x \in A \cap C$, so $x \in (A \cap B) \cup (A \cap C)$.

Hence $\boxed{A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C).}$

For the other containment, let $x \in (A \cap B) \cup (A \cap C)$.
Then $x \in A \cap B$ or $x \in A \cap C$.
If $x \in A \cap B$ then $x \in A$ and $x \in B$, so $x \in A$ and $x \in B \cup C$.
Thus $x \in A \cap (B \cup C)$.
Similarly, $x \in A \cap C$ implies $x \in A \cap (B \cup C)$ again. Hence

$$\boxed{(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C).}$$

Combining the two boxes proves the claim.

We list a few other set theory identities. All are either clear from definitions, or can be proved 'elementwise' as in the proof at the end of the previous lecture.

Let $A, B, C \subseteq U$.

- Commutative laws: $A \cup B = B \cup A$, $A \cap B = B \cap A$
- Associative laws: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$
- Distributive laws: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Double negation: $A'' = A$
- De Morgan laws: $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$.

Boolean algebra

Sets, together with the operations $\cup$, $\cap$, $'$, and the constants $\emptyset$, U, behave similarly to propositions, together with analogous operations $\vee$, $\wedge$, $\neg$ and the constants F, T.

Both are examples of an algebraic structure with operations $\cdot$, $+$, $'$ and constants 0, 1, called a *Boolean algebra*.

Each *logical equivalence* translates to a corresponding *set identity*, and vice versa.

Duality The *dual* of a set identity is obtained by swapping $\cup$ with $\cap$ and swapping $\emptyset$ with U (check the previous page!).

The dual of a valid set identity is also a valid set identity. Thus only one of them needs to be proved.

The two definitions on this page and the next fall under the theme of 'sets of sets'.

Let A be a set. The *power set* of A, denoted $P(A)$, is the set $\{X \mid X \subseteq A\}$.

That is, $P(A)$ is the set of all subsets of A.

Example. Let $A = \{1, 2, 3\}$.

Then $\emptyset \in P(A)$ (the unique subset of size 0).

Subsets of size 1: $\{1\}$, $\{2\}$, $\{3\}$.

Subsets of size 2: $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

A unique subset of size 3: A itself.

Hence $P(A)$ has size $1 + 3 + 3 + 1 = 8 = 2^3 = 2^{\text{size of } A}$.

$P(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

Let A be a set.

A *partition* of A is a set $P = \{P_1, P_2, \ldots\}$ where $P_1, P_2, \ldots$ are **nonempty** subsets of A satisfying the following conditions:

- $P_i \cap P_j = \emptyset$ for all $i \neq j$.
- $A = P_1 \cup P_2 \cup \cdots$

That is

- different P_i, P_j are *disjoint* (i.e., have no elements in common)
- every element of A is in at least one (hence only one, by the previous) P_i.

The subsets P_i are called *parts* of the partition P of A.

Example. $A = \{1, 2, 3\}$ has the following partitions:

$$\{\{1\}, \{2\}, \{3\}\},\ \{\{1, 2\}, \{3\}\},\ \{\{1\}, \{2, 3\}\},\ \{\{1, 3\}, \{2\}\},\ \{\{1, 2, 3\}\}.$$

The *Cartesian product* of sets A and B is the set

$$A \times B = \{(a, b) \mid a \in A,\, b \in B\}$$

of all (ordered) pairs (a, b).

Examples

- If $A = \{1, 2, 3\}$ and $B = \{x, y\}$ then $A \times B = \{(1, x), (1, y), (2, x), (2, y), (3, x), (3, y)\}$. Note: A has size 3, B has size 2, and $A \times B$ has size $3.2 = 6$.
- If $A = \{1, 2, 3\}$ then A^2, i.e., $A \times A$, is $\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$. Size $9 = 3^2$.

More generally, the Cartesian product of n sets $S_1, S_2, \ldots, S_n$ is $S_1 \times S_2 \times \cdots \times S_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in S_i,\, 1 \leq i \leq n\}$.
Elements of $S_1 \times S_2 \times \cdots \times S_n$ are called *n-tuples*.
Write A^n for $A \times A \times \cdots \times A$ (n factors).

Relations

Let X, Y be sets. A *relation* from X (called the *domain*) to Y (called the *codomain*) is a subset R of $X \times Y$.

If $X = Y$ then $R \subseteq X \times X$ is said to be a *relation on* X.

Example. If $A = \{1, 2, 3\}$ and $B = \{x, y\}$ then here are some relations from A to B: $\{(1, x)\}$, $\{(2, y), (3, x)\}$, $\{(1, y), (2, y), (3, y)\}$, $\{(1, x), (1, y), (2, x), (2, y)\}$,...
(How many relations from this A to this B are there?)

If R is a relation from X to Y, and $(x, y) \in R$, then we say x *is related to* y, and write xRy.

Let R be a relation on the set X.

- R is *reflexive* if xRx for all $x \in X$.
- R is *symmetric* if, for all $x, y \in X$, whenever xRy then yRx too.
- R is *transitive* if, for all $x, y, z \in X$, whenever xRy and yRz then xRz.

A relation R on a set X that is reflexive, symmetric, and transitive is called an *equivalence relation* on X.

Example. If $A = \{a, b, c\}$ then $\{(a, a), (b, b), (c, c)\}$, $\{(a, a), (b, b), (c, c), (a, b), (b, a)\}$, $\{(a, a), (b, b), (c, c), (a, b), (b, a), (b, c), (c, b), (a, c), (c, a)\}$ are all equivalence relations on A. In the first, each element is related only to itself; in the second, a and b are related, but not related to c; in the third, a, b, c are all related.

Example. Consider the relation $\leq$ on $\mathbb{R}$.

$\leq$ is reflexive: $a \leq a$ for all $a \in \mathbb{R}$.

$\leq$ is **not** symmetric: if $a \leq b$ then $b \leq a$ if and only if $a = b$.

$\leq$ is transitive: if $a \leq b$ and $b \leq c$ then $a \leq c$.

Example. Define a relation R on $\mathbb{R}$ by: xRy if and only if $x = \pm y$. Then R is an equivalence relation on $\mathbb{R}$ (confirm all three properties on the checklist).

Example. Let $A = \mathbb{R}^2 \setminus \{(0,0)\}$, i.e., the x-y plane without the origin.

Define a relation R on A by: pRq if and only if p, q are on the same straight line through $(0,0)$.

Quickly confirm that R is reflexive and symmetric.

If p, q are on the line ℓ_1 through $(0,0)$, and q, r are on the line ℓ_2 through $(0,0)$, then $\ell_1 = \ell_2$ ($q \neq (0,0)$ cannot be on two different lines through the origin).

Thus p and r lie on the same line $\ell_1 = \ell_2$: R is transitive.

Consequently R is an equivalence relation on A.

Equivalence relations = partitions

Suppose that R is an equivalence relation on a set X.

For $x \in X$, denote by $[x]$ *the equivalence class* of x: this is the set of all $y \in X$ such that yRx. N.B. xRy if and only if $[x] = [y]$.

Note that $[x] \neq \emptyset$: since R is reflexive, i.e., xRx, the equivalence class $[x]$ always contains at least x itself.

Example. Let T be the equivalence relation on the set $\mathbb{Z}$ of integers defined by $x\,T\,y$ if and only if $x + y$ is even. (Check that $T = \{(x, y) \mid x, y \in \mathbb{Z},\, x + y \text{ even}\}$ is indeed an equivalence relation on $\mathbb{Z}$.)

Then for this relation T, $[0] =$ all even integers; $[1] =$ all odd integers. Note that $\mathbb{Z} = [0] \cup [1]$, disjoint union: a partition of $\mathbb{Z}$.

Denote by X/R the set $\{[x] \mid x \in X\}$ of all **different** equivalence classes (also called the *quotient set* for the equivalence relation R on X).

Suppose that P is a partition of X. For $x \in X$, denote by $P(x)$ the (unique) part of P that contains x.

Theorem

1. *If R is an equivalence relation on the set X, then X/R is a partition of X.*
2. *Conversely, if P is a partition of a set X, then the relation $R = \{(x, y) \in X^2 \mid P(x) = P(y)\}$ is an equivalence relation on X.*

Proof.

The proofs of both 1. and 2. just walk through the definitions.

Prove only 1. here; proof of 2. is left as an exercise.

We are assuming that R is an equivalence relation on X.
Since R is reflexive, each $x \in X$ is an element of its equivalence class $[x]$. Therefore $X = \cup_{x \in X} [x]$.

Next, suppose that $a \in [x] \cap [y]$ for $a, x, y \in X$.

Then xRa and aRy. Transitivity of R implies that xRy, i.e., $[x] = [y]$. Hence, *different* equivalence classes must have *empty* intersection—they are disjoint.

Together with $X = \cup_{x \in X} [x]$, this proves that the equivalence classes comprise a partition of X. □

Example. Consider the partition $\{\{1\},\{2,3\},\{4,5,6\}\}$ of $A = \{1,2,3,4,5,6\}$.

The equivalence relation on A corresponding to this partition is

$\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(2,3),(3,2),(4,5),(5,4),$
$(4,6),(6,4),(5,6),(6,5)\} \subset A \times A.$

Example. Define a relation $\equiv$ ('congruence modulo 3') on $\mathbb{Z}$ by: $a \equiv b$ if and only if $a - b$ is exactly divisible by 3. (Check that this is an equivalence relation on $\mathbb{Z}$.) Then the partition of $\mathbb{Z}$ corresponding to this equivalence relation is $[0] \cup [1] \cup [2]$ where

$$[0] = \{\ldots, -9, -6, -3, 0, 3, 6, \ldots\}$$
$$[1] = \{\ldots, -8, -5, -2, 1, 4, 7, \ldots\}$$
$$[2] = \{\ldots, -7, -4, -1, 2, 5, 8, \ldots\}.$$

Functions

Let A, B be sets.

A *function* f from A (the *domain*) to B (the *codomain*) is a subset $f \subseteq A \times B$ such that

$$\text{for each } a \in A \ \ \exists \ \textbf{unique } b \in B \text{ such that } (a, b) \in f.$$

(In particular, if A is finite, then so too is every function f with domain A; f has the same size as A.)

We write $f : A \to B$ for a function f from A to B; also we write $f(a) = b$ if b is the unique element of B such that $(a, b) \in f$.

Example. Let $A = \{a, b, c, d, e\}$ and $B = \{\alpha, \beta, \gamma, \delta\}$. Then $f = \{(a, \beta), (b, \alpha), (c, \delta), (d, \delta), (e, \delta)\}$ is a function $f : A \to B$; $f(a) = \beta$, $f(b) = \alpha$, $f(c) = f(d) = f(e) = \delta$.

