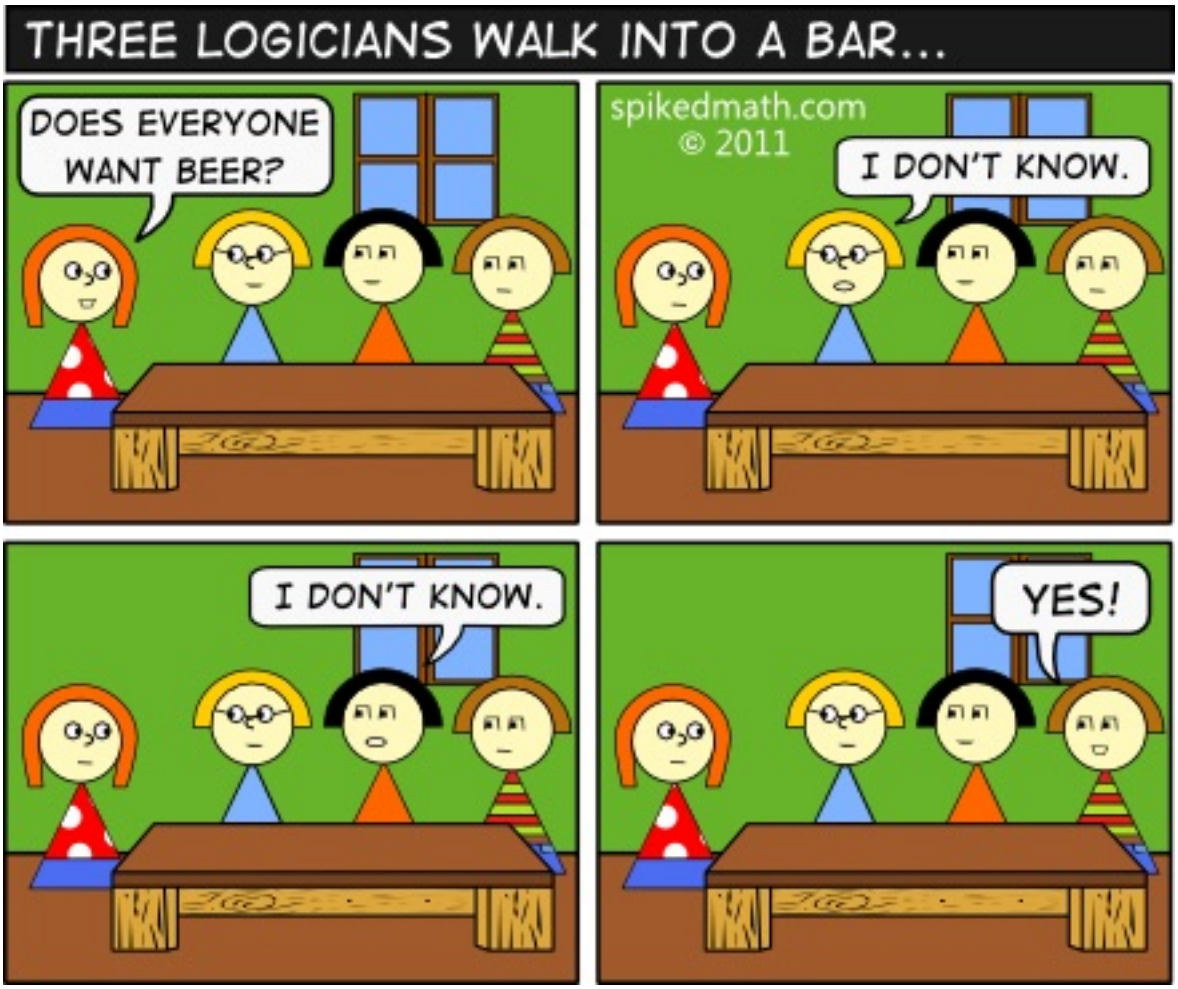
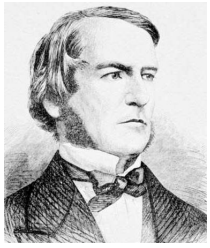
Chapter 1: Logic, truth tables and methods of proof.



Propositions

A proposition is a statement which has a truth value; it is either true (T) or false (F). Compare Boolean data types in programming. Here's a picture of Boole.

Boole: 

He lived in Cork. Look him up. In particular find out how he died.

Examples of Propositions are

the earth is flat
Sarah is a doctor
29 is a prime number.

The first is false and the third is true; the second depends on which 'Sarah' we are talking about, but assuming we agree on who she is we should be able to say if she is a doctor or not.

Sentences of the form

Ireland will win UEFA Euro 2020.
When is the next bus?

are not propositions. The first is a prediction whose truth value we cannot know now and the second is a question.

Since we will be doing operations on propositions and do not want to write the full statements each time we usually abbreviate propositions using capital letters. For example,

P: the earth is flat
Q: Sarah is a doctor
R: 29 is a prime number.

Logical operators
Logical operators allow us to construct new propositions called compound propositions from existing ones.

Negation: The simplest logical operator is negation. The negation of the general proposition P is the proposition **not** P. This is sometimes written as ¬ P. This proposition has truth value T when P is false and truth value F when P is true. For example,
not Q: Sarah is not a doctor

Conjunction: The conjunction logical operator produces a compound proposition P **and** Q from a general pair of propositions P and Q. This is sometimes written as P ∧ Q. This proposition is true if both P and Q are true, otherwise it is false. For example,
P **and** R: the earth is flat and 29 is a prime number

Example: In the logicians joke the statement being considered was a conjunction:
(logician 1 wants beer) **and** (logician 2 wants beer) **and** (logician 3 wants beer)

Disjunction: The disjunction logical operator produces a compound proposition P **or** Q from a general pair of propositions P and Q. This is sometimes written as P ∨ Q. This proposition is only false if both P and Q are false, otherwise it is true. For example,
P **or** R: the earth is flat or 29 is a prime number

Note that this **or** is an 'inclusive or', meaning that P **or** R is true if P is true or R is true or both P and R are true. (Look up the **XOR** logical operator on wikipedia.)

Example: The following example may help to remember the values of **and** , **or** :
Suppose I want to get home by bus:
(a) I will get home if bus A **or** bus B arrives. Of the four possibilities only one means I do not get home, namely when neither bus arrives.

(b) The journey is in two parts. I will get home if bus C arrives **and** bus D arrives where bus C lets me off. Of the four possibilities only one means I get home, namely when both buses arrive.

Python agrees: Python understands 'True', 'False', 'not', 'and' and 'or'. Here's what it understands by 'and' and 'or'.

```
Python 2.7.10 (default, Feb 22 2019, 21:55:15)
[GCC 4.2.1 Compatible Apple LLVM 10.0.1 (clang-1001.0.37.14)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> True and True
True
>>> True and False
False
>>> False and True
False
>>> False and False
False
>>> True or True
True
>>> True or False
True
>>> False or True
True
>>> False or False
False
>>>
```

Truth tables

The truth value of a compound statement can be deduced from that of the original propositions using a truth table. This table has a row for each possible combination of truth values of the simple propositions and will have a column for the truths of each proposition constructed along the way to the final compound proposition. The tables for negation, conjunction and disjunction are

P	**not** P
T	F
F	T

P	Q	P **and** Q
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	P **or** Q
T	T	T
T	F	T
F	T	T
F	F	F

Note: If the compound proposition has one simple proposition as in the case of negation there are two rows in the truth table. If the compound proposition involves two simple propositions as in the case of conjunction or disjunction there are four rows in the truth table. If the compound proposition involves three simple propositions there are eight rows in the truth table, etc.

Example: Compute the truth table for **not** (P **and** Q). Two simple propositions so 4 rows.

P	Q	P **and** Q	**not** (P **and** Q)
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T

Example: Compute the truth table for (**not** P) **or** (**not** Q).

P	Q	**not** P	**not** Q	(**not** P) **or** (**not** Q)
T	T	F	F	F
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T

Note: It seems as if the two previous examples are different descriptions of the same compound proposition.

Logical Equivalence: It can happen that two combinations of applications of the logical operators give compound statements which have the same truth values for every possible set of truth values of the original propositions. Such compound statements are said to be logically equivalent. This can be checked using a combined truth table. We write $R \equiv S$ to denote the fact that R and S are equivalent statements.

Note: It should not be too much of a surprise that logical equivalences arise. Suppose we just consider compound statements using two simple propositions P and Q. The truth table has four rows so the last column is one of 16 possibilities. (T or F in the first row, T or F in the second row, etc.) Therefore lots of compound statements in P and Q must be equivalent.

Example: Perhaps the simplest equivalence is **not** (**not** P) $\equiv$ P.

Example: **not** (P **or** Q) is logically equivalent to (**not** P) **and** (**not** Q):

P	Q	P **or** Q	**not** (P **or** Q)	**not** P	**not** Q	(**not** P) **and** (**not** Q)
T	T	T	F	F	F	F
T	F	T	F	F	T	F
F	T	T	F	T	F	F
F	F	F	T	T	T	T

Here column 3 is the disjunction of columns 1 and 2, column 4 is the negation of column 3, column 5 is the negation of column 1, column 6 is the negation of column 2 and column 7 is the conjunction of columns 5 and 6. Since the fourth and seventh columns are the same the corresponding propositions are logically equivalent.

Example: In the logicians joke the negation of
(logician 1 wants beer) **and** (logician 2 wants beer) **and** (logician 3 wants beer)
is the statement
(logician 1 does not want beer) **or** (logician 2 does not want beer) **or** (logician 3 does not want beer)

Implication

Sometimes we wish to state that the truth of proposition P guarantees the truth of proposition Q. This is known as a conditional proposition and written $P \Rightarrow Q$. Some books may use $P \rightarrow Q$. As an example of implication, consider the propositions

P: It is raining

Q: I drive to work

Here $P \Rightarrow Q$ is the conditional statement 'If it is raining, I drive to work.' Note that it does not mean that I only drive to work when it is raining. I could drive to work for some other reason but I will definitely drive to work if it is raining. This conditional statement is false if P is true and Q is false, otherwise it is true. Thus it is equivalent to (**not** P) **or** Q. The table for $P \Rightarrow Q$ is

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

Exercise: Check for yourself using a truth table that $(P \Rightarrow Q) \equiv ((\textbf{not } P) \textbf{ or } Q)$.

Example: The proposition $((\textbf{not } Q) \Rightarrow (\textbf{not } P))$ is equivalent to $(P \Rightarrow Q)$ and is called the contrapositive of $(P \Rightarrow Q)$.

P	Q	$(P \Rightarrow Q)$	**not** P	**not** Q	$((\textbf{not } Q) \Rightarrow (\textbf{not } P))$
T	T	T	F	F	T
T	F	F	F	T	F
F	T	T	T	F	T
F	F	T	T	T	T

Predicates
Sometimes we want our statements to contain variables and their truth to depend on the value of the variable. Such statements are called predicates.

For example, the statements
x is a computing student who likes mathematics,
x is a mathematician who can program,
x is an integer satisfying $x^2 > 3$
are predicates.

The first can be true or false depending on which computing student x is. Similarly, the second can be true or false depending on which mathematician x is. Finally, the third statement can be true or false depending on which integer x is. Specifically, the third statement is true for $x \leq -2$ or $x \geq 2$ but false for $x =, -1, 0, 1$.
(Remember that the integers are the whole numbers $\ldots -2, -1, 0, 1, 2, 3, \ldots$)

As such, predicates are not propositions since they do not have a truth value.

Predicates arise in programming in code such as

while $(x < 5)$ ...

Here the statement $x < 5$ is a predicate and its truth depends on the value of x which changes as the programme runs.

Quantifiers
These predicates can be turned into propositions using quantifiers such as 'for all' and 'there exists'.

For example, the third predicate above
x is an integer satisfying $x^2 > 3$
can be turned into the proposition
For all integers x, $x^2 > 3$.

Now we can assign a truth value. Specifically, this new proposition is false since there are some integers, $x = 0$ for example, with $x^2 \leq 3$.

Similarly, the predicate can be turned into the proposition
There is an integer x, satisfying $x^2 > 3$

This new proposition is true, since $x = 2$ is an integer with $x^2 > 3$.

Since the expressions 'for all x ' and 'there exists an x' come up so regularly we have shorthand notation for them:

$\forall x$ 'for all x ' $\quad$ $\exists x$ 'there exists an x'

More complicated predicates
When several quantifiers are involved care needs to be taken in determining the truth values.

Example: If x and y are integers and P(x,y) is the statement $xy = 1$, express the following propositions in words and determine their truth value:
$\exists\, x(\exists\, y\ (\mathrm{P}(x,\, y)))$ $\quad$ $\forall\, x(\exists\, y\ (\mathrm{P}(x,\, y)))$

Solution The first says 'there exists an x with the property that there exists a y satisfying $xy = 1$'.

This proposition is true. We could use the values $x = -1$ and $y = -1$.

The second says 'for all x there is a y with $xy = 1$'.

This proposition is false, since we can take $x = 2$ so that $xy = 1$ gives $y = 1/2$ which is not an integer.

Negating Propositions involving Quantifiers
If we construct a proposition from a predicate using a quantifier we can apply the logical operation **not** to it.

The negation of the proposition
For all integers x, $x^2 > 3$
is the proposition
There is an integer x with $x^2 \leq 3$.

The negation of the proposition
There is an integer x satisfying $x^2 > 3$
is the proposition
For all integers x, $x^2 \leq 3$.

In general, we use the following logical equivalences for their negations

not $(\exists\, x\,(\mathrm{P}(x))) \equiv \forall\; x\;$ (**not** $(\mathrm{P}(x))$)

not $(\forall\; x\,(\mathrm{P}(x))) \equiv \exists\; x\;$ (**not** $(\mathrm{P}(x))$)

Note: It may help you remember these negations if you consider the case where x belongs to a finite set $X = \{x_1, x_2, \ldots, x_n\}$. Now the proposition $\exists\; x$ in X satisfying $\mathrm{P}(x)$ is equivalent to $\mathrm{P}(x_1)$ **or** $\mathrm{P}(x_2)$ **or** $\ldots$ **or** $\mathrm{P}(x_n)$ whose negation should be

not$(\mathrm{P}(x_1))$ **and not**$(\mathrm{P}(x_2))$ **and** $\ldots$ **and not**$(\mathrm{P}(x_n))$.

To spell this out further, suppose $X = \{x_1, x_2, x_3\}$. Now $\exists x(P(x))$ is equivalent to $\mathrm{P}(x_1)$ **or** $\mathrm{P}(x_2)$ **or** $\mathrm{P}(x_3)$. Using the rule

not(P **or** Q) $\equiv$ (**not**P) **and** (**not**Q)

we deduce

not$[\mathrm{P}(x_1)$ **or** $\mathrm{P}(x_2)$ **or** $\mathrm{P}(x_3)]$ $\equiv$ **not**$(\mathrm{P}(x_1))$ **and not**$[\,\mathrm{P}(x_2)$ **or** $\mathrm{P}(x_3)]$

$\equiv$ **not**$(\mathrm{P}(x_1))$ **and not**$(\mathrm{P}(x_2))$ **and not**$(\mathrm{P}(x_3))$

Methods of proof
We use logical arguments to prove theorems. In computing, these proofs are used in the verification of algorithms.

We will consider three methods of proof of a proposition of the form (P $\Rightarrow$ Q):

1. Direct argument: Here we simply assume that P is true and show that Q must be true. Since (P $\Rightarrow$ Q) is only false if P is true and Q is false this will complete the proof.

2. Contrapositive argument: Here we assume Q is false and prove P is also false. This uses the equivalence
(P $\Rightarrow$ Q) $\equiv$ ((**not** Q) $\Rightarrow$ (**not** P)).

3. Proof by contradiction: To show (P $\Rightarrow$ Q) we assume P is true and Q is false and derive a contradiction. That is, we show (P **and** (**not** Q)) is false.

Example: Use a direct argument to show that the sum of two even integers has to be even.

Solution: Recall that an integer is even if it is a multiple of 2, that is, an integer x is even if $x = 2y$ for some integer y.

Now suppose a and b are even integers. So $a = 2c$ and $b = 2d$ for some integers c and d.

Now form their sum $a + b$:
$a + b = 2c + 2d = 2(c + d)$

But $c + d$ is an integer so that $a + b$ is an even integer.

Here P was the statement
a is even **and** b is even
while Q was the statement
$a + b$ is even.

Methods of proof

Recall that we considered three methods of proof of a proposition of the form (P $\Rightarrow$ Q):

1. Direct argument: Here we simply assume that P is true and show that Q must be true. Since (P $\Rightarrow$ Q) is only false if P is true and Q is false this will complete the proof.

2. Contrapositive argument: Here we assume Q is false and prove P is also false. This uses the equivalence
(P $\Rightarrow$ Q) $\equiv$ ((**not** Q) $\Rightarrow$ (**not** P)).

3. Proof by contradiction: To show (P $\Rightarrow$ Q) we assume P is true and Q is false and derive a contradiction. That is, we show (P **and** (**not** Q)) is false.

Example: Use a contrapositive argument to show that if the square of a positive integer is an even number then the integer itself must have been even.

Solution: Suppose n is a positive integer and n is not even. We will show that n^2 is also not even.

Recall that an integer which is not even is odd and has the form $2x + 1$ for some integer x.

So $n = 2x + 1$ and therefore

$n^2 = (2x + 1)^2 = 4x^2 + 4x + 1 = 2(2x^2 + x) + 1.$

But $2x^2 + x$ is an integer so that n^2 is an odd integer. Here P was the statement
n^2 is even
while Q was the statement
n is even.
Thus **not** Q was the statement
n is odd
while **not** P was the statement
n^2 is odd.

Example: Use a proof by contradiction argument to show that there is no $x \in \mathbb{Q}$ (fractions) satisfying $x^2 = 2$.

Note: This fact really puzzled the ancient Greeks. Since fractions can be made arbitrarily small and moved around they expected one of them to be

$\sqrt{2}$. They might proceed as follows: Since $1^2 = 1 < 2$ and $2^2 = 4 > 2$ look for a fraction between 1 and 2. Check that $(3/2)^2 = 9/4 > 2 = 8/4$ so look between 1 and 3/2. Check that $(5/4)^2 = 25/16 < 2 = 32/16$ so look between 5/4 and 3/2. Check that $(11/8)^2 = 121/64 < 2 = 128/64$ so look between 11/8 and 3/2, etc.

Solution: Recall that $\mathbb{Q}$ is the set of rational numbers or fractions, that is, numbers of the form p/q where p and q are integers with $q \neq 0$.

Suppose $x \in \mathbb{Q}$ has form $x = p/q$ where p and q are integers with $q \neq 0$. We can also assume that after cancelling common factors of p and q that p and q are not both even.

Now $(p/q)^2 = 2$ means that $p^2 = 2q^2$. This makes p^2 even. Using the result from the last example this means p is also even. So $p = 2r$ for some integer r. Substituting to get $(2r)^2 = 2q^2$ or $4r^2 = 2q^2$ or $2r^2 = q^2$ allows to deduce that q^2 is even and hence that q is even. This gives the contradiction since we assumed that p and q were not both even.

Here P was the statement
$x^2 = 2$
while Q was the statement
x cannot be expressed as p/q, p and q integers with $q \neq 0$
and p and q are not both even.
Thus **not** Q was the statement
$x = p/q$, p and q integers with $q \neq 0$ and p and q are not both even.

Mathematical Induction

Example: What happens when we sum consecutive odd positive integers? Let's check:

1	1
$1+3$	4
$1+3+5$	9
$1+3+5+7$	16
$\vdots$	$\vdots$

It looks like we get perfect squares. Can we come up with a predicate proposition for this and prove that it is always true? Let's start with the proposition.

A positive odd integer is of the form $2k-1$ for some positive integer k so that the odd positive integers are

$$1 = 2(1) - 1, \quad 3 = 2(2) - 1, \quad 5 = 2(3) - 1, \quad 7 = 2(4) - 1, \ldots$$

Now it looks like our rule should be

$$\mathrm{P}(n) : 1 + 3 + 5 + \ldots + (2n-1) = n^2.$$

We would like to prove that this is always true, that is $\forall n, \mathrm{P}(n)$. How can we prove this for infinitely many values of n in a finite time?

The Principle of Mathematical Induction: Let $\mathrm{P}(n)$ be a predicate that is defined for all integers $n \geq 1$. Suppose that

1. $\mathrm{P}(1)$ is true, and
2. $\forall k \geq 1$, $(\mathrm{P}(k) \Rightarrow \mathrm{P}(k+1))$ is true.

Then $\mathrm{P}(n)$ is true for all $n \geq 1$.

The reasoning is that $\mathrm{P}(1)$ is true by 1 and, by 2 applied repeatedly, $\mathrm{P}(2)$ is true, $\mathrm{P}(3)$ is true, $\mathrm{P}(4)$ is true, etc.

The Principle of Mathematical Induction: Let P(n) be a predicate that is defined for all integers $n \geq 1$. Suppose that

1. P(1) is true, and

2. $\forall k \geq 1$, (P(k) $\Rightarrow$ P($k+1$)) is true.

Then P(n) is true for all $n \geq 1$.

The reasoning is that P(1) is true by 1 and, by 2 applied repeatedly, P(2) is true, P(3) is true, P(4) is true, etc.

Example: Show that, for all integers $n \geq 1$:

$$\mathrm{P}(n) : 1 + 3 + 5 + \ldots + (2n-1) = n^2.$$

Base case: ($n = 1$) LHS is $1 = 1$ while RHS is $(1)^2 = 1$ also.