However, $\{(a, \beta), (a, \alpha), (b, \delta), (c, \gamma), (d, \delta)(e, \beta)\}$ is *not* a function from A to B (a occurs twice in the first component);

neither is $\{(a, \delta), (b, \gamma), (c, \alpha), (e, \beta)\}$ ($d \in A$ does not appear).

Example. $f = \{(x, x^2) \mid x \in \mathbb{R}\}$ is a function with domain $\mathbb{R}$ and codomain the set of non-negative real numbers;

$f = \{(x, x^3) \mid x \in \mathbb{R}\}$ is a function with domain = codomain = $\mathbb{R}$.

Example. Let $A \subseteq B$. The *characteristic function* $c_A : B \to \{0, 1\}$ is defined by: $c_A(b) = 1$ if and only if $b \in A$.

Example. Let X be a set. The relation $\mathrm{id}_X = \{(x, x) \mid x \in X\}$ is a function $\mathrm{id}_X : X \to X$ (the *identity function*). □

Every function $f : A \to B$ is a relation from A to B; *not* vice versa.

Example. Let $A = \{(x, y) \mid x, y \in \mathbb{R},\ x^2 + y^2 = 1\}$. This is a relation, *not* a function on $\mathbb{R}$, e.g., $(0, 1)$ and $(0, -1)$ are both in A. Draw the graph: it fails the 'vertical line test'. □

Remember: a function is a triple: its domain, its codomain, and the 'rule' that assigns to each element of the domain a unique element of the codomain.

Thus, two functions are equal if and only if they have the same domain, D, the same codomain, C, and they are equal as subsets of $D \times C$ (have the same rule).

Injections, surjections, bijections

A function $f : A \to B$ is *injective* if

$$\forall\, a_1, a_2 \in A,\ \ f(a_1) = f(a_2) \ \text{ implies } a_1 = a_2$$

(roughly, we have a cancelation law for injective f). Thus $f : A \to B$ is injective if for every $b \in B$, there is *at most one* $a \in A$ such that $f(a) = b$.

Injective functions are also called *one-to-one* and *injections*.

Example. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not injective, e.g., $f(-1) = 1 = f(1)$ but $1 \neq -1$.

Example. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is injective.

Graphs of the previous examples fail/pass the 'horizontal line test'.

A function $f : A \to B$ is *surjective* if

$$\forall\, b \in B \;\; \exists\, a \in A \text{ such that } f(a) = b.$$

That is, $f : A \to B$ is surjective if for every $b \in B$, there is *at least* one $a \in A$ such that $f(a) = b$.

Surjective functions are also said to be *onto* and are called *surjections*.

Example. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not surjective, e.g., there is not real number x such that $f(x) = -1$.

Example. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3$. Then f is surjective.

$f : A \to B$ is a *bijection*, or is *bijective*, if it is injective and surjective: for each $b \in B$, $\exists$ *unique* $a \in A$ such that $f(a) = b$. Bijections are also called *one-to-one correspondences*.

Example. Let $X = \{a, b, c, d, e\}$, $Y = \{1, 2, 3, 4\}$.
Define functions $f : X \to Y$, $g : X \to X$, $h : Y \to X$ by

f :

a	b	c	d	e
3	4	2	4	1

, g :

a	b	c	d	e
b	d	e	a	c

, h :

1	2	3	4
a	c	b	e

Then

- f is surjective, but not injective ($f(b) = f(d)$), hence not a bijection;
- g is injective and surjective, hence a bijection;
- h is injective, but not surjective (there is no $y \in Y$ such that $h(y) = d$) hence not a bijection.

Compare definition of function $f : A \to B$,

for each $a \in A$, $\exists$ *unique* $b \in B$ *such that* $f(a) = b$

with definition of bijection,

for each $b \in B$, $\exists$ *unique* $a \in A$ *such that* $f(a) = b$.

Example. The identity map $\mathrm{id}_X : X \to X$ is a bijection.

Example. Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) = 2n$.

Then f is injective: if $f(m) = f(n)$ then $2m = 2n$, so $m = n$.

However f is not surjective: there is no natural number m such that $f(m) = 2m = 1$. Hence f is not a bijection from $\mathbb{N}$ to $\mathbb{N}$.

Example. If $a, b \in \mathbb{R}$ and $a \neq 0$ then $f(x) = ax + b$ defines a function $f : \mathbb{R} \to \mathbb{R}$ that is a bijection. (Check.)

Let A and B be finite sets; say $A = \{a_1, \ldots, a_n\}$ has size n and B has size m.

If there exists an *injection* $f : A \to B$ then $f(a_1), \ldots, f(a_n)$ are all different elements of B. Hence $n \leq m$.

If there exists a *surjection* $f : A \to B$ then each element of B must occur at least once on the list $f(a_1), \ldots, f(a_n)$. Hence $n \geq m$.

Thus, if there is a bijection between A and B then these sets must have the same size. Conversely, if finite sets have the same size then there is a bijection between them.

Let $f : X \to Y$ be a function.

The *image* $f(X) = \{f(x) \mid x \in X\}$ is a subset of Y.

The relation $\sim_f$ on X defined by $x_1 \sim_f x_2$ if $f(x_1) = f(x_2)$ is an equivalence relation (check).

The equivalence classes $[x] = \{x' \in X \mid f(x') = f(x)\}$ for $\sim_f$ form (as usual) a partition $X/\sim_f$ of X, called the *kernel* of f.

Theorem

1. *The function $F : X/\sim_f \to f(X)$ given by $F([x]) = f(x)$ is a bijection between the kernel of f and the image $f(X)$.*
2. *Conversely, if $Y_1 \subseteq Y$, $\approx$ is any equivalence relation on X, and $G : X/\approx \to Y_1$ is a bijection, then $g(x) = G([x])$ for $x \in X$ defines a function $g : X \to Y$.*

Proof of part 1.

First, we check that F is *well-defined*. That is, the definition of F makes a choice of element from each equivalence class to define each output of F; need to see that changing the choice does not change F's output. So, suppose that $x' \in [x]$; then $f(x') = f(x)$, and hence $F([x]) = f(x)$ does not change if we choose x' from $[x]$ instead of x.

Now suppose that $x_1, x_2 \in X$ and $F([x_1]) = F([x_2])$. Then $f(x_1) = f(x_2)$ by definition of F. Hence x_1, x_2 are in the same $\sim_f$-equivalence class, i.e., $[x_1] = [x_2]$. Thus F is injective.

That F is surjective is obvious. $\square$

Composition and inverse functions

Let $f : A \to B$ and $g : B \to C$ be functions. The *composition* of f and g, denoted $g \circ f$ (note the order), is defined by

$$(g \circ f)(a) = g(f(a)) \quad \forall a \in A.$$

Note that the composite $g \circ f$ is a function, $g \circ f : A \to C$.

Composition is a kind of *multiplication* of functions.

Theorem

Composition of functions is associative, i.e., if $f : A \to B$, $g : B \to C$, $h : C \to D$ *are functions, then* $h \circ (g \circ f) = (h \circ g) \circ f$.

Proof.

First, $h \circ (g \circ f)$, $(h \circ g) \circ f$ have the same domain A and the same codomain D. Secondly, for all $a \in A$ we have
$h \circ (g \circ f)(a) = h((g \circ f)(a)) = h(g(f(a)))$; and
$(h \circ g) \circ f(a) = (h \circ g)(f(a)) = h(g(f(a)))$. □

Theorem

Let $f : A \to B$, $g : B \to C$ be functions.

(i) *If f, g are injective then $g \circ f : A \to C$ is injective.*

(ii) *If f, g are surjective then $g \circ f : A \to C$ is surjective.*

Proof.

(i) If $(g \circ f)(a_1) = (g \circ f)(a_2)$ then $g(f(a_1)) = g(f(a_2))$, so $f(a_1) = f(a_2)$ because g is injective. Then $a_1 = a_2$ because f is injective.

(ii): Exercise. □

Corollary

The composition of bijections is a bijection.

Neither converse of (i), (ii) in the theorem is true.

Example. Let $A = \{a_1, a_2\}$, $B = \{b_1, b_2, b_3\}$, $C = \{c_1, c_2\}$.
Define $f : A \to B$ and $g : B \to C$ by

$$f : \begin{array}{|c|c|} \hline a_1 & a_2 \\ \hline b_1 & b_2 \\ \hline \end{array}, \qquad g : \begin{array}{|c|c|c|} \hline b_1 & b_2 & b_3 \\ \hline c_1 & c_2 & c_2 \\ \hline \end{array}.$$

Then f is injective, not surjective.

And g is surjective, not injective.

But $(g \circ f)(a_1) = g(f(a_1)) = g(b_1) = c_1$
and $(g \circ f)(a_2) = g(f(a_2)) = g(b_2) = c_2$,
showing that $g \circ f : A \to C$ is a bijection.

Theorem

Let $f : A \to B$ and $g : B \to C$ be functions.

(i) *If $g \circ f$ is injective then f is injective.*

(ii) *If $g \circ f$ is surjective then g is surjective.*

Proof.

(i) Suppose that $f(a_1) = f(a_2)$ for $a_1, a_2 \in A$.

Then $(g \circ f)(a_1) = g(f(a_1)) = g(f(a_2)) = (g \circ f)(a_2)$.

Hence $a_1 = a_2$ because $g \circ f$ is injective.

(ii): Exercise. □

Let $f : X \to Y$ be a bijection. Define a function $g : Y \to X$ by

$$g(y) = x \quad \text{if and only if} \quad f(x) = y.$$

That is, $g = \{(y, x) \in Y \times X \mid (x, y) \in f\}$. (Check that g as defined *is* a function $Y \to X$; use injectivity and surjectivity of f.)

g is a bijection too:

Suppose that $g(y_1) = g(y_2)$ for some $y_1, y_2 \in Y$. Since f is surjective, $f(x_1) = y_1$ and $f(x_2) = y_2$ for some $x_1, x_2 \in X$. By definition of the function g we have $g(y_1) = x_1$ and $g(y_2) = x_2$. But $g(y_1) = g(y_2)$, so $x_1 = x_2$. Hence $y_1 = f(x_1) = f(x_2) = y_2$. Thus g is injective.

Next, let x be any element of X. Then $f(x) = y \in Y$ say. By definition of g we have $g(y) = x$. This shows that $X = g(Y)$, i.e., g is surjective too.

So: there is a bijection f from a set X to a set Y if and only if there is a bijection g from Y to X.

Moreover: $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$.

To see this (only need to prove one of the identities, by symmetry), let $x \in X$; then $f(x) = y \in Y$ says that $g(y) = x$.
Hence $(g \circ f)(x) = g(f(x)) = g(y) = x$. True $\forall\, x \in X$, so $g \circ f = \mathrm{id}_X$ as claimed.

g is called the *inverse* of f, and is denoted f^{-1}.