Inductive step: Prove (P(k) $\Rightarrow$ P($k+1$)). So assume P(k) is true, that is

$$1 + 3 + \ldots + (2k-1) = k^2$$

and prove that P($k+1$)) follows. Note that P($k+1$)) states

$$1 + 3 + \ldots + (2k-1) + (2(k+1)-1) = (k+1)^2$$

and the LHS of P($k+1$)) contains the LHS of P(k)). This is the key to this proof.

$$\begin{aligned} 1 + 3 + \ldots + (2k-1) + (2(k+1)-1) &= k^2 + (2(k+1)-1) \\ &= k^2 + 2k + 1 \\ &= (k+1)^2 \end{aligned}$$

However this is the RHS of P($k+1$).

Note: In order to prove the inductive step, we must find some relationship between the statement $P(k)$ and the statement $P(k+1)$. You must use $P(k)$ in your proof of $P(k+1)$. Otherwise it is not an inductive proof.

Example: If a is a positive real number with $a \neq 1$ then

$$1 + a + a^2 + a^3 + \ldots + a^{n-1} = \frac{1 - a^n}{1 - a}.$$

Proof: Base case: $n = 1$, LHS is 1 (only one term) while RHS is $(1-a)/(1-a) = 1$.

Inductive step: Assume $P(k)$, that is,

$$1 + a + a^2 + a^3 + \ldots + a^{k-1} = \frac{1 - a^k}{1 - a}.$$

and try to deduce $P(k+1)$, that is

$$1 + a + a^2 + a^3 + \ldots + a^{k-1} + a^k = \frac{1 - a^{k+1}}{1 - a}.$$

Here we notice again that the LHS of $P(k+1)$ contains the LHS of $P(k)$ and we can substitute for this the RHS of $P(k)$. Thus

$$\begin{aligned} 1 + a + a^2 + a^3 + \ldots + a^{k-1} + a^k &= (1 + a + a^2 + a^3 + \ldots + a^{k-1}) + a^k \\ &= \frac{1 - a^k}{1 - a} + a^k \\ &= \frac{1 - a^k + a^k(1 - a)}{1 - a} \\ &= \frac{1 - a^{k+1}}{1 - a} \end{aligned}$$

which is the RHS of $P(k+1)$.

Note: In the next example, we define a sequence of numbers. Rather than call them
the first number, the second number , the third number, ...
we will write
x_1, x_2, x_3, ...
where x denotes a number and the subscript tells us where in the sequence the number is. This allows to write
'each term in the sequence is obtained by dividing the previous term by the previous term plus three'
as

$$x_{k+1} = \frac{x_k}{x_k + 3}$$

but we have to be careful to distinguish between x_{k+1} which is the value of the term immediately after the kth term and $x_k + 1$ which is the sum of the value of the kth term and the number 1.

Example: If a sequence of numbers $x_1, x_2, \ldots, x_n$ is defined recursively by

$$x_1 = 1 \text{ and } \; x_{k+1} = \frac{x_k}{x_k + 3}$$

compute the first 4 terms and prove, by induction, that

$$x_n = \frac{2}{3^n - 1}$$

For the first part we compute

$$\begin{aligned}
x_1 &= 1 \\
x_2 &= \frac{x_1}{x_1 + 3} \\
&= \frac{1}{4} \\
x_3 &= \frac{x_2}{x_2 + 3} \\
&= \frac{1/4}{1/4 + 3} \\
&= \frac{1}{13} \\
x_4 &= \frac{x_3}{x_3 + 3} \\
&= \frac{1/13}{1/13 + 3} \\
&= \frac{1}{40}
\end{aligned}$$

Base step: $x_1 = 1$ and, from the formula $x_n = 2/(3^n - 1)$ with $n = 1$ we get $2/(3^1 - 1) = 2/2 = 1$ also.

Inductive step: We assume

$$\mathrm{P}(k) : x_k = 2/(3^k - 1)$$

and we want to prove

$$\mathrm{P}(k+1) : x_{k+1} = 2/(3^{k+1} - 1)$$

using the recursion $x_{k+1} = x_k/(x_k + 3)$.

$$
\begin{aligned}
x_{k+1} &= \frac{x_k}{x_k+3} \\
&= \frac{2/(3^k-1)}{2/(3^k-1)+3} \text{ (by P}(k)\text{)} \\
&= \frac{2}{2+3(3^k-1)} \text{ (multiplying above and below by } 3^k-1\text{)} \\
&= \frac{2}{2+3^{k+1}-3} \\
&= \frac{2}{3^{k+1}-1}
\end{aligned}
$$

Note: If the manipulation of fractions causes concern do it slowly

$$
\frac{2/(3^k-1)}{2/(3^k-1)+3} = \frac{2/(3^k-1)}{(2+3(3^k-1))/(3^k-1)} = \frac{2}{3^k-1} \cdot \frac{3^k-1}{2+3(3^k-1)}
$$

where we have used the rule

$$
\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a}{b} \cdot \frac{d}{c}
$$

Webwork

For semester 1, part of the MS121 Continuous assessment is webwork homework.

So do it.

Set 1 due on Thursday week (17th Oct. 2019)

"I cannot get in!"

website	http://webwork.dcu.ie/webwork2/MS121/
username	joseph.bloggs7@mail.dcu.ie (your DCU e-mail address)
password	19202122 (your DCU student ID)

Change the password once you are in.

Test 1

This will take place in tutorial in week 3. Go to the correct tutorial for you based on the first letter of your surname (A-G, H-M, N-Z). Here is the solution to last year's test:

MS121, Test 1, 9th. Oct. 2018

Name: _____	Student No.: _____

?. Let P and Q be propositions defined as follows:
P: I studied hard. Q: I passed my exam.
The compound proposition 'If I studied hard, then I passed my exam.' can be expressed as

(a) Q $\Rightarrow$ P, (b) Q $\Rightarrow$ (**not** P), (c) (**not** Q) $\Rightarrow$ P, (d) (**not** Q) $\Rightarrow$ (**not** P)
Answer: [d] The proposition can be expressed as P $\Rightarrow$ Q which we saw is equivalent to (d).

?. Suppose that R and S are propositions, R has value T and S has value F. Then the two compound statements $[R \Rightarrow (\textbf{not } S)]$ and $[S \Rightarrow (\textbf{not } R)]$ take the following values respectively.

(a) T and T, (b) T and F, (c) F and T, (d) F and F
Answer: [a] **not** S has value T so $[R \Rightarrow (\textbf{not } S)]$ is already true. Similarly S has value F so $[S \Rightarrow (\textbf{not } R)]$ is already true.

?. The negation of the statement 'McGregor won all his fights.' is the following:

(a) McGregor won all his fights. (b) McGregor won some of his fights.
(c) McGregor failed to win any of his fights. (d) McGregor failed to win at least one of his fights.
Answer: [d] If P_i is the statement 'McGregor won his ith fight' then
not $(\forall i\colon P_i)$ is equivalent to $\exists i\colon (\textbf{not } P_1)$

?. A sequence of numbers $x_1, x_2, \ldots, x_n, \ldots$ is defined inductively by
$x_1 = 1$ and $x_{k+1} = x_k/(x_k + k)$ for $k > 1$.
The numbers x_4 and x_5 take the following values respectively.

(a) 1/16 and 1/65, (b) 1/16 and 1/64, (c) 1/15 and 1/61, (d) 1/15 and 1/60
Answer: [a]
$x_2 = (x_1)/(x_1 + 1) = 1/(1 + 1) = 1/2$
$x_3 = (x_2)/(x_2 + 2) = (1/2)/((1/2) + 2) = 1/5$
$x_4 = (x_3)/(x_3 + 3) = (1/5)/((1/5) + 3) = 1/16$
$x_5 = (x_4)/(x_4 + 4) = (1/16)/((1/16) + 4) = 1/65$

Example: For every integer n bigger than 1, $n^2 > n + 1$.

Proof: Base case: $n = 2$ (the first integer bigger than 1), LHS is $2^2 = 4$ while RHS is 3.

For the inductive step we will need some facts about inequalities:
(a) If $a < b$ then $a + c < b + c$ for any number c.
(b) If $a < b$ and $c > 0$ then $ca < cb$.

Inductive step: Assume $P(k)$, that is,

$$k^2 > k + 1.$$

and try to deduce $P(k + 1)$, that is

$$(k + 1)^2 > k + 1 + 1 = k + 2.$$

Bearing in mind that $k > 1$ we argue

$$\begin{aligned} (k+1)^2 &= k^2 + 2k + 1 \\ &> (k+1) + 2k + 1 \text{ (by } P(k) using(a)) \\ &> k + 1 + 2(1) + 1 \text{ (since } k > 1 using(b)) \\ &> k + 2 \end{aligned}$$

Example: All cars have the same colour.
See `https://en.wikipedia.org/wiki/All_horses_are_the_same_color`

This famous false proof by induction goes like this. Let P(n) be the proposition

In any set of n cars all the cars have the same colour

Thus we have to prove P(n) for all positive integers n and induction is the proof method we try. For the base case P(1)

In any set of 1 cars all the cars have the same colour

the statement is true so we just need to prove P(k) $\Rightarrow$ P($k + 1$). Assume P(k) is true and a set of $k + 1$ cars is given. Line them up in a row:

$c_1, c_2, \ldots, c_k, c_{k+1}$

By P(k), the first k cars

$(c_1, c_2, \ldots, c_k), c_{k+1}$

have the same colour, colour 1 say. Also by P(k), the last k cars

$c_1, (c_2, \ldots, c_k, c_{k+1})$

have the same colour, colour 2 say. But the cars in the overlap

c_1, $(c_2, \ldots, c_k)$, c_{k+1}
are the same cars, so colour 1 and colour 2 are the same.

The proof is false since the argument given for $\mathrm{P}(k) \Rightarrow \mathrm{P}(k+1)$ makes the unspoken (false) assumption that there is an 'overlap' so that $k > 1$. In fact the argument for $\mathrm{P}(1) \Rightarrow \mathrm{P}(2)$ is false.

Chapter 2: Sets

A set is a collection of objects. In fact, we do not define the term 'set'. All that matters about the set is what its elements are, and that we can decide if a given object is or is not an element of a given set. There is much similarity between the theory of sets and that of logic. The two subjects should complement each other.

Notation: If an object x belongs to a set A we write $x \in A$, if not we write $x \notin A$. Two sets are equal if they have exactly the same elements. Write $X = Y$ if for every z, $z \in X \Rightarrow z \in Y$ and $z \in Y \Rightarrow z \in X$. Otherwise write $X \neq Y$.

Note: So for a set A and an element x the statement $x \in A$ should be a proposition, either true or false.

Notation: We write our sets by enclosing a description of the elements between braces '{' on left and '}' on right. We can give a finite or infinite list or a predicate description using properties.

Example: $P = \{\text{Norwich City}, \text{Sheffield United}, \text{Aston Villa}\}$ or

$$P = \{T \text{ a Premiership team} \mid T \text{ was promoted this year}\}.$$

Example: $E = \{2, 4, 6, 8, \ldots\}$ or

$$E = \{x \text{ a positive integer } \mid x = 2k \text{ for some integer } k\}$$

Note: You have to be a little careful with the predicate notation in order to avoid things like Russell's Paradox but we will not worry about such things.

Example: There are some standard sets of numbers

$$\begin{aligned}
\mathbb{N} &= \text{the set of all natural numbers} = \{1, 2, 3, \ldots\} \\
\mathbb{Z} &= \text{the set of all integers} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} \\
\mathbb{Q} &= \text{the set of all rational numbers or fractions} \\
\mathbb{R} &= \text{the set of all real numbers} \\
\mathbb{C} &= \text{the set of all complex numbers}
\end{aligned}$$

Definition: There is a special set with no elements called the empty set. There are two notations $\emptyset$ and $\{\}$. The definition is

$$\emptyset = \{x \mid x \neq x\}.$$

Definition: We say that a set A is a subset of a set B or A is contained in B and write $A \subseteq B$ if

$$x \in A \Rightarrow x \in B.$$

Other notation for this situation is $B \supseteq A$, and we say B is a superset of A or B contains A.

Example: The even integers form a subset of $\mathbb{Z}$. We also have

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$$

Proposition: Notice the following three properties

$$A \subseteq A$$

$$A \subseteq B \text{ and } B \subseteq A \Leftrightarrow A = B$$

$$A \subseteq B \text{ and } B \subseteq C \Rightarrow A \subseteq C$$

Definition: If A is a set then the set of all subsets of A is called the power set of A. We will write $P(A)$ for this power set, although some people write 2^A.

Example: If $A = \{1, 2, 3\}$ then

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

Note that A has 3 elements while $P(A)$ has $8 = 2^3$. This is where the notation 2^A comes from. In general a set with n elements will have 2^n subsets, one subset B for each combination of truth values of the n propositions

object one $\in B$, object two $\in B$, ..., object $n \in B$.

For this example, we get

$1 \in B$	$2 \in B$	$3 \in B$	B
T	T	T	$\{1, 2, 3\}$
T	T	F	$\{1, 2\}$
T	F	T	$\{1, 3\}$
T	F	F	$\{1\}$
F	T	T	$\{2, 3\}$
F	T	F	$\{2\}$
F	F	T	$\{3\}$
F	F	F	$\emptyset$

Definition: If A and B are sets, we define their union by

$$A \cup B = \{x \mid x \in A \textbf{ or } x \in B\}$$

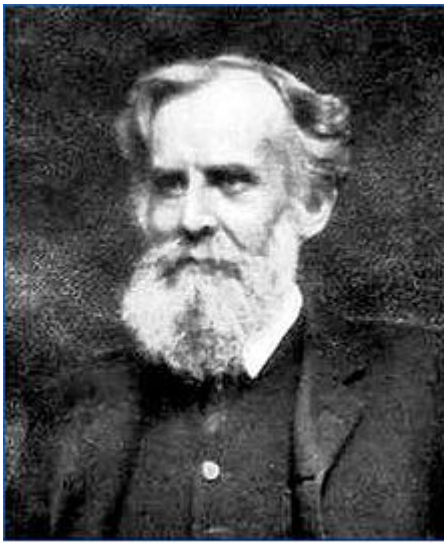
and their intersection by

$$A \cap B = \{x \mid x \in A \textbf{ and } x \in B\}.$$

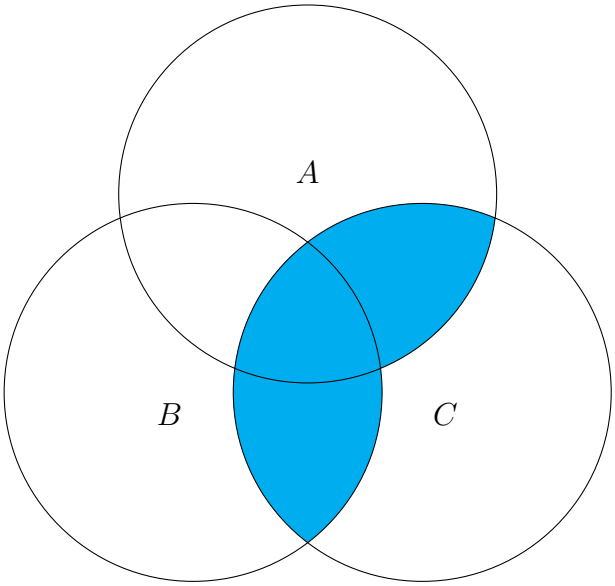
Example: If $A = \{1, 2, 3\}$ and $B = \{0, 2, 5\}$ then

$$A \cap B = \{2\} \quad A \cup B = \{0, 1, 2, 3, 5\}.$$

Venn Diagrams
A Venn diagram is a device for pictorially representing relationships between sets. It is named after John Venn (1834-1923) who popularised its use.



Elliptic regions are drawn to represent sets. The overlapping ellipses define regions corresponding to intersections and several regions will combine to represent unions.



The blue region is $(A \cup B) \cap C$. However it can also be described as $(A \cap C) \cup (B \cap C)$. The equality of these two sets is one of the identities of set theory.

Proposition: The following properties hold

$$A \cap B \subseteq A, \;\; A \cap B \subseteq B, \;\; A \subseteq A \cup B, \;\; B \subseteq A \cup B$$

$$A \cup A = A, \;\; A \cap A = A, \;\; A \cup B = B \cup A, \;\; A \cap B = B \cap A$$

$$A \cup B = B \Leftrightarrow A \subseteq B, \;\; A \cap B = A \Leftrightarrow A \subseteq B$$

$$A \cup (B \cup C) = (A \cup B) \cup C, \;\; A \cap (B \cap C) = (A \cap B) \cap C$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \;\; A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Each of these can be translated into a logical equivalence. We consider the one illustrated by the Venn diagram above, namely

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

Let x be an element and consider the three propositions

$$P : x \in A, \quad Q : x \in B, \quad R : x \in C.$$

Then $x \in (A \cup B) \cap C$ is the compound proposition (P **or** Q) **and** R, while $x \in (A \cap C) \cup (B \cap C)$ is the compound proposition (P **and** R) **or** (Q **and** R). So establishing the equality is the same as establishing the logical equivalence

$$(\text{P } \textbf{or } \text{Q}) \textbf{ and } \text{R} \equiv (\text{P } \textbf{and } \text{R}) \textbf{ or } (\text{Q } \textbf{and } \text{R})$$

We do this in the usual way with a truth table. To save space let X be P **or** Q, Y be X **and** R, U be P **and** R, V be Q **and** R and W be U **or** V.

P	Q	R	X	Y	U	V	W
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	F
T	F	T	T	T	T	F	T
T	F	F	T	F	F	F	F
F	T	T	T	T	F	T	T
F	T	F	T	F	F	F	F
F	F	T	F	F	F	F	F
F	F	F	F	F	F	F	F

Since the Y and W columns are the same the logical equivalence holds. We note that the eight rows of the table correspond to the eight regions of our Venn diagram. For example, the TFT row corresponds to P and R true but Q false which translates to $x \in A$ and $x \in C$ but $x \notin B$ and corresponds to the region in the intersection of A with C which is not in B.

Definition: If A and X are sets we define the complement of A in X to be

$$X - A = \{x \mid x \in X \textbf{ and } x \notin A\}.$$

That is the set consisting of those elements in X which are not also in A.

Note: In applications there is usually a universal set in the background denoted U. This could be the set of all the people in the world, the set of all real numbers, etc.,

Definition: If there is a universal set U, we define the complement of A to be

$$\sim A = U - A = \{x \mid \textbf{not } (x \in A)\}.$$

Note: Since
inclusions, unions, intersections and complements
are defined using
$\Rightarrow$, **or**, **and** and **not**
many identities about sets follow from logical equivalences.

Example: We know **not** (P **and** Q) $\equiv$ (**not** P) **or** (**not** Q). Letting P and Q stand for the statements
P: $x \in A$
Q: $x \in B$

gives **not** $(x \in A$ **and** $x \in B) \equiv$ (**not** $x \in A$) **or** (**not** $x \in B$)

which is the same as

$$\sim(A \cap B) = (\sim A) \cup (\sim B)$$

Here are more properties:

$$A \cup \sim A = U, \;\; A \cap \sim A = \emptyset.$$

If A is not necessarily a subset of X,

$$X - \emptyset = X, \;\; X - X = \emptyset, \;\; A \cap (X - A) = \emptyset.$$

De Morgan's Laws

$$X - (A \cup B) = (X - A) \cap (X - B), \;\; X - (A \cap B) = (X - A) \cup (X - B).$$

De Morgan's Laws for $X = U$

$$\sim(A \cup B) = (\sim A) \cap (\sim B), \ \sim(A \cap B) = (\sim A) \cup (\sim B).$$

Definition: The symmetric difference of two sets is defined by

$$A \triangle B = \{x \mid (x \in A \textbf{ and } x \notin B) \textbf{ or } (x \notin A \textbf{ and } x \in B)\}.$$

Example: If $A = \{1, 2, 3\}$ and $B = \{0, 2, 5\}$ then

$$A \triangle B = \{0, 1, 3, 5\}.$$

Cardinality

Definition: If A is a finite set we write $|A|$ for the number of elements in A and call $|A|$ the cardinality of A.

Proposition: If A and B are finite sets then

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

Proof: From a Venn diagram we see

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

The first two terms give $|A|$ while the last term gives $|B| - |A \cap B|$.

Example: In a survey of 100 phoneowners 55 have iphones, 40 have Samsungs and 3 have both. (!!!) How many have neither?

Solution: Let I be the set of iphone owners in the survey and S be the set of Samsung phone owners in the survey. Here the universal set U is the set of people surveyed. The set of surveyed people with neither an iphone nor a Samsung is $\sim(I \cup S)$. So we want to find $|\sim(I \cup S)|$. We know

$$|\sim(I \cup S)| = |U| - |I \cup S| = 100 - |I \cup S|$$

so we should find $|I \cup S|$ and the last proposition gives

$$|I \cup S| = |I| + |S| - |I \cap S|.$$

Thus we can say

$$\begin{aligned} |I \cup S| &= |I| + |S| - |I \cap S| \\ &= 55 + 40 - 3 = 92 \end{aligned}$$

so that the number with neither phone is $100 - 92 = 8$.

Proposition: If A, B and C are finite sets then

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\quad -|A \cap B| - |A \cap C| - |B \cap C| \\ &\quad +|A \cap B \cap C| \end{aligned}$$

Note: This proposition and the earlier one for two sets are examples of the Inclusion-Exclusion Principle. See its Wikipedia page.

Proof: Look at Venn diagram and record how often regions are counted. I'll write $\sim D$ as D' and $D \cap E$ as DE to save space.