Summing up: every bijection $f : X \to Y$ has an inverse, $f^{-1} : Y \to X$, which is also a bijection, and

$$f \circ f^{-1} = \mathrm{id}_Y, \qquad f^{-1} \circ f = \mathrm{id}_X.$$

(Remember: in general, the composition of bijections is a bijection—corollary on p. 3 of this lecture.)

Note: $(f^{-1})^{-1} = f$ (!).

Multiplying (composing) an invertible function with its inverse gives the identity (function); cf. (for example) multiplying non-zero $x \in \mathbb{R}$ by x^{-1} to get 1.

In the other direction, if $f : A \to B$ is a bijection, and $\alpha : B \to A$ is a function such that

$$f \circ \alpha = \mathrm{id}_B \qquad \text{and} \qquad \alpha \circ f = \mathrm{id}_A$$

then $\alpha = f^{-1}$.

Example. Let $f : \mathbb{R} \to \mathbb{R}$ be the function such that $f(x) = 5x + 8$.

We know f is a bijection from $\mathbb{R}$ to $\mathbb{R}$. What is f^{-1}?

Since $(f \circ f^{-1})(x) = x$, we have

$x = (f \circ f^{-1})(x) = f(f^{-1}(x)) = 5f^{-1}(x) + 8$, so

$5f^{-1}(x) = x - 8$ and thus $f^{-1}(x) = \frac{1}{5}(x - 8)$.

Inverse of a (non-horizontal, non-vertical) straight line is a straight line.

Example. If $f : A \to A$ is a function such that $f^2 := f \circ f = \mathrm{id}_A$, then f is a bijection. Furthermore, $f^{-1} = f$. (*Exercise*: prove both claims; second claim is clear by definition of inverse of bijection after the first claim is proved.)

Permutations

A *permutation* of a set X is a bijection from X to itself.

Usually $X = \{1, 2, \ldots, n\}$ for some $n \geq 1$.

Then a permutation of X can be thought of as a *rearrangement* of the *ordered* list $1, 2, \ldots, n$.

Example. Let $X = \{1, 2, 3\}$. The function $\alpha : X \to X$ defined by $\alpha(1) = 2$, $\alpha(2) = 3$, $\alpha(3) = 1$ is a permutation. Write this as $\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Here are some others: $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = \mathrm{id}_X$,

$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \alpha^{-1}$.

Is this all permutations of $\{1, 2, 3\}$?

If $X = \{1, 2, \ldots, n\}$ then the set of all permutations of X is denoted $\mathrm{Sym}(X)$, or $\mathrm{Sym}(n)$, or S_n. It is called the *symmetric group of* X, or *the symmetric group of degree* n.

Theorem

S_n is a set of size $n!$.

Proof.

$n!$, read 'n factorial' is the product of the first n positive integers: $n \cdot (n-1) \cdots 2 \cdot 1$. By convention, $0! = 1$.

An element of S_n can be written as $\begin{pmatrix} 1 & 2 & \cdots & n \\ * & * & \cdots & * \end{pmatrix}$. Fill in the asterisks: n choices in position 1, then $n-1$ choices in position 2, then ..., then 2 choices in position $n-1$. Product of these numbers of choices $=$ total no. ways of filling in the bottom row. $\square$

Example. S_3 has size $3! = 6$. Hence we did write down all permutations of $\{1, 2, 3\}$ in the first example.

Why is S_n called a *group*?

We can 'multiply' two elements $\alpha, \beta \in S_n$ to get an element of S_n, by composing them, i.e., $\alpha \circ \beta \in S_n$ (also $\beta \circ \alpha \in S_n$; usually $\alpha \circ \beta \neq \beta \circ \alpha$). Remember: composition of bijections is a bijection.

Since function composition is associative, this 'multiplication' defined on S_n is associative: $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$, for all $\alpha, \beta, \gamma \in S_n$.

There is an identity element for the multiplication, the identity permutation id of $\{1, 2, \ldots, n\}$: $\mathrm{id} \circ \alpha = \alpha \circ \mathrm{id} = \alpha$, for all $\alpha \in S_n$.

Each element $\alpha \in S_n$ has an inverse $\alpha^{-1} \in S_n$ such that $\alpha \circ \alpha^{-1} = \alpha^{-1} \circ \alpha = \mathrm{id}$.

The fact that these properties hold for S_n make it an example of an algebraic structure called a *group*.

Example. Let $\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$.

Then $\alpha \circ \beta(1) = \alpha(2) = 3$, $\alpha \circ \beta(2) = \alpha(1) = 2$,
$\alpha \circ \beta(3) = \alpha(3) = 1$.

That is, $\alpha \circ \beta = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

Check also that $\beta \circ \alpha = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, so $\alpha \circ \beta \neq \beta \circ \alpha$;

$\alpha^2 := \alpha \circ \alpha = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, $\alpha^3 = \beta^2 = \mathrm{id}$. □

We need a more compact notation to specify permutations.

Commonly used 'in-line' notation: e.g., $(1, 2)$ denotes the permutation that swaps 1 and 2, leaving any other element of $X = \{1, 2, \ldots, n\}$ fixed. Note $(i, j) = (j, i)$.

$(2, 4, 5)(7, 8) \in S_8$ sends 2 to 4, 4 to 5, 5 to 2 (loop around), 7 to 8, 8 to 7; fixes 1, 3, 6.

Note: $(2, 4, 5)(7, 8) = (2, 4, 5) \circ (7, 8) = (7, 8) \circ (2, 4, 5) = (7, 8)(2, 4, 5)$.

Example. $(1, 2, 3) \circ (1, 2) = (1, 3)(2) = (1, 3)$;
$(1, 2) \circ (1, 2, 3) = (1)(2, 3) = (2, 3)$.

Example. $(1, 4, 6) \circ (4, 5, 7) = (4, 5, 7, 6, 1) = (1, 4, 5, 7, 6)$.

Cycles

An m-*cycle* has the form $(x_1, x_2, \ldots, x_m)$. That is, this m-cycle takes x_1 to x_2, x_2 to $x_3, \ldots, x_{m-1}$ to x_m, and then x_m to x_1.

Example. $(1,2) \circ (1,2) = \mathrm{id}$.
$(1,2,3)^3 = (1,2,3) \circ [(1,2,3) \circ (1,2,3)] = (1,2,3) \circ (1,3,2) = \mathrm{id}$.
$(1,2,3,4)^4 = \mathrm{id}$ (check). □

Two cycles are said to be *disjoint* if they share no common points. E.g., $(2,4)$ and $(3,5,7)$ are disjoint. $(1,2)$, $(3,4)$, $(9,10)$ are pairwise disjoint. $(1,2,3)$ and $(2,5,6)$ are not disjoint.

Fact: two disjoint cycles α, β commute with each other, i.e., $\alpha \circ \beta = \beta \circ \alpha$. (Observed in some previous examples.)

Theorem

Each permutation is a product of disjoint cycles, which are uniquely determined by the permutation up to order of the cycles.

Disjoint cycles commute.

Example.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 3 & 6 & 1 & 4 \end{pmatrix} = (1,2,5)(3)(4,6) = (1,2,5)(4,6).$$

Note: in the example, we omitted '$\circ$'. This is because the cycles are disjoint, hence commute, so the product is equal in either order.

The *order* of a permutation $\alpha \in S_n$ is the least positive integer r such that $\alpha^r := \alpha \circ \alpha \circ \cdots \circ \alpha$ (r times) $= \mathrm{id}$.

Example. id has order 1 (unique permutation of order 1). $(1,2)$ has order 2. So too do $(3,5)$, $(6,2)$, $(3,11)$...

Example. What is the order of $\gamma = (1,2)(7,8,9)$?

$\gamma^2 = (1,2)^2(7,8,9)^2$ [disjoint cycles commute] $= (7,8,9)^2 = (7,9,8)$,

$\gamma^3 = \gamma\gamma^2 = (1,2)(7,8,9)(7,9,8) = (7)(9)(8)(1,2) = (1,2)$,

$\gamma^4 = \gamma\gamma^3 = (1,2)(7,8,9)(1,2) = (1,2)^2(7,8,9) = (7,8,9)$,

$\gamma^5 = \gamma\gamma^4 = (1,2)(7,8,9)(7,8,9) = (1,2)(7,9,8)$,

$\gamma^6 = \gamma\gamma^5 = (1,2)(7,8,9)(1,2)(7,9,8) = (1,2)^2(7,8,9)(7,9,8)$
$= \mathrm{id}$.

Hence, γ has order 6.

Example. An m-cycle (cycle of length m) has order m.

To compute the order of a permutation $\alpha \in S_n$, first write α as a product of disjoint cycles; say these are of lengths $m_1, \ldots, m_r$. Then the order of α is the *least common multiple* of $m_1, \ldots, m_r$.

Reasoning: a cycle of length m has order m and $\therefore$ has m'th power equal to id for an integer m' iff m' is divisible by m.

Disjoint cycles commute with each other.

Thus, the least integer $r > 0$ such that $\alpha^r = \mathrm{id}$ is the least r such that each of the cycles powers to the identity: the least common multiple of the m_is.

Example. Find the order of $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 5 & 2 & 1 \end{pmatrix} \in S_6$.

Solution. $\sigma = (1,6)(2,4,5)(3)$.

Thus σ has order $\mathrm{lcm}(2,3) = 6$. □

The same method applies to the second example on p. 1: much quicker calculation of the order of $\gamma = (1,2)(7,8,9)$ as 6.

A *transposition* is a cycle of length 2, e.g., $(1,2)$, $(98,1001)$, etc.

Of course, a transposition has order 2.

Each cycle is a product of transpositions; e.g.,
$(1,2,3) = (1,2)(2,3)$, $(1,3,13,24) = (1,3)(3,13)(13,24)$.

In general, $(x_1, x_2, \ldots, x_m) = (x_1, x_2)(x_2, x_3) \cdots (x_{m-1}, x_m)$.

Theorem

Every permutation is a product of transpositions.

Proof.

Write the permutation as a product of disjoint cycles. Write each cycle as a product of transpositions. □

Parity of permutations

A permutation is called *even* if it is the product of an even number of transpositions; if it is the product of an odd number of transpositions then it is called *odd*.

A permutation cannot be both even and odd (why?).

Fact: the product of any two even permutations is even.

Fact: the product of any two odd permutations is even.

Fact: the product of an odd and an even permutation is odd.

Fact: the order of an odd permutation is even.

Fact: the order of an even permutation is even or odd.

Example. Previously we worked out all elements of S_3: $\mathrm{id}, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)$. Note $(1,2,3)^{-1} = (1,3,2)$.

$(1,2)$, $(1,3)$, $(2,3)$ are odd. id, $(1,2,3)$, $(1,3,2)$ are even.

All the odd permutations square to the even permutation id.

The product of a pair of different odd permutations (transpositions) is even (one of the 3-cycles); check.

And $(1,2,3)^2 = (1,3,2)$, even; $(1,3,2)^2 = (1,2,3)$, even; $(1,3,2)(1,2,3) = (1,2,3)(1,3,2) = \mathrm{id}$, even. Of course, multiplying any permutation by id doesn't change its parity.

And $(1,2)(1,2,3) = (2,3)$, odd; $(1,2)(1,3,2) = (1,3)$, odd; $(1,3)(1,2,3) = (1,2)$, odd; $(1,3)(1,3,2) = (2,3)$, odd; $(2,3)(1,2,3) = (1,3)$, odd; $(2,3)(1,3,2) = (1,2)$, odd. Check that each product in reverse order is odd too.

Groups

As discussed previously, S_n is (an example of) a *group*. We now define this concept in general.

Definition. Let G be a (non-empty) set on which there is defined a *binary operation*, denoted $\star$ say.That is, $\star$ is a function from $G \times G$ to G; we write the image of $(a, b) \in G \times G$ under $\star$ as $a \star b$.

Then G is called a *group* if the following all hold.

- $a \star (b \star c) = (a \star b) \star c, \ \forall\, a, b, c \in G.$ ASSOCIATIVITY
- $\exists\ e \in G$ (called *the identity of* G) such that $a \star e = e \star a = a, \ \forall\, a \in G.$ IDENTITY
- $\forall\, a \in G, \exists\ a^{-1} \in G$ (called *the inverse of* a) such that $a \star a^{-1} = a^{-1} \star a = e.$ INVERSES

The identity of a group G is often written 1_G, or just 1.

Note: it is 'the' identity: if $f \in G$ and $f \star a = a \star f = a$ for all $a \in G$, then $1 = 1 \star f$ (take $a = 1$ previously) $= f$ (because of the axiom defining 1).

Similarly, each element a of a group G has a unique inverse.

The binary operation in a group is sometimes called 'multiplication'. It is also standard to omit any special symbol for the operation and simply juxtapose elements, i.e., $ab := a \star b$.

Caution: the multiplication may not be commutative: we do not stipulate that $ab = ba$, $\forall\, a, b \in G$.

(However, an arbitrary pair of elements of a group *may* commute.)

A group that does have a commutative binary operation is called *abelian* (after the Norwegian mathematician Niels Henrik Abel, 1802–1829).

The binary operation of an abelian group might be called *addition*, its identity might be called *zero* and denoted 0.

Example. $\{0, 1\}$ under addition modulo 2 is an (abelian!) group.

Example. S_n is a group; non-abelian if $n > 2$. The set of all even permutations in S_n is a group; non-abelian if $n > 3$.

Example. The set $\mathbb{R}$ of real numbers under addition is an abelian group.

The set of non-zero real numbers under multiplication is an abelian group.

The analogous statements hold for the rationals $\mathbb{Q}$ and the complex numbers $\mathbb{C}$.

Example. The integers $\mathbb{Z}$ under addition form an abelian group. Under multiplication, $\mathbb{Z} \setminus \{0\}$ does not form a group; e.g., the multiplicative inverse of 2 is not an integer.

Example. The set of natural numbers $\mathbb{N}$ is not a group under $+$; e.g., $-1 \notin \mathbb{N}$.

Example. The set of 2×2 matrices with entries in $\mathbb{R}$ is an abelian group under matrix addition. The set of 2×2 *invertible* real matrices is a group under matrix multiplication.

Note $\det(xy) = \det(x)\det(y)$, so the product of two invertible matrices is invertible.

The identity here is the 2×2 identity matrix: 1s down the main diagonal, zeros elsewhere.

Rings

Some sets have *two* separate binary operations, satisfying separate axioms, and interacting with each other. Such a structure is called a *ring*. The premier class of examples for us is *polynomial rings* (next major topic).

Definition. Let R be a set on which there are two binary operations defined, $+$ and $\star$. Write the image of $(a, b) \in R \times R$ under $\star$ as $a \star b$, and the image of (a, b) under $+$ as $a + b$. Then R is called a *ring* if the following all hold.

- R under $+$ is an abelian group (with identity $0 = 0_R$).
- $\star$ is associative: $a \star (b \star c) = (a \star b) \star c, \ \forall\, a, b, c \in R$.
- Distributivity: $a \star (b + c) = a \star b + a \star c,$
  $(a + b) \star c = a \star c + b \star c \quad \forall\, a, b, c \in R.$

$\star$ is usually omitted in notation (we would write ab instead of $a \star b$), and is called the 'multiplication of R'.

If $\star$ is commutative, i.e., $a \star b = b \star a$ for all $a, b \in R$, then R is a *commutative ring*.

All rings of interest 'have a 1', i.e., possess an identity for $\star$, an element e of R such that $a \star e = e \star a = a$ for all $a \in R$.

Example. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all rings (commutative, with 1). Note how every non-zero element of each of the rings $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ has a multiplicative inverse. However, that is not the case in $\mathbb{Z}$.

Example. The set of all 2×2 real matrices is a ring. *Not* commutative. Has a 1 (the identity matrix).

Polynomial rings

Let R be a commutative ring (with 1). A *polynomial over R in the indeterminate x* is an expression of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n$$

where $a_0, \ldots, a_n \in R$ (the *coefficients of $p(x)$*). More compactly, $p(x) = \sum_{i=0}^{n} a_i x^i$.

If $a_n \neq 0$ then $p(x)$ *has degree n*.

Note: $p(x)$ is a 'formal expression'. It does not denote a function; x is just a symbol.

Two polynomials $f(x)$, $g(x)$ are equal, $f(x) = g(x)$, if and only if they have the same coefficients on every power of x.

We can take, e.g., R to be any of the familiar number rings: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Or, e.g., $R = \mathbb{Z}_m$, the integers $\{0, 1, \ldots, m-1\}$ under addition and multiplication modulo m: this is a commutative ring with 1 (check).

Notation: $R[x]$ is the (infinite) set of all polynomials over the ring R in the indeterminate x.

Observe $R \subset R[x]$; $r \in R$ is a *constant polynomial* in $R[x]$.

We can define an addition, denoted $+$, on $R[x]$ in 'componentwise' fashion:

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{m} b_i x^i = \sum_{i=0}^{\max(m,n)} (a_i + b_i) x^i$$

where $a_j = 0$ if $j > n$ and $b_j = 0$ if $j > m$. Note how this *extends* addition in R.

Also, we can define a multiplication on $R[x]$, as follows. Set $a_i x^i \, b_j x^j = a_i b_j x^{i+j}$ (makes sense 'formally'), then extend to all polynomials by using distributivity.

That is, $\left(\sum_{i=0}^{n} a_i x^i\right)\left(\sum_{i=0}^{m} b_i x^i\right) =$

$$a_0 b_0 + (a_1 b_0 + a_0 b_1)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \cdots$$

In this product, x^k has coefficient $\sum_{i=0}^{k} a_i b_{k-i}$.

Theorem

With addition and multiplication of polynomials as defined above, $R[x]$ is a ring. Furthermore, $R[x]$ is commutative because R is commutative, and $R[x]$ has a 1 because R has a 1.

Example. Let $p(x) = x^2 + 2x + 1$, $q(x) = x^3 + x + 2 \in \mathbb{Z}[x]$.

Then $p(x) + q(x) = x^3 + x^2 + 3x + 3$ and

$$\begin{aligned} p(x)q(x) &= (x^2 + 2x + 1)(x^3 + x + 2) \\ &= x^5 + x^3 + 2x^2 + 2x^4 + 2x^2 + 4x + x^3 + x + 2 \\ &= x^5 + 2x^4 + 2x^3 + 4x^2 + 5x + 2. \end{aligned}$$

Now consider $p(x), q(x)$ as elements of $\mathbb{Z}_3[x]$; i.e., read all coefficients modulo 3.

Then $p(x) + q(x) = x^3 + x^2$ and

$$p(x)q(x) = x^5 + 2x^4 + 2x^3 + x^2 + 2x + 2.$$

A *field* F is a commutative ring with 1 such that every element of F apart from 0_F has an inverse under the field multiplication; i.e., the non-zero elements of F form an (abelian) group under the multiplication of F.

Example. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are fields.

Example. $\mathbb{Z}$ is not a field; e.g., $2^{-1} \notin \mathbb{Z}$.

Example. $\mathbb{Z}_3$, the ring of integers modulo 3, is a field: 1 is its own inverse, as is 2: $2.2 = 4 \equiv 1$ modulo 3.

Example. $\mathbb{Z}_4$, the ring of integers modulo 4, is not a field: $2.1 \equiv 2$, $2.2 \equiv 0$, $2.3 \equiv 2$ modulo 4. So 2 has no multiplicative inverse in $\mathbb{Z}_4$ (2 does not multiply with any element to give 1 modulo 4).

In general, $\mathbb{Z}_m$ is a field if and only if m is a prime. (Why?)

We can divide one polynomial by another (non-zero) polynomial in a polynomial ring over a field.

Theorem

Let F be a field, and let $f(x), g(x)$ be non-zero elements of $F[x]$. Then there exist $q(x), r(x) \in F[x]$ such that

(i) $f(x) = g(x)q(x) + r(x)$, *and*

(ii) *the degree of $r(x)$ is strictly less than the degree of $g(x)$.*

Furthermore, the quotient $q(x)$ and remainder $r(x)$ are uniquely determined by (i) *and* (ii).

Important: compare this with the *Integer Division Theorem*:

Theorem

Let $a, b \in \mathbb{Z}$, $b > 0$. Then there exist unique $q \in \mathbb{Z}$ and $r \in \mathbb{Z}$ where $0 \leq r < b$ such that $a = bq + r$.

Proof of the polynomial division theorem is by *induction* on the degree of $f(x)$. (Induction is the next major topic after this one.)

Example. In $\mathbb{Q}[x]$,

$$x^3 + 2x^2 + 4x - 7 = (x^2 + x - 2)(x + 1) + (5x - 5).$$

(Check.) Thus, division of $f(x) = x^3 + 2x^2 + 4x - 7$ by $g(x) = x^2 + x - 2$ in $\mathbb{Q}[x]$ gives quotient $q(x) = x + 1$ and remainder $r(x) = 5x - 5$.

Note: $5x - 5$ has degree 1, less than the degree 2 of $x^2 + x - 2$.

How might we find $q(x)$ and $r(x)$ in practice, given any $f(x)$ and non-zero $g(x)$?