Summand	ABC	ABC'	$AB'C$	$AB'C'$	$A'BC$	$A'BC'$	$A'B'C$	$A'B'C'$
$+\|A\|$	1	1	1	1	0	0	0	0
$+\|B\|$	1	1	0	0	1	1	0	0
$+\|C\|$	1	0	1	0	1	0	1	0
$-\|AB\|$	-1	-1	0	0	0	0	0	0
$-\|AC\|$	-1	0	-1	0	0	0	0	0
$-\|BC\|$	-1	0	0	0	-1	0	0	0
$+\|ABC\|$	1	0	0	0	0	0	0	0
Total	1	1	1	1	1	1	1	0

Example: A group of 100 families are surveyed about where they shop. They all shop at at least one of Tesco, Supervalu or Dunnes Stores. Suppose 45 shop at Tesco, 50 shop at Supervalu and 60 shop at Dunnes, while 22 shop at Tesco and Supervalu, 17 shop at Tesco and Dunnes and 19 shop at Supervalu and Dunnes. How many shop only at Tesco? Only at Supervalu? Only at Dunnes?

Idea: $100 = |T \cup S \cup D|$. Use the proposition to find $|T \cap S \cap D|$. Fill in the rest of the numbers.

$$\begin{aligned} |T \cup S \cup D| &= |T| + |S| + |D| \\ &\quad -|T \cap S| - |T \cap D| - |S \cap D| \\ &\quad +|T \cap S \cap D| \end{aligned}$$

This gives

$$100 = 45 + 50 + 60 - 22 - 17 - 19 + |T \cap S \cap D|$$

so that $|T \cap S \cap D| = 3$. So

$$\begin{aligned}
|T \cap S \cap \sim D| &= |T \cap S| - 3 = 19 \\
|T \cap \sim S \cap D| &= |T \cap D| - 3 = 14 \\
|\sim T \cap S \cap D| &= |D \cap S| - 3 = 16 \\
|T \cap \sim S \cap \sim D| &= |T| - |T \cap S \cap \sim D| - |T \cap \sim S \cap D| - 3 = 9 \\
|\sim T \cap S \cap \sim D| &= |S| - |T \cap S \cap \sim D| - |\sim T \cap S \cap D| - 3 = 12 \\
|\sim T \cap \sim S \cap D| &= |D| - |T \cap S \cap \sim D| - |\sim T \cap S \cap D| - 3 = 27
\end{aligned}$$

An old MS121 final Exam question:

(c) (i) **State** the inclusion-exclusion principle for three finite sets A, B and C.
(ii) Suppose that A, B and C are finite sets with the following properties:

B has two more elements than A; C is twice as big as B;

$A \cap B$ is the same size as $A \cap C$; B and C have no elements in common.

Prove that $|A \cup B \cup C|$ is divisible by 2 (i.e. has an even number of elements).

$$\begin{aligned}
|A \cup B \cup C| &= |A| + |B| + |C| \\
&\quad -|A \cap B| - |A \cap C| - |B \cap C| \\
&\quad +|A \cap B \cap C| \\
&= |A| + (|A| + 2) + 2(|A| + 2) \\
&\quad -|A \cap B| - (|A \cap B|) - (0) \\
&\quad +(0) \\
&= 2(2|A| + 3 - |A \cap B|)
\end{aligned}$$

Here we are not told the sizes of the sets, just some relationships. We translate
B has two more elements than A: $|B| = |A| + 2$
C is twice as big as B: $|C| = 2|B|$
$A \cap B$ is the same size as $A \cap C$: $|A \cap B| = |A \cap C|$
B and C have no elements in common: $|B \cap C| = 0$.
There is still the problem of knowing the size of $A \cap B \cap C$. However, set theory facts help us here. Since B and C have no elements in common, $B \cap C = \emptyset$. Furthermore, since $X \cap Y \subseteq X$ for any two sets X and Y we get

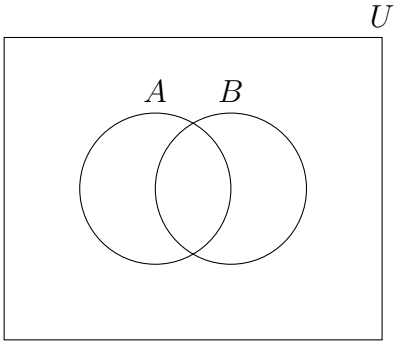
$$A \cap B \cap C = A \cap (B \cap C) \subseteq B \cap C = \emptyset$$

and $A \cap B \cap C = \emptyset$ which gives $|A \cap B \cap C| = 0$.

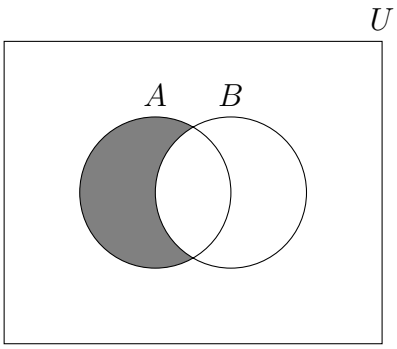
Set theory and methods of proof.

Suppose we want to prove $P \Rightarrow Q$. Let U be the universal set of all things under discussion (Integers, real numbers, triangles, pairs of integers, etc.,) Set $A = \{x \in U \mid P(x)\}$, $B = \{x \in U \mid Q(x)\}$. Proving $P \Rightarrow Q$ is equivalent to showing $A \subseteq B$. Showing $x \in A \Rightarrow x \in B$ is the direct proof.

Now look at a Venn diagram of two general sets A and B in a universal set U.



To say $A \subseteq B$ is equivalent to saying $\sim B \subseteq \sim A$. So we could approach proving $P \Rightarrow Q$ by proving **not** $Q \Rightarrow$ **not** P. This is the contrapositive approach.



Finally, again looking at the Venn diagram we see that $A \subseteq B$ is equivalent to saying $\sim B \cap A = \emptyset$. This gives the proof by contradiction approach.

Product sets

Definition: If A and B are sets their Cartesian product, denoted $A \times B$, is the set whose elements are all possible ordered pairs (a, b), where $a \in A$ and $b \in B$. That is,

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

Example: If $A = \{1, 2\}$ and $B = \{p, q, r\}$ then

$$A \times B = \{(1, p), (1, q), (1, r), (2, p), (2, q), (2, r)\}$$

If A and B are finite we can picture the Cartesian product as a grid. In the case of the last example this would be

$$. (1, r) \qquad . (2, r)$$

$$. (1, q) \qquad . (2, q)$$

$$. (1, p) \qquad . (2, p)$$

Note: If A and B are finite, then $|A \times B| = |A||B|$. (That is the product of the two cardinalities.)

Note: We already view the plane $\mathbb{R}^2$ as $\mathbb{R} \times \mathbb{R}$.

Note: If $A = B$ then we write $A \times A$ as A^2.

Note: We can extend this definition to more than two sets and to a Cartesian product of a set with itself several times.

Example: If $A = \{0, 1\}$, then

$$A^3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

This is the same as the set of binary strings of length 3. We usually write these as

$$000, 001, 010, 011, 100, 101, 110, 111$$

and binary strings of length 3 can be used to represent subsets of a three element set.

Example: The number of elements in A^3 is $|A|^3$. In the above example, $|A| = 2$ and A^3 has $8 = 2^3$ elements. In general, the number of binary strings of length n is 2^n.

Definition: If A and B are sets their Cartesian product, denoted $A \times B$, is the set whose elements are all possible ordered pairs (a, b), where $a \in A$ and $b \in B$. That is,

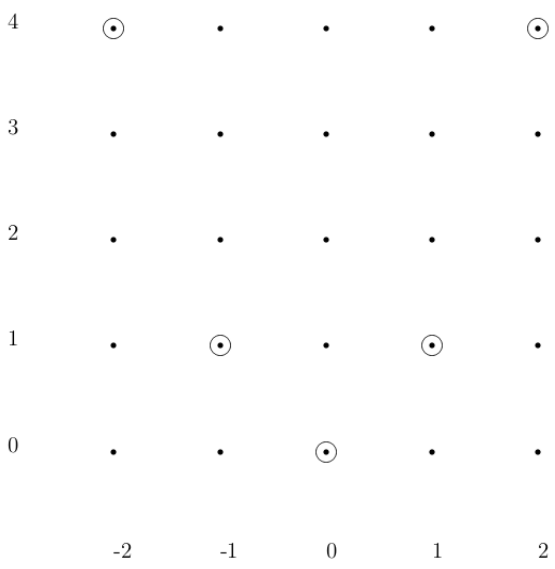
$$A \times B = \{(a, b) \mid a \in A \textbf{ and } b \in B\}$$

Note: In the next two sections we will use Cartesian products to define relations and functions. A relation between sets A and B will simply be a subset of $A \times B$ and a function will be a special type of relation.

Example: Suppose $A = \{-2, -1, 0, 1, 2\}$ and $B = \{0, 1, 2, 3, 4\}$. So $A \times B$ is a 25-element set. It contains the subset

$$S = \{(x, y) \in A \times B \mid y = x^2\}$$

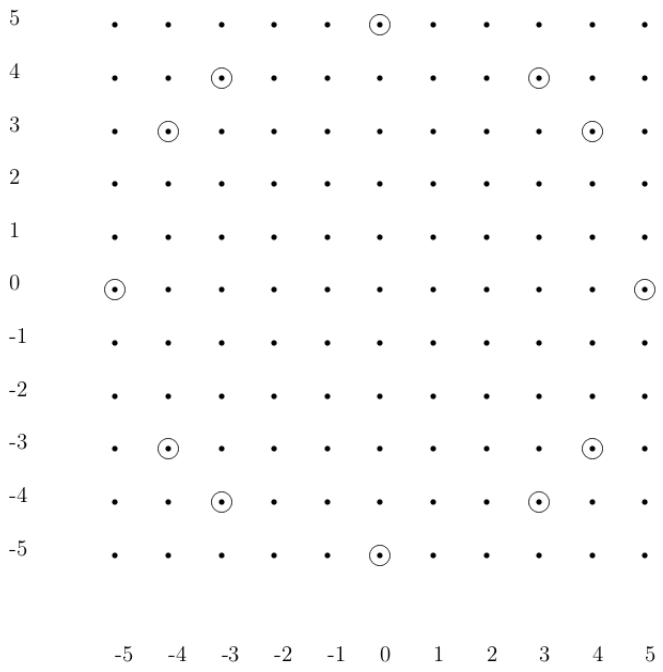
which, in the grid picture, looks like (part of) the graph of the function $y = x^2$.



Example: Suppose $A = B = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$. So $A \times B$ is a 121-element set. It contains the subset

$$S = \{(x, y) \in A \times B \mid x^2 + y^2 = 25\}$$

which, in the grid picture, looks like (part of) the graph of the circle with equation $x^2 + y^2 = 25$. (Recall that $3^2 + 4^2 = 5^2$.)



Chapter 3: Relations.

Here is a table (spreadsheet) describing the processes used in the making of products.

Product	Process
A	2
A	4
B	1
B	4
C	2
C	3

If process i is used in the making of product p we see a row with p in the first column and i in the second column. The table simply lists the elements

of the subset
$\{(A,2),(A,4),(B,1),(B,4),(C,2),(C,3)\}$
of the set $\{A,B,C\} \times \{1,2,3,4\}$.