We discuss some consequences of the polynomial division theorem: recall that if $f(x)$ and $g(x)$ are non-zero polynomials over a field F, then $\exists$ unique $q(x), r(x) \in F[x]$ such that $f(x) = g(x)q(x) + r(x)$ and $\deg(r(x)) < \deg(g(x))$.

If $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$ and $c \in F$ then we write $f(c)$ for the element $\sum_{i=0}^{n} a_i c^i$ of F.

An element c of F is a *root* of $f(x) \in F[x]$ if $f(c) = 0$.

Theorem

Let $c \in F$ and $f(x) \in F[x]$. Then $f(c)$ is the remainder after division of $f(x)$ by $x - c$.

Proof.

By the PDT, $f(x) = (x - c)q(x) + r(x)$ where $\deg(r(x)) < \deg(x - c) = 1$. Hence $r(x)$ is a constant. Since $f(c) = 0 + r(c)$, we have $r(x) = r(c) = f(c)$. □

If a polynomial $p(x)$ exactly divides another polynomial $q(x)$, i.e., it has zero remainder after division, then $p(x)$ is a *factor* of $q(x)$.

Corollary

$c \in F$ is a root of $f(x) \in F[x]$ if and only if $x - c$ is a factor of $f(x)$.

Proof.

By definition, $x - c$ is a factor iff the remainder after division of $f(x)$ by $x - c$ is 0; by the theorem, this remainder is $f(c)$. □

Example. Let $f(x) = x^3 + x^2 - 5x + 3 \in \mathbb{R}[x]$. Then $f(1) = 1 + 1 - 5 + 3 = 0$. Thus $x - 1$ is a factor of $f(x)$. Indeed, $f(x) = (x - 1)(x^2 + 2x - 3)$. Since $x^2 + 2x - 3 = (x + 3)(x - 1)$, by the corollary $f(-3) = 0$ [check].

A gcd (*greatest common divisor*) of non-zero $f(x), g(x) \in F[x]$ is $d(x) \in F[x]$ such that $d(x)$ is a factor of $f(x)$ and $g(x)$, and if $h(x)$ is a factor of $f(x)$ and $g(x)$ then $h(x)$ is a factor of $d(x)$.

Note: a gcd is determined only up to multiplication by non-zero constants; i.e., if $d(x)$ is a gcd then so is $ad(x)$, $\forall\, a \in F \setminus \{0\}$. Make the gcd unique by insisting it is *monic* (leading coefficient 1).

Example. Let $F = \mathbb{Z}_3 = \{0, 1, 2\}$, $f(x) = x^3 + 2x^2 + 2$, $g(x) = x^2 + 2x + 1$.

Divide f by g: $f(x) = g(x)x + (2x + 2)$; remember, modulo 3.

Divide the previous divisor by the previous (non-zero) remainder: $g(x) = (2x + 2)(2x + 2)$. Zero remainder here.

$x + 1$ is a common divisor of f and g. Any other common divisor must divide $2(x + 1) = f(x) - g(x)x$. Thus $x + 1$ is a gcd.

$f(x), g(x) \in F[x]$ are *coprime* if they have 1 as a gcd.

Example. Let $F = \mathbb{Z}_3$, $f(x) = x^5 + 1$, $g(x) = x^2 + 1$.

$f(x) = g(x)(x^3 + 2x) + (x + 1)$.

Hence any common divisor $d(x)$ of $f(x)$ and $g(x)$ must divide $x + 1$.

However, $g(x) = (x + 1)(x + 2) + 2$; so $d(x)$ divides 2.

Therefore this $f(x)$ and $g(x)$ are coprime.

Repeating the division theorem: computing gcds

Computing a gcd of two non-zero polynomials (over a field) is analogous to computing the gcd of a pair of positive integers. The method is known as the *Euclidean algorithm*. It avoids factorization of the polynomials in question, using *polynomial long division*.

The algorithm is recursive. At each stage, we have a pair of non-zero polynomials $a(x), b(x)$ (in $\mathbb{Q}[x]$, say), and we perform polynomial long division to divide $a(x)$ by $b(x)$.

That is, we compute polynomials $q(x), r(x)$ such that $a(x) = b(x)q(x) + r(x)$ where $\deg(r(x))$ is less than $\deg(b(x))$: PDT.

If $r(x) = 0$ then we $\textsc{Stop}$, and output the last non-zero remainder in the recursion.

If $r(x) \neq 0$ then we proceed to the next stage: divide this remainder $r(x)$ into the previous divisor $b(x)$.

The process must eventually terminate, since the degrees are decreasing as we move from stage to stage.

However, why is the output correct?

Exercise. prove that the output—the very last non-zero remainder—really is a gcd of the input pair of polynomials. Compare with the Euclidean algorithm in $\mathbb{Z}$.

Example. Find a gcd of $x^3 + 2x^2 + 4x - 7$ and $x^2 + x - 2$ in $\mathbb{Q}[x]$.

Solution.

$$\begin{array}{rl} & x + 1 \\ \hline x^2 + x - 2 & \big| \; x^3 + 2x^2 + 4x - 7 \\ & x^3 + x^2 - 2x \\ \hline & x^2 + 6x - 7 \\ & x^2 + x - 2 \\ \hline & 5x - 5 \end{array}$$

Thus $x^3 + 2x^2 + 4x - 7 = (x^2 + x - 2)(x + 1) + (5x - 5)$.

Non-zero remainder, so must continue to next stage.

Divide previous divisor $x^2 + x - 2$ by previous remainder $5x - 5$.

$$\begin{array}{rl} & \frac{1}{5}x + \frac{2}{5} \\ \hline 5x - 5 & \big| \; x^2 + x - 2 \\ & x^2 - x \\ \hline & 2x - 2 \\ & 2x - 2 \\ \hline & 0 \end{array}$$

Remainder is 0: stop. Last non-zero remainder, i.e., $5x - 5$, is a gcd of the input pair.

Exercise. Check independently that $x - 1$ is a gcd of $x^3 + 2x^2 + 4x - 7$ and $x^2 + x - 2$. Can at least check quickly that $x - 1$ is a common divisor!

Let F be a field and let $f(x) \in F[x]$ be a non-constant polynomial. $f(x)$ is *irreducible* (over F) if, for any $a(x), b(x) \in F[x]$ such that $f(x) = a(x)b(x)$, either $a(x)$ or $b(x)$ is a constant.

Every non-constant polynomial can be factorized into a product of irreducible polynomials.

Moreover the factorization is *unique*, up to unimportant changes such as altering the order of factors, or multiplying a factor by some non-zero $a \in F$ (and multiplying another factor by a^{-1}).

Example. Every linear polynomial $x - a \in F[x]$ where $a \in F$ is irreducible.

Example. $x^2 + 1$ is irreducible over $\mathbb{R}$. However, over $\mathbb{C}$, it is *not* irreducible (it is *reducible*): $x^2 + 1 = (x - \mathrm{i})(x + \mathrm{i})$ where $\mathrm{i} = \sqrt{-1}$ as usual.

The following is the *Fundamental Theorem of Algebra*.

Theorem

Each non-constant polynomial in $\mathbb{C}[x]$ has a root in $\mathbb{C}$.

Corollary

Each polynomial of degree $n > 0$ over $\mathbb{C}$ factorizes into the product of n linear (hence irreducible) factors over $\mathbb{C}$ (counting repeats).

We use calculus to prove a specialisation of the FTA: every real polynomial of odd degree has at least one real root. (This fails for even degree; e.g., consider $f(x) = x^2 + 1$.)

Theorem

Let $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{R}[x]$ where $a_n \neq 0$ and n is odd. Then $f(r) = 0$ for some $r \in \mathbb{R}$.

Proof.

We may assume that a_n is positive (if not, multiply $f(x)$ by -1 and note that $f(r) = 0$ if and only if $-f(r) = 0$).

For large enough positive b, $f(b) \approx a_n b^n$ is positive.

On the other hand, there will be negative c such that $f(c) \approx a_n c^n < 0$. (We are using n odd here: x^n is negative if x is negative.)

So: $f(b) > 0 > f(c)$: $f(x)$ takes on positive and negative values after substituting for x.

Since $f(x)$ as a function of real numbers is continuous, the Intermediate Value Theorem implies that there must be some r (real) such that $f(r) = 0$. □

The specialisation of the Fundamental Theorem of Algebra proved above is also a consequence of the following (check this claim).

Theorem

Let $f(x) \in \mathbb{R}[x]$, where the degree of $f(x)$ is greater than 2. Then $f(x)$ is reducible over $\mathbb{R}$.

That is, every polynomial of degree greater than 2 in $\mathbb{R}[x]$ factorizes as a product of irreducibles in $\mathbb{R}[x]$, each of degree 1 (linear) or two (quadratic).

Proof. By the FTA (in general), $f(x)$ has a root $\alpha \in \mathbb{C}$.

Denote complex conjugation by overline: $\overline{a+b\mathrm{i}} = a - b\mathrm{i}$, $a, b \in \mathbb{R}$. Then $f(\alpha) = 0$ implies that $\overline{f(\alpha)} = \overline{0} = 0$, and then $f(\overline{\alpha}) = \overline{f(\alpha)} = 0$ ($f(\overline{\alpha}) = \overline{f(\alpha)}$ because conjugation is additive and multiplicative: $\overline{\mu+\nu} = \overline{\mu} + \overline{\nu}$ and $\overline{\mu\nu} = \overline{\mu}\,\overline{\nu}$).

If $\alpha = \overline{\alpha}$ then α is real. That is, $x - \alpha \in \mathbb{R}[x]$ is a factor of $f(x)$ (by our root theorem: $c \in F$ is a root of $p(x) \in F[x]$ if and only if $x - c$ is a factor of $p(x)$). Also, $x - \alpha$ has degree 1, less than $\deg(f(x))$, which is greater than 2 by hypothesis. Done.

Otherwise, $x - \alpha$ and $x - \overline{\alpha}$ are both factors of $f(x)$ (by our root theorem); hence their product is also a factor ($x - \alpha$ and $x - \overline{\alpha}$ are different and thus coprime).

Thus $g(x) := (x-\alpha)(x-\overline{\alpha}) = x^2 - (\alpha+\overline{\alpha})x + \alpha\overline{\alpha}$ is a factor of $f(x)$. Now note that $\alpha+\overline{\alpha}$ and $\alpha\overline{\alpha}$ are both real:

$$(a+b\mathrm{i}) + (a-b\mathrm{i}) = 2a \in \mathbb{R},$$
$$(a+b\mathrm{i})(a-b\mathrm{i}) = a^2 + b^2 \in \mathbb{R};$$

remember that $\mathrm{i}^2 = -1$.

Hence $g(x) \in \mathbb{R}[x]$ of degree $2 < \deg(f(x))$ is a factor of $f(x)$. This completes the proof. $\square$

The principle of mathematical induction

A powerful proof technique, commonly used to prove statements about (positive) *integers*.

Example. The sum of the first n odd positive integers is n^2.

Example. The sum of the first n positive integers is $n(n+1)/2$.

Example. For $n \geq 3$, the sum of the angles (in radians) of an n-sided polygon is $(n-2)\pi$.

Each of the above examples is a statement $P(n)$ that is made for an infinite set of positive integers n. ($P(n)$ is a predicate!)

Example. Let $P(n)$ be 'the sum of the first n odd positive integers is n^2'. Using sigma notation we can rewrite this as $P(n) : \sum_{k=1}^{n}(2k-1) = n^2$.

Then $P(1) : (2.1 - 1) = 1$ which is T. $P(2) : 1 + 3 = 2^2$, T. $P(3) : 1 + 3 + 5 = 3^2$, T. $P(4) : 1 + 3 + 5 + 7 = 4^2$, T...

Example. Let $P(n)$ be 'the sum of the angles of an n-sided polygon is $(n-2)\pi$', for $n \geq 3$.

Then $P(3)$: 'the sum of the angles of a triangle is $(3-2)\pi = \pi$, which is T. $P(4)$: 'the sum of the angles of a quadrilateral is $(4-2)\pi = 2\pi$, T...

Exercise. Check that $P(n) : \sum_{k=1}^{n} k = n(n+1)/2$ is T for five or six random choices of n.

Principle of mathematical induction. *Let $P(n)$ be a statement about each positive integer n. If we prove that*

1. *$P(1)$ is true, and*
2. *$P(k)$ true implies that $P(k+1)$ is true, for all positive integers k,*

then $P(n)$ is true for all positive integers n.

Reason as follows: given $P(1) \equiv \mathsf{T}$ by 1. of the Principle. Since $P(1) \equiv \mathsf{T}$, so $P(2) = P(1+1) \equiv \mathsf{T}$ by 2. of the Principle. Since $P(2) \equiv \mathsf{T}$, so $P(3) = P(2+1) \equiv \mathsf{T}$ by 2. of the Principle,...

...a chain of dominoes falls over!

N.B. Note how proving 1. is vital; without 1., the domino effect cannot start.

Example. Prove $P(n) : \sum_{i=1}^{n}(2i-1) = n^2$.

Solution. We previously observed that $P(1)$ is true (and $P(2), P(3), P(4)$ too).

Assume that $P(k)$ is true, for some positive integer k (the *inductive hypothesis*).
That is, we **assume** that $1 + 3 + \cdots + (2k-1) = k^2$. $(*)$

Add the next odd positive integer after $2k-1$ to both sides of $(*)$:
$1 + 3 + \cdots + (2k-1) + (2k+1) = k^2 + (2k+1)$.

Now $k^2 + 2k + 1 = (k+1)^2$. Thus, the inductive hypothesis implies that $1 + 3 + \cdots + (2k-1) + (2k+1) = (k+1)^2$; which is exactly $P(k+1)$.

We have now fulfilled both parts of the POMI, hence proving that this statement $P(n)$ is true for all positive integers n.

Usually just say ‘proof by induction’.

Example. Define a sequence $\{a_n\}$ of integers as follows: $a_1 = 1$, and if $n \geq 2$ then $a_n = a_{n-1} + n$. Use induction to prove that $P(n) : a_n = n(n+1)/2$ for all $n \geq 1$.

Solution.

Base step. $n(n+1)/2 = 1.2/2 = 1$ for $n = 1$. Also $a_1 = 1$. Thus $P(1)$ is true.

Inductive hypothesis. Assume that $P(k)$ is true, i.e., assume that $a_k = k(k+1)/2$. We have to use this hypothesis to prove that $P(k+1)$ is true, i.e.,

$$a_{k+1} = (k+1)(k+1+1)/2 = (k+1)(k+2)/2. \qquad (\dagger)$$

Substituting $k+1$ for n in the definition $a_n = a_{n-1} + n$ gives $a_{k+1} = a_k + (k+1)$.

Thus $a_{k+1} = a_k + (k+1) = k(k+1)/2 + (k+1)$, using the inductive hypothesis.

Now a bit of algebra: $a_{k+1} = (k+1)(\frac{k}{2} + 1) = (k+1)\frac{k+2}{2}$, which is $(\dagger)$, exactly what we want.

This proves the formula by induction.

Example. Find and prove a formula in terms of n for $s_n := 1^2 + 2^2 + \cdots + n^2$, the sum of the squares of the first n positive integers.

Solution. Try some values of n, try to spot a pattern. $n = 1$: $1^2 = \underline{1}$. $n = 2$: $1^2 + 2^2 = \underline{5}$. $n = 3$: $5 + 3^2 = \underline{14}$. $n = 4$: $14 + 4^2 = \underline{30}$.

Looking at the sequence $s_1 = 1, s_2 = 5, s_3 = 14, s_4 = 30, \ldots$, not easy to guess a formula in terms of n for s_n. Ingenuity required. Note that

$$\begin{array}{rcl} \left(\sum_{i=1}^{n}(i+1)^3\right) - \left(\sum_{i=1}^{n} i^3\right) & = & (2^3 + \cdots + n^3 + (n+1)^3) \\ & & -(1^3 + 2^3 + \cdots + n^3) \\ & = & (n+1)^3 - 1 = n^3 + 3n^2 + 3n. \end{array}$$

Also, the left hand side above is

$$\textstyle\sum_i \big((i+1)^3 - i^3\big) = \sum_i (i^3 + 3i^2 + 3i + 1 - i^3) = 3s_n + 3\sum_{i=1}^{n} i + n.$$

Solution (continued). Equating, we get

$$3s_n + 3\sum_{i=1}^{n} i + n = n^3 + 3n^2 + 3n.$$

Solving yields $s_n = \frac{1}{3}(n^3 + 3n^2 + 2n) - \sum_{i=1}^{n} i$.

Recall from an earlier example that $\sum_{i=1}^{n} i = n(n+1)/2$.

Plugging this into the previous line, we finally get (check!!)

$$s_n = \tfrac{1}{6}(2n^3 + 3n^2 + n) = \tfrac{n(n+1)(2n+1)}{6}.$$

(This agrees with the values of s_1, s_2, s_3, s_4 computed at the beginning; again, check). We have actually proved the required formula! However, we now (re-)prove that $\boxed{s_n = \frac{n(n+1)(2n+1)}{6}}$ independently, by induction.

Solution (continued). The base case has been established, so we make the inductive hypothesis that the box for $n = k$ is true, i.e., we assume that

$$s_k = \tfrac{1}{6}k(k+1)(2k+1)$$

for some $k \geq 1$. Then

$$\begin{aligned} s_{k+1} = s_k + (k+1)^2 &= \tfrac{1}{6}k(k+1)(2k+1) + \tfrac{1}{6}(6k^2 + 12k + 6) \\ &= \tfrac{1}{6}(2k^3 + 3k^2 + k + 6k^2 + 12k + 6) \\ &= \tfrac{1}{6}(2k^3 + 9k^2 + 13k + 6). \end{aligned}$$

Now we have to factorize the cubic. We check that -1 is a root, so $(k+1)$ is a factor. Dividing the cubic by $k+1$, we get

$$\begin{aligned} 2k^3 + 9k^2 + 13k + 6 &= (k+1)(2k^2 + 7k + 6) \\ &= (k+1)(k+2)(2k+3). \end{aligned}$$

Solution (continued).

Assuming the inductive hypothesis, and using basic algebra, we have shown that

$$s_{k+1} = \tfrac{1}{6}(k+1)(k+2)(2k+3).$$

Also $\boxed{s_n = \frac{n(n+1)(2n+1)}{6}}$ for $n = k+1$ reads

$$\begin{aligned} s_{k+1} &= \tfrac{1}{6}(k+1)(k+1+1)(2(k+1)+1) \\ &= \tfrac{1}{6}(k+1)(k+2)(2k+3). \end{aligned}$$

This completes the proof that $s_n = \frac{n(n+1)(2n+1)}{6}$ for all $n \geq 1$, by induction.

Sometimes we start with a base case different from 1.

Example. For which positive integers n is $n! > 3^n$? Prove your answer by induction.

Solution. We check $1! = 1 < 3 = 3^1$, $2! = 2 < 9 = 3^2$, $3! = 6 < 27 = 3^3$, $\ldots$ $6! = 720 < 729 = 3^6$, $7! = 5040 > 2187 = 3^7$.

Thus, we surmise that $n! > 3^n$ for all $n \geq 7$.

We prove this by induction, the base case having already been established.

Solution (continued).

Inductive hypothesis: suppose that $k! > 3^k$ for some $k \geq 7$.

Then $(k+1)! = (k+1)k! > (k+1)3^k$, by definition of factorial and using the inductive hypothesis.

Now $k \geq 7$, so certainly $k + 1 > 3$. Hence

$$(k+1)! > (k+1)3^k > 3.3^k = 3^{k+1}.$$

We have obtained the inequality $n! > 3^n$ for $n = k + 1$, proving that $n! > 3^n$ for all $n \geq 7$, by induction.

Probability

A *sample space* is the set of all possible outcomes of a random process (experiment).

An *event* is a subset of a sample space.

Example. Consider the experiment of rolling a (6-sided) die once.
Outcome $=$ number on uppermost face.
So sample space $= \{1, 2, 3, 4, 5, 6\}$.

The event of rolling a number greater than $2 = \{3, 4, 5, 6\}$.

The event of rolling a $5 = \{5\}$.

The event of rolling an odd number $= \{1, 3, 5\}$.

Example. Consider selecting a card randomly from a standard 52-card deck.

The event of picking a $\heartsuit = \{\mathrm{A}\heartsuit, 2\heartsuit, \ldots, \mathrm{J}\heartsuit, \mathrm{Q}\heartsuit, \mathrm{K}\heartsuit\}$; size 13.

The event of picking a numbered $\clubsuit = \{\mathrm{A}\clubsuit, 2\clubsuit, \ldots, 10\clubsuit\}$; size 10.

The event of picking a $\diamondsuit$ face card $= \{\mathrm{J}\diamondsuit, \mathrm{Q}\diamondsuit, \mathrm{K}\diamondsuit\}$; size 3. $\square$

Empirical probability is determined by observation of an experiment; it is a relative frequency.

If E is an event in an experiment, then empirical probability $P(E)$ of E occurring is

(no. times E occurs)/(no. times experiment is performed).

Example. A coin is tossed 100 times. It comes up Heads 44 times. For this experiment, $P(\text{Heads}) = 0.44$.