Relations, from a mathematical point of view, are simply subsets of product sets. The reason they are interesting is that they describe precisely what we mean when we say that some elements of one set are related to some elements of another. For example, if A is the set of students in a university and B is the set of modules taught in the university, then the huge set $A \times B$ contains the much smaller subset

$$R = \{(a,b) \in A \times B \mid \text{student } a \text{ is registered for module } b\}$$

and the subset R is a relation. It completely describes the relationship 'is registered for' between students and modules. In web advertising such relationships are very important. When web user A clicks on link L then that relationship sparks exposure to particular advertising.

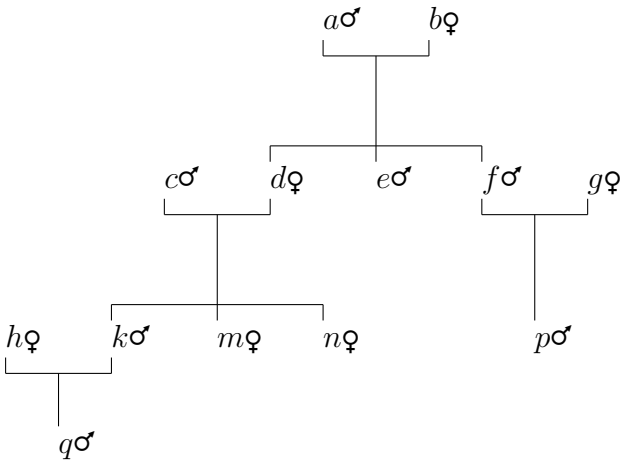
Definition: A binary relation between two sets A and B is a subset R of the Cartesian product $A \times B$.

Note: We could have $A = B$ in which case we refer to R as a relation on A.

Example: If $A = \{1,2\}$ and $B = \{p,q,r\}$, then $A \times B$ has 6 elements $(1,p),(1,q),(1,r),(2,p),(2,q),(2,r)$, so that $A \times B$ has $2^6 = 64$ subsets. This means there are 64 different relations between A and B, ranging from $\emptyset$, where no elements of A are related to elements in B to $A \times B$, where every element of A is related to every element of B.

In between are the other 62 relations, including 6 where exactly one element of A is related to one element of B, 2 where one element of A is related to all elements of B, 3 where all elements of A are related to a single element of B, etc.,

Example: In English, we think of the word relation as most closely connected to family relationships. Here's a family tree.



What are the relations
(a) R_1: Is an uncle of,
(b) R_2: Is a grandchild of?

(a) $R_1 = \{(e,k),(e,m),(e,n),(e,p),(f,k),(f,m),(f,n)\}$

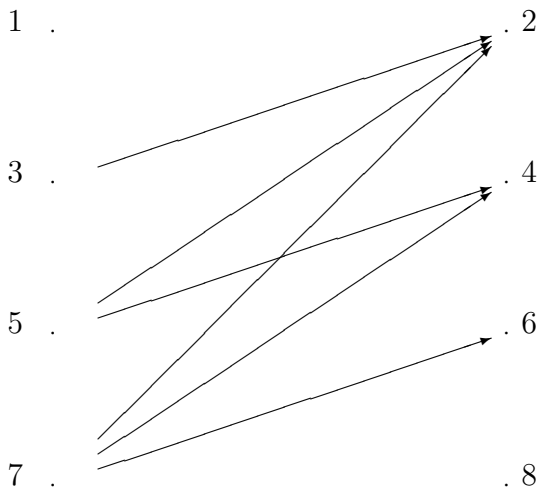
(b) $R_2 = \{(k,a),(k,b),(m,a),(m,b),(n,a),(n,b),(p,a),(p,b),(q,c),(q,d)\}$

Example: For the sets of integers $A = \{1,3,5,7\}$ and $B = \{2,4,6,8\}$ what are the relations
(a) $R_1 = \{(x,y) \in A \times B \mid x > y\}$
(b) $R_2 = \{(x,y) \in A \times B \mid x - y \text{ is even }\}$.
For (a): $R_1 = \{(3,2),(5,2),(5,4),(7,2),(7,4),(7,6)\}$. For (b) $R_2 = \emptyset$.

Note: For small examples, there are devices which help us visualise relations.

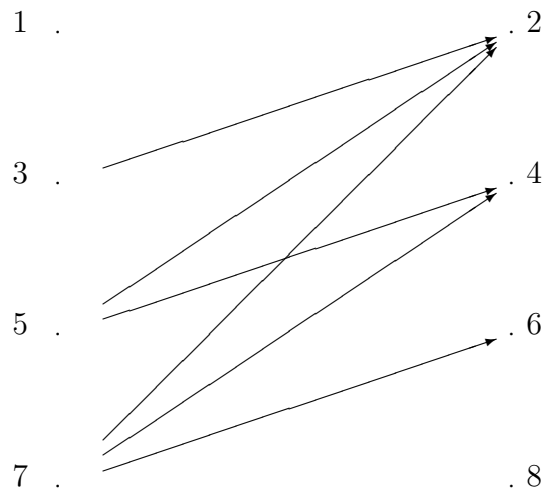
Definition: If R is a binary relation between sets A and B the digraph of the relation is a graph with vertex set $A \cup B$ and an edge joining a to b if $(a,b) \in R$.

Example: For R_1 in the last example, the digraph is


1
3
5
7
2
4
6
8

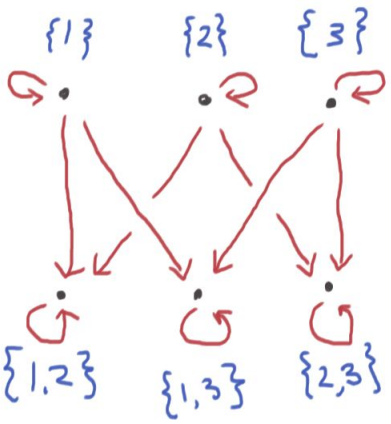
Definition: If R is a binary relation between sets A and B the digraph of the relation is a graph with vertex set $A \cup B$ and an edge joining a to b if $(a, b) \in R$.

Example: For the last example, with $A = \{1, 3, 5, 7\}$ and $B = \{2, 4, 6, 8\}$ and R_1 the relation $x > y$, the digraph is

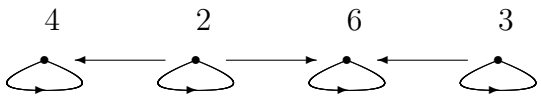


Note: For the case $A = B$, we just use A as the vertex set of the digraph. This is because $A \cup A = A$.

Example: The set of 1 or 2 element subsets of $\{1, 2, 3\}$ with the relation $A \subseteq B$.



Example: If $A = \{2, 3, 4, 6\}$ and R is the relation given by 'is a divisor of' then the corresponding digraph is



Note: Recall how we visualised the Cartesian product of A and B by an array. This required ordering the elements of A and B by $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$. This array allows us to specify a relation R by placing a T or F in the appropriate entry of the array.

Definition: If R is a relation between A and B and A and B are ordered by $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$, then the matrix of R is the array whose (i, j)th entry is T if $(a_i, b_j) \in R$ and F if $(a_i, b_j) \notin R$.

Example: For $A = \{1, 3, 5, 7\}$ and $B = \{2, 4, 6, 8\}$ the relation $x > y$ is given by the matrix

$$\begin{pmatrix} F & F & F & F \\ T & F & F & F \\ T & T & F & F \\ T & T & T & F \end{pmatrix}$$

Example: If $A = B$ is the set of all pages on the web and R is the relation 'web page A links to web page B' a variation on the corresponding matrix

is used to compute the pagerank of each page on the web. This is done regularly by Google in what is called the largest matrix calculation in the world.

Note: Sometimes a relation R between A and B is simply specified by writing the predicate statement aRb whenever $(a, b) \in R$. (Recall that a predicate statement is one involving variables.)

Example: We could just write
$3 > 2$, $5 > 2$, $5 > 4$,$7 > 2$, $7 > 4$, $7 > 6$
to specify the relation $a > b$ between $A = \{1, 3, 5, 7\}$ and $B = \{2, 4, 6, 8\}$.

Properties of a relation on a single set.

To illustrate important properties that a relation on a single set might have we will consider the following examples of relations.

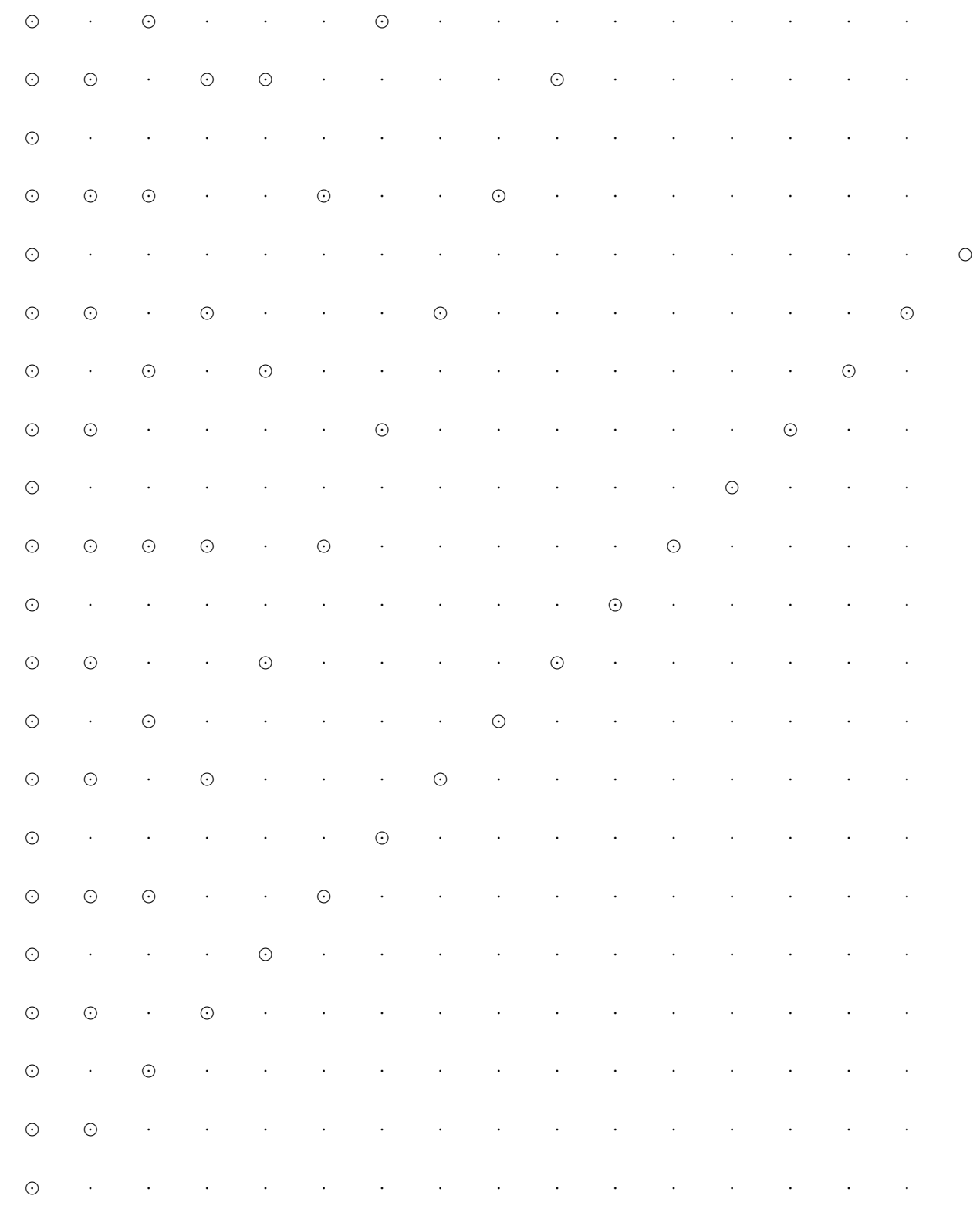
Example 1: If $A = \{1, 2, 3, 4\}$ and $R = \{(2, 1), (2, 3), (3, 2), (4, 4)\}$.