Cf. theoretical probability 0.5 of a fair coin coming up heads on a single toss. □

Law of large numbers: the relative frequency (empirical probability) of an event approaches the theoretical probability as the number of times an experiment is performed $\to \infty$.

So, if we perform a coin toss 1000, 10^4, $10^5, \ldots$ times, we expect the relative frequency of Heads and Tails each to get closer and closer to $\frac{1}{2}$.

Theoretical probability. Suppose that each outcome in a (finite) sample space S is equally likely to occur (e.g., H or T in toss of a fair coin). If E is an event in S, then

$$P(E) = \frac{|E|}{|S|}$$

where we write $|X|$ for the size of a finite set X.

Example. An (unbiased, cubic) die is thrown. Find the probability of throwing (i) a 5; (ii) an even number; (iii) a number greater than 2; (iv) a 7; (v) a number less than 7.

Solution. There are six possible outcomes, each equally likely.

(i) Throwing a five can happen in just one way. Hence probability here is $\frac{1}{6}$.

Solution (continued).

(ii) An even number can be thrown in three different ways $(2, 4, 6)$, so probability here is $\frac{3}{6} = \frac{1}{2}$.

(iii) Event has size 4: $\{3, 4, 5, 6\}$, so probability here is $\frac{4}{6} = \frac{2}{3}$.

(iv) This event is empty: it has size 0, so probability is $\frac{0}{6} = 0$.

(v) This event is certain: it has size 6, so probability is $\frac{6}{6} = 1$. $\square$

Note:

- $P(\text{impossible event}) = 0$,
- $P(\text{certain event}) = 1$,
- For any event E, $0 \leq P(E) \leq 1$,
- $P(E) + P(S \setminus E) = 1$.

E and $S \setminus E$ are *complementary* events. e.g., in the above die throw example, (probability of throwing a 2 or less) $= (1 -$ the probability of throwing a 3 or more) $= 1 - \frac{2}{3} = \frac{1}{3}$.

Example. We select a card randomly from a standard 52-card deck. Find the probability that the selected card is (i) a 5; (ii) not a 5; (iii) a $\diamondsuit$; (iv) a J, Q, or K; (v) greater than 6 and less than 9.

Solution.

(i) There are four 5s (one in each suit), so $P(5) = \frac{4}{52} = \frac{1}{13}$.

(ii) Probability not a $5 = 1 - P(5) = \frac{12}{13}$.

(iii) There are exactly 13 $\diamondsuit$, so $P(\diamondsuit) = \frac{13}{52} = \frac{1}{4}$.

(iv) There are 3 face cards in each of the 4 suits, so 12 face cards in total. Hence $P(\text{J or Q or K}) = \frac{12}{52} = \frac{3}{13}$.

Solution (continued).

(v) If greater than 6 but less than 9, the card must be a 7 or 8.

There are four 7s and four 8s, so this event has size 8.

Hence $P(> 6 \text{ and } < 9) = \frac{8}{52} = \frac{2}{13}$.

Odds

Odds is a ratio of probabilities.

Odds *against* an event E is defined to be

$$P(E \text{ fails to occur}) \big/ P(E \text{ occurs});$$

i.e., $P(\text{failure}) \big/ P(\text{success})$.

Example. The odds against throwing a 4 on a single throw of a die are $\frac{5}{6} \big/ \frac{1}{6} = 5/1$, i.e., 'five-to-one' or $5:1$.

Equivalently, the odds *in favor* of throwing a 4 are 1 to 5. □

Conversely, probabilities may be inferred from odds.

Example. The odds *against* Adam winning the next round of poker are $9:2$. What is the probability that Adam wins the next round of poker?

Solution. These are quite long odds, so the probability will be reasonably small.

Remember odds $= \frac{\ell}{w}$ where $\ell = P(\text{A loses})$ and $w = P(\text{A wins})$.

Note $\ell + w = 1$.

We have $\frac{9}{2} = \frac{\ell}{w} = \frac{1-w}{w}$. Cross multiplying:

$$9w = 2(1-w) = 2 - 2w,$$

so $11w = 2$. Therefore $P(\text{Adam wins}) = w = 2/11$.

Expected value

Or *expectation*—used in decision-making; indicates expected net result of an experiment over the long term.

Example. Bert rolls a die.
Alf will give Bert one euro if Bert rolls an even;
otherwise, Bert must give Alf one euro.

Who is better off in the long run?

Expected gain or loss for Alf $= P(\text{Alf wins}).(\text{amount Alf wins})$
$+P(\text{Alf loses}).(-\text{amount Alf loses}) = \frac{1}{2}.1 + \frac{1}{2}.-1 = 0.$

That is, Alf (and Bert) expects to break even: fair game.

If an experiment comprises n events, with respective probabilities $p_1, \ldots, p_n$ and outcomes $a_1, \ldots, a_n$ (could be win/loss, i.e., positive or negative), then expectation of the experiment $= \sum_{i=1}^{n} p_i a_i$.

Example. Find the expectation for one throw of an unbiased die.

Solution.
$\frac{1}{6}.1 + \frac{1}{6}.2 + \frac{1}{6}.3 + \frac{1}{6}.4 + \frac{1}{6}.5 + \frac{1}{6}.6 = \frac{1+2+3+4+5+6}{6} = 6.7/2.6 = 3.5.$

Example. In a multiple choice quiz, there are five possible answers to each question. Marking scheme: 2 marks for a correct answer, $\frac{1}{2}$ mark deducted for an incorrect answer; 0 marks for blank.

(i) If Alice doesn't know an answer, should she guess?

(ii) If Alice can eliminate one of the choices as answer to a question, should she guess for that question?

Solution.

(i) $P(\text{correct guess}) = \frac{1}{5}$. $P(\text{incorrect guess}) = \frac{4}{5}$.

Hence expectation is $\frac{1}{5}.2 + \frac{4}{5}(-\frac{1}{2}) = \frac{2}{5} - \frac{4}{10} = 0$.

In the long run, expect no disadvantage (nor advantage!) from guessing.

(ii) Now $P(\text{correct guess}) = \frac{1}{4}$. $P(\text{incorrect guess}) = \frac{3}{4}$.

Expectation is $\frac{1}{4}.2 - \frac{3}{4} \cdot \frac{1}{2} = \frac{1}{2} - \frac{3}{8} = \frac{1}{8} > 0$.

Yes, she should guess.

Example. One thousand raffle tickets are sold at one euro per ticket. First prize is 500 euro and there are two consolation prizes of 100 euro each.

(i) Irene buys one ticket. Expectation?

(ii) Expectation if Irene buys five tickets?

Solution.

(i) $E = p_1a_1 + p_2a_2 + p_3a_3 = \frac{1}{1000}(500 - 1)$ [wins 1st prize] $+\frac{2}{1000}(100 - 1)$ [wins consolation prize] $+\frac{997}{1000}(-1)$ [wins nothing].

So $E = \frac{499+198-997}{1000} = -300/1000 = -0.3$, i.e., a 30 cent loss.

(ii) Expect a loss of $5 \times 0.3 = 1.50$ euro. □

What would be a fair ticket price in the above example?

Fair price = expected value + cost to play.

This is $-0.3 + 1 = 0.7$, i.e., fair price is 70 cent per ticket.

Indeed, at this price, Irene's expectation for buying a single ticket is

$$\frac{1}{1000}(499.3) + \frac{2}{1000}(99.3) + \frac{997}{1000}(-0.7) = \frac{499.3+198.6-697.9}{1000} = 0.$$

Example. One of the games at a funfair is spinning the pointer on a wheel with the following prizes: 10 euro in one quarter of the wheel, 5 euro in another quarter of the wheel, 2 euro in a third quarter of the wheel, 20 euro in one eighth of the wheel, 5 euro in the remaining eighth of the wheel. A ticket for one spin of the pointer is 8 euro.

Is this a fair game?

Solution.

You can guess that it is not a fair game! i.e., that the expectation of a single spin is negative. We do the calculation:

Result	Probability	Win/loss
2	$\frac{1}{4}$	$2-8=-6$
5	$\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$	$5-8=-3$
10	$\frac{1}{4}$	$10-8=2$
20	$\frac{1}{8}$	$20-8=12$

(In the second row, there are two ways of spinning 5: one in a quarter of the wheel, the other in an eighth of it.)
Expectation is therefore

$$\tfrac{1}{4}(-6)+\tfrac{3}{8}(-3)+\tfrac{1}{4}(2)+\tfrac{1}{8}(12)=\tfrac{-12-9+4+12}{8}=-\tfrac{5}{8}.$$

As predicted, expect to lose playing this game (62.5 cent per spin).

Or & And problems

'Or' problems ask for the probability of success for at least one event from a list.

For two events A, B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Also write $P(A \text{ or } B)$ as $P(A \cup B)$, and $P(A \text{ and } B)$ as $P(A \cap B)$.

('A and B' means 'A and B occurring together at the same time'.)

Note that we have to subtract $P(A \cap B)$ because it is counted twice in adding $P(A)$ to $P(B)$.

The 'or' rule generalizes to more than two events, e.g.,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad +P(A \cap B \cap C). \end{aligned}$$

The general result is known as the *inclusion/exclusion principle*. It can be proved by induction on the number of events.

Example. Chips labeled $1, 2, \ldots, 10$ are placed in a hat. Then a chip is randomly selected. Calculate the probability that the selected chip is even or greater than 6.

Solution. $P(E) = \frac{5}{10} = \frac{1}{2}$. $G = \{7, 8, 9, 10\}$ so $P(G) = \frac{4}{10} = \frac{2}{5}$. $G \cap E = \{8, 10\}$ so $P(G \cap E) = \frac{2}{10} = \frac{1}{5}$. Thus

$$P(E \cup G) = P(E) + P(G) - P(E \cap G) = \tfrac{1}{2} + \tfrac{2}{5} - \tfrac{1}{5} = 0.7.$$

Example. A single card is drawn from a standard 52-card deck. Calculate $P(A \cup B)$ for (i) $A =$ ace, $B =$ J; (ii) $A =$ ace, $B = \heartsuit$; (iii) $A =$ red, $B =$ black; (iv) $A =$ face card, $B =$ red.

Solution.

(i) Here $A \cap B = \emptyset$, so that $P(A \cup B) = P(A) + P(B) = \frac{4}{52} + \frac{4}{52} = \frac{2}{13} \approx 0.1538$.

(ii) Here $P(A \cap B) = \frac{1}{52}$ (only one ace of hearts). Thus $P(A \cup B) = P(A) + P(B) - \frac{1}{52} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} \approx 0.3077$.