Example 2: If A is the set of all pages on the web and R is the relation 'web page X links to web page Y'.

Example 3: If $A = \mathbb{N} = \{1, 2, 3, \ldots\}$ and R is the relation given by 'is a divisor of'. The subset R of $A \times A$ is shown in a figure below. $A \times A$ is an infinite grid with lower left corner at $(1, 1)$ and the elemeents of R are circled.

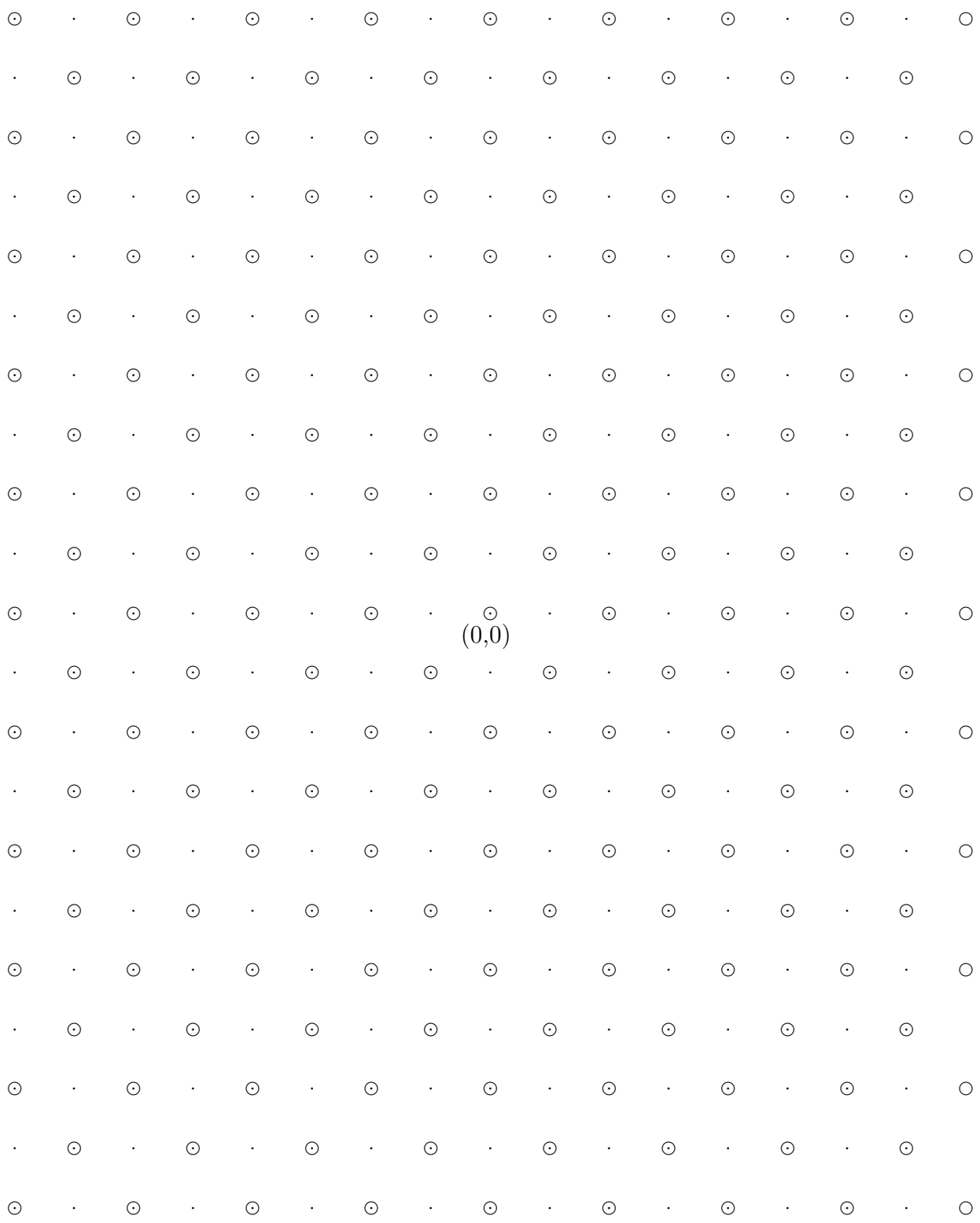
Example 4: If $A = \mathbb{Z}$ and R is the relation given by $(x, y) \in R$ if and only if $y - x$ is even. The subset R of $A \times A$ is shown in a figure below. $A \times A$ is an infinite grid with no corners this time and the elements of R are circled.

Example 5: Let A be the set of first year students in DCU and define the relation R on A by $(a, b) \in R$ if b is in the same programme as a.



(1,1)

The relation ‘a divides b’ inside $\{1, 2, 3, 4, \ldots\} \times \{1, 2, 3, 4, \ldots\}$.



The relation '$y - x$ is even' inside $\mathbb{Z} \times \mathbb{Z}$.

Properties of a relation on a single set.

To illustrate important properties that a relation on a single set might have we will consider the following examples of relations.

Example 1: If $A = \{1, 2, 3, 4\}$ and $R = \{(2, 1), (2, 3), (3, 2), (4, 4)\}$.

Example 2: If A is the set of all pages on the web and R is the relation 'web page X links to web page Y'.

Example 3: If $A = \mathbb{N} = \{1, 2, 3, \ldots\}$ and R is the relation given by 'is a divisor of'. The subset R of $A \times A$ is shown in a figure below. $A \times A$ is an infinite grid with lower left corner at $(1, 1)$ and the elements of R are circled. (The parallel lines we see are explained as follows: if $(a, b) \in R$ then $b = ca$ so that $b + kc = ca + kc = c(a + k)$ giving $(a + k, b + kc) \in R$.)

Example 4: If $A = \mathbb{Z}$ and R is the relation given by $(x, y) \in R$ if and only if $y - x$ is even. The subset R of $A \times A$ is shown in a figure below. $A \times A$ is an infinite grid with no corners this time and the elements of R are circled. (The recurring pattern of circles is due to: even - even = even, even - odd = odd, odd - even = odd, odd - odd = even.)

Example 5: Let A be the set of first year students in DCU and define the relation R on A by $(a, b) \in R$ if b is in the same programme as a.

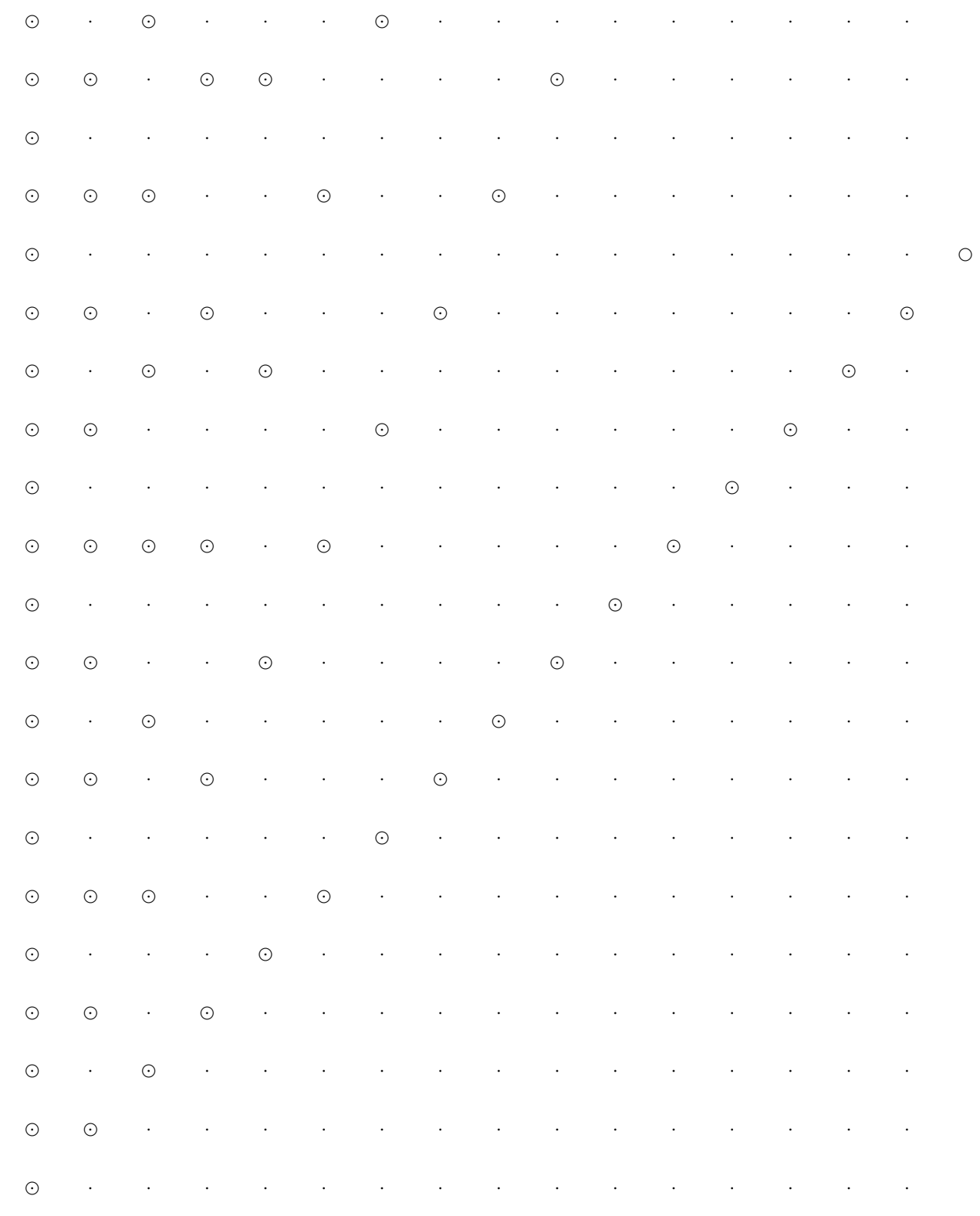
Definition: A relation R on a set A is called reflexive if $(a, a) \in R$ for all $a \in A$.