(iii) Certain! so probability $= 1$. Or: $P(\text{r}) = \frac{1}{2} = P(\text{b})$ and $P(\text{r \& b}) = 0$. Thus $P(A \cup B) = P(A) + P(B) - 0 = \frac{1}{2} + \frac{1}{2} = 1$.

(iv) 3 face cards in each of the 4 suits. Hence $P(A) = \frac{12}{52}$. $P(\text{r}) = \frac{1}{2}$. $P(\text{r and face}) = \frac{3+3}{52}$ (3 in $\heartsuit$, 3 in $\diamondsuit$) $= \frac{6}{52}$. Thus $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{12+26-6}{52} = \frac{32}{52} \approx 0.6154$.

'And' problems

These involve compound event probabilities.

Let A, B be events. Then $P(A \text{ and then } B) = P(A) \cdot P(B)$.

That is, we consider A occurring first, and then B.

Example. Suppose that a bag contains one red token, one green token, one blue token.

Two tokens are selected from the bag; the first is replaced before the second is drawn.

The sample space (can be written down from a tree diagram) is $\{rr, rb, rg, br, bb, bg, gr, gb, gg\}$; size 9.

$P(r \text{ and then } b) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$. $P(r, b \text{ in any order}) = \frac{2}{9}$ $(rb + br)$.

Example. Two cards are selected randomly from a standard 52-card deck. Find the probability that two queens are drawn (i) with replacement; (ii) without replacement.

Solution. (i) The probability of a Q on each draw is $\frac{4}{52} = \frac{1}{13}$. Hence $P(\text{Q and then 2nd Q}) = \frac{1}{13^2} = \frac{1}{169}$.

(ii) Here, the probability on the second draw changes: without replacement of the first drawn card it is $\frac{3}{51} = \frac{1}{17}$. Hence the probability of the compound event is $\frac{1}{13} \cdot \frac{1}{17} = \frac{1}{221}$.

As previous example indicates, for a given series of events we need to determine whether they are *dependent* or *independent*.

e.g., in part (i), the two events are independent: probability of the second event did not rely on the outcome of the first event.
In part (ii), the events are dependent: not replacing the first card drawn changed the sample space going into the second event.

Counting techniques

Example. A keypad uses passwords of length 4: two letters followed by two decimal digits. How many passwords are possible?

Solution. e.g., AB13, XY45, etc. etc. are all valid passwords. Think about how we would make up such a password $\alpha\beta\gamma\delta$. Begin by choosing the first symbol, α: can be done in 26 ways.

For each fixed α, we are counting the number of possible $\beta\gamma\delta$. Same (similar) problem, shorter length. There are 26 choices for β.

Now, for each of the $26^2 = 676$ possible $\alpha\beta$, we choose a γ: can be done in 10 ways.

Finally, for each of the $676 \times 10 = 6760$ possible $\alpha\beta\gamma$, we choose a δ: can be done in 10 ways. Hence number of passwords is 67600.

Tree diagram?!

The above example illustrates a general counting principle (which translates to 'and' problems in probability—and vice versa):

Multiplication principle: *if P is a process consisting of n (independent) stages, where stage i can be done in d_i ways, then the total number of ways of carrying out P is $d_1 d_2 \cdots d_n$.*

In the above example, $n = 4$, $d_1 = d_2 = 26$, $d_3 = d_4 = 10$.

One direct application of the MP is familiar: the number of permutations of an n-element set is $n! = n(n-1)(n-2)\cdots 2 \cdot 1$: pick the first element in the permutation (n ways), pick the second element ($n-1$ ways), ..., pick the kth element ($n-k+1$ ways)...

Define $\binom{n}{k}$ (read 'n choose k') as the number of ways of choosing a k-element subset from an n-element set.

For example, let $S = \{1, 2, 3, 4\}$; the 2-element subsets of S are $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{3, 4\}$. Hence $\binom{4}{2} = 6$.

Theorem $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. We count the no. permutations of $S = \{1, 2, \ldots, n\}$ as follows.

Choose a k-element subset T of S: can be done in $\binom{n}{k}$ ways.

Permute the elements of T, to get the first k entries in a permutation of S: can be done in $k!$ ways.

Permute the elements of $S \setminus T$, to get the remaining $n - k$ entries in a permutation of S: can be done in $(n - k)!$ ways.

Proof (continued).

MP: no. permutations of S is $\binom{n}{k} \times k! \times (n-k)!$. This number is $n!$. Rearranging $n! = \binom{n}{k} k!(n-k)!$ to solve for $\binom{n}{k}$ proves the theorem. □

The positive integer $\binom{n}{k}$ is known as a *binomial coefficient*. The following 'binomial theorem' (for expanding a power of a binomial, i.e., two-term expression) indicates a reason for the name.

Theorem $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

Proof. Exercise...

At least check that the theorem is correct for $n = 1, 2, 3, 4, \ldots$

Corollary $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Proof. Take $a = b = 1$ in the binomial theorem. □

Corollary $\binom{n}{k} = \binom{n}{n-k}$.

Proof. Equate the expressions for $(a+b)^n$ and $(b+a)^n$ from the binomial theorem. □

Now we apply counting techniques to probability problems.

Example. In a previous example, calculated that the probability of drawing two Qs from a standard deck without replacement is $\frac{1}{221}$.

Can also calculate this using binomial coefficients.

Total number of two card draws is $\binom{52}{2} = \frac{52!}{2!50!} = 26 \times 51 = 1326$.

There are 4 Qs in the deck. Two Qs may be chosen in $\binom{4}{2} = 6$ ways.

Hence, the probability that we choose two Qs if we draw two cards from the deck is $\frac{6}{1326} = \frac{1}{221}$.

Example. A club comprises four men and five women. Three members are selected randomly to form a committee. What is the probability that the committee is all women?

Solution. If the committee is all women, then the 3 members must be chosen from among the 5 women. This can be done in $\binom{5}{3} = \frac{5!}{3!2!} = 10$ ways.

Next we need the total number of 3-member committees, choosing from among the total pool of $4 + 5 = 9$ members. This is $\binom{9}{3} = \frac{9!}{3!6!} = 84$.

Hence the desired probability is $\frac{10}{84} = \frac{5}{42} \approx 0.119$.

Conditional probability

The probability of an event A occurring, *given that* an event B has occurred (or will occur), is called a *conditional probability*, and is denoted $P(A \mid B)$.

Example. A family has two children, each being a boy or girl with equal probability 0.5. You know that at least one of the children is a boy. What is the probability that both children are boys?

Solution. In all, there are 4 possibilities: BG, GB, BB, GG, each with equal probability of $0.25 = 0.5 \times 0.5$ (MP).

From what you know, GG is not possible, and we are choosing only from $\{$BB, BG, GB$\}$.

There is a single possibility BB in this set of 3, so the required probability is $\frac{1}{3}$.

In general, if A, B are events in the sample space S, then

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

equivalently, $P(A \cap B) = P(A \mid B) \cdot P(B)$.

A justification for this formula is as follows (cf. the opening example).

The favorable outcomes for the event $E =$ 'A, given B' lie in B.

That is, E has $|A \cap B|$ favorable outcomes. Hence by the definition of probability,

$$P(E) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|S|}{|B|/|S|} = \frac{P(A \cap B)}{P(B)}.$$

Example. In the opening example, we found
$P(\text{both B} \,|\, \text{at least one B}) = \frac{1}{3}$. Also
$P(\text{both B} \,|\, \text{at least one B}) = \frac{1}{4}\big/\frac{3}{4}$, with
$P(\text{both B and at least one B}) = P(\text{both B}) = \frac{1}{4}$,
$P(\text{at least one B}) = \frac{3}{4}$.

Example. A single card is selected from a standard deck. Find the probability that it is a club, given that it is black.

Solution. $P(\clubsuit \,|\, \text{b}) = P(\clubsuit \cap \text{b})\big/P(\text{b}) = P(\clubsuit)\big/P(\text{b}) = \frac{1}{4}\big/\frac{1}{2} = \frac{1}{2}$.

Example. A bag contains 4 red and 8 white balls. Two balls are drawn, without replacement. What is the probability that both are white? Of different colors?

Solution.

Let $W_i = i$th draw is a white, $R_i = i$th draw is a red. (W_1, W_2 are dependent, as are R_1, R_2.)

$P(W_1) = \frac{8}{12} = \frac{2}{3}$.

$P(W_2 \,|\, W_1) = \frac{7}{11}$.

Thus $P(W_1 \cap W_2) = P(W_2 \,|\, W_1)P(W_1) = \frac{2}{3} \cdot \frac{7}{11} = \frac{14}{33}$.

We can work out the probability of different colors in two ways.

In one way, first note

$P(R_1 \cap W_2) = P(W_2 \,|\, R_1)P(R_1) = \frac{8}{11} \cdot \frac{1}{3} = \frac{8}{33}$;

also $P(W_1 \cap R_2) = P(R_2 \,|\, W_1)P(W_1) = \frac{4}{11} \cdot \frac{2}{3} = \frac{8}{33}$.

Thus $P(\text{different colors}) = \frac{8}{33} + \frac{8}{33} = \frac{16}{33}$.

The Monty Hall Problem

On a certain game show, hosted by Monty Hall, you have to choose one of three doors.

Behind only one door is a new car. Behind each of the other two doors is a goat.

You pick a door.

Monty opens a door that you didn't pick (Monty knows what is behind each door).

Behind the door that Monty opened stands a goat.

Monty then asks you whether you want to switch your choice to the other of the two remaining unopened doors.

Should you switch your choice?

In fact, you would *double* your chances of winning the car by switching.

Here is a justification, using conditional probability.

Let C_i be the event that the car is behind door number i, so $P(C_i) = \frac{1}{3}$ for $i = 1, 2, 3$.

Let M_i be the event that Monty chooses door number i to open.

Suppose that you choose door 1 (say, without loss of generality).

If the car is behind door 1 then Monty can pick door 2 or 3 to open; so $P(C_1 \cap M_2) = P(C_1 \cap M_3) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$.

However, if the car is *not* behind door 1, then Monty has only one choice of door to open, and $P(C_2 \cap M_3) = P(C_3 \cap M_2) = \frac{1}{3}$.

If Monty opens door 3, then $P(\text{keep and win}) = P(C_1 \mid M_3) = P(C_1 \cap M_3)/P(M_3) = \frac{1}{6}/P(M_3)$, and

$P(\text{switch and win}) = P(C_2 \mid M_3) = P(C_2 \cap M_3)/P(M_3) = \frac{1}{3}/P(M_3)$.

Thus, if Monty opens door 3, then you are *twice as likely to win if you switch*.

It is the same reasoning if Monty opens door 2; hence, in all cases, you double your chances if you switch.

A large number of trials (and the law of large numbers) bear out the above decision.