Note: The digraph of a reflexive relation will have loops at every element of A and the matrix will have T's on the diagonal.

Examples: The R in Example 1 is not reflexive. $(1, 1) \notin R$. The R in Example 2 is not reflexive. Most webpages do not link to themselves. The R in Example 3 is reflexive. If $a \in \{1, 2, 3, \ldots\}$ then a divides a since $a = a(1)$. The R in Example 4 is reflexive. If $x \in \mathbb{Z}$ then $x - x = 0 = 2(0)$ is even so $(x, x) \in R$. The R in Example 5 is reflexive. Every student is in the same programme as himself/herself.

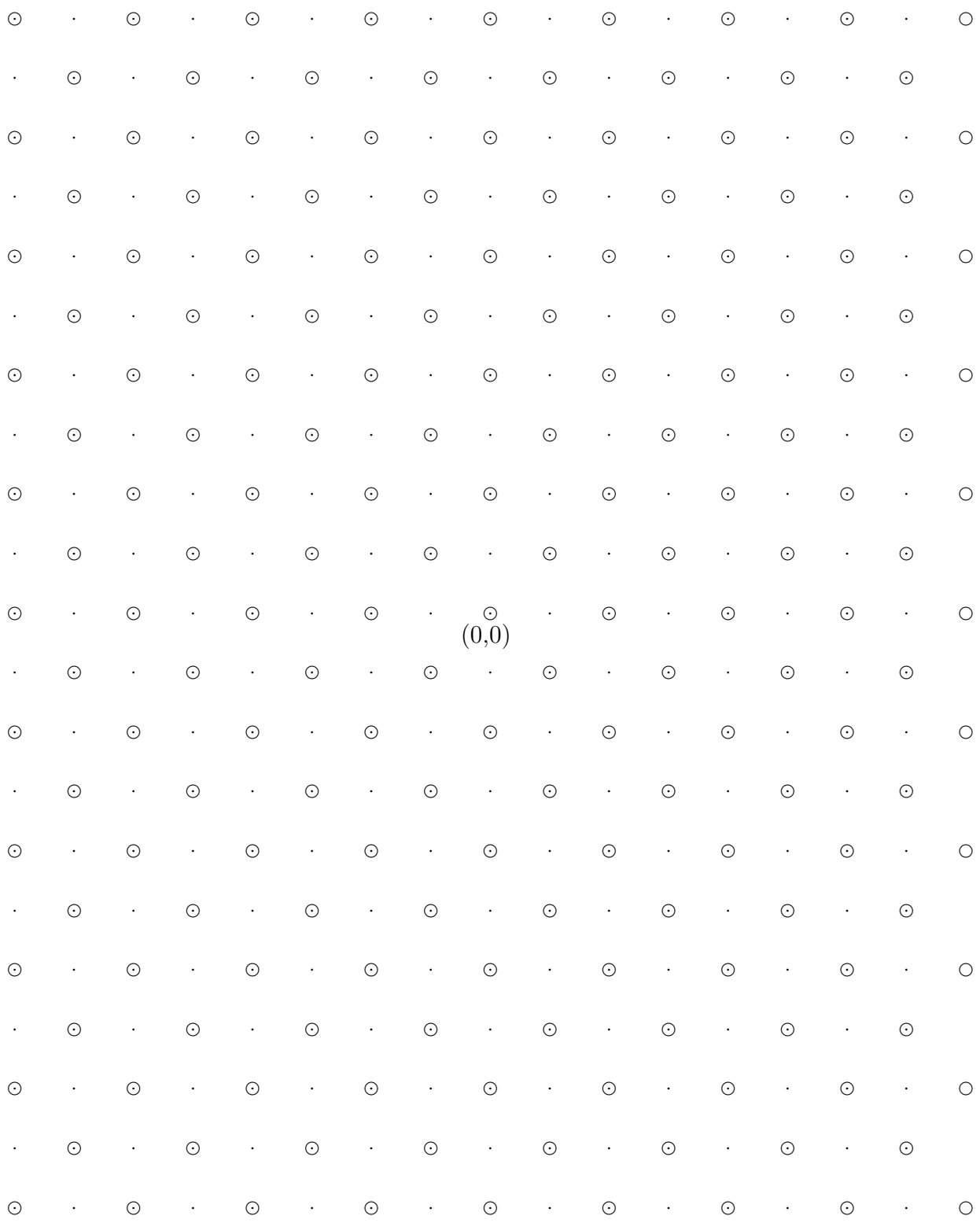
Definition: A relation R on a set A is called symmetric if whenever we have $(a, b) \in R$ we also have $(b, a) \in R$.

Note: The digraph of a symmetric relation will have, for each edge from a to b, an edge going in the opposite direction from b to a, and the matrix will have a T below the diagonal if and only if there is a T above the diagonal in



(1,1)

The relation ‘a divides b’ inside $\{1, 2, 3, 4, \ldots\} \times \{1, 2, 3, 4, \ldots\}$.



The relation '$y - x$ is even' inside $\mathbb{Z} \times \mathbb{Z}$.

the corresponding position.

Examples: The R in Example 1 is not symmetric. $(2,1) \in R$ but $(1,2) \notin R$. The R in Example 2 is not symmetric. Your page may link to an Irish Times article page but the page will not link back to you. The R in Example 3 is not symmetric. $(2,4) \in R$ since $4 = 2(2)$ but $(4,2) \notin R$ since $2 = (1/2)4$ and $1/2$ is not an integer. The R in Example 4 is symmetric. If $(x,y) \in R$ then $y - x = 2k$ for some integer k. However this gives $x - y = (-1)(y - x) = -2k = 2(-k)$ and $(y,x) \in R$. The R in Example 5 is symmetric. If student A is in the same programme as student B then student B is in the same programme as student A.

Definition: A relation R on a set A is called antisymmetric if whenever we have $(a,b) \in R$ and $(b,a) \in R$, then $a = b$.

Note: The digraph of an antisymmetric relation will have, for each edge from a to b with $b \neq a$, no edge going in the opposite direction from b to a, and the matrix will have for every T off the diagonal an F across the diagonal in the corresponding position.

Examples: The R in Example 1 is not antisymmetric. $(2,3) \in R$ and $(3,2) \in R$ but $2 \neq 3$. The R in Example 2 is not antisymmetric. There are many pairs of friends who link to each others webpages. The R in Example 3 is antisymmetric. If $(a,b) \in R$ and $(b,a) \in R$ the $b = ka$ and $a = lb$ for some positive integers k and l. However this means $a = lb = l(ka) = (lk)a$ so that $lk = 1$ and hence $k = 1$ and $l = 1$. This gives $a = b$. The R in Example 4 is not antisymmetric. $(2,6) \in R$ since $6 - 2 = 4$ is even and $(6,2) \in R$ since $2 - 6 = -4$ is even but $2 \neq 6$. The R in Example 5 is not antisymmetric as long as there are two different students in the same programme.

Definition: A relation R on a set A is called transitive if whenever we have $(a, b) \in R$ and $(b, c) \in R$ we also have $(a, c) \in R$.

Note: If, in the digraph of a transitive relation R, one can get from a to b following arrows then there must be an arrow from a to b. (If we see two edges following each other the third side of the directed triangle must also be present.)

Examples: The R in Example 1 is not transitive. $(3, 2) \in R$ and $(2, 1) \in R$ but $(3, 1) \notin R$. Here $a = 3$, $b = 2$ and $c = 1$. Note that we also have $(3, 2) \in R$ and $(2, 3) \in R$ but $(3, 3) \notin R$. (a and c could be equal.) The R in Example 2 is not transitive. If I link to your webpage, I do not necessarily link to the crazy sites that you link to. The R in Example 3 is transitive. If $(a, b) \in R$ and $(b, c) \in R$ then $b = ka$ and $c = lb$ for some positive integers k and l. However this means $c = lb = l(ka) = (lk)a$ so that $(a, c) \in R$. The R in Example 4 is transitive. If $(x, y) \in R$ and $(y, z) \in R$ then $y-x = 2k$ and $z-y = 2l$ for some integers k and l. However this gives $z-x = z-y+(y-x) = 2l+2k = 2(l+k)$ and $(x, z) \in R$. The R in Example 5 is transitive. If student A is in the same programme as student B and student B is in the same programme as student C then all three are in the same programme and student A is in the same programme as student C.

Definition: If A is a set and R is a relation on A which is not reflexive we define the reflexive closure of R to be the relation consisting of R together with $\{(a, a) \mid a \in A\}$. (Recall that this is just a union of sets.)

Definition: If A is a set and R is a relation on A which is not symmetric we define the symmetric closure of R to be the relation consisting of R together with $\{(a, b) \mid (b, a) \in R\}$.

Definition: If A is a set and R is a relation on A which is not transitive we define the transitive closure of R to be the relation R^* on A defined by aR^*b if and only if there is a sequence of elements of A, $a_1, a_2, \ldots, a_k$ with $a = a_1$, $b = a_k$ and $a_i R a_{i+1}$.

Example: Transitive closures of relations are the most applicable. If A is a set of websites and R is the relation define by 'is linked to', then the transitive closure of R is the relation 'is connected to by a sequence of links'.

Example: If $A = \{1, 2, 3\}$ and R is the relation $\{(1, 3), (2, 1), (2, 2), (2, 3), (3, 2)\}$, then R is not reflexive, not symmetric and not transitive. What are the corresponding closures?

The reflexive closure is

$$\{(1,1),(1,3),(2,1),(2,2),(2,3),(3,2),(3,3)\}$$

Here we had to add $(1,1)$ and $(3,3)$ to get $(a,a) \in R^*$ for all $a \in A$.

The symmetric closure is

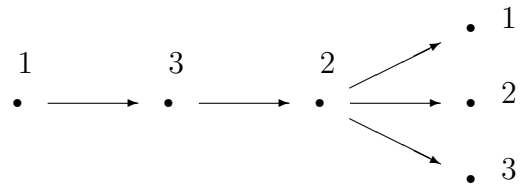
$$\{(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2)\}$$

Here we had to add $(1,2)$ and $(3,1)$ to get $(b,a) \in R^*$ for all $(a,b) \in R$.

The transitive closure is

$$\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),\}$$

Here the construction is more complicated. We see that $1R3$ and $3R2$ so that transitivity of R^* will force $(1,2) \in R^*$. However $1R^*2$ and $2R^*1$ will force $(1,1) \in R^*$. Thus 1 is related to every element of A by R^*. Continue in this way for 2 and 3.



Definition: A relation R on a set A is called an equivalence relation if R is reflexive, symmetric and transitive.

Example: Let A be a set of first year DCU students and define the relation R on A by $(a, b) \in R$ if b is in the same programme as a. R is an equivalence relation.

Note: Both the digraph of R and the matrix of R have special properties. We'll look at them in a smaller example.

Example: Suppose $A = \{1, 2, 3, 4, 5, 6\}$ and let $A_1 = \{1, 2, 5\}$, $A_2 = \{4, 6\}$ and $A_3 = \{3\}$. Then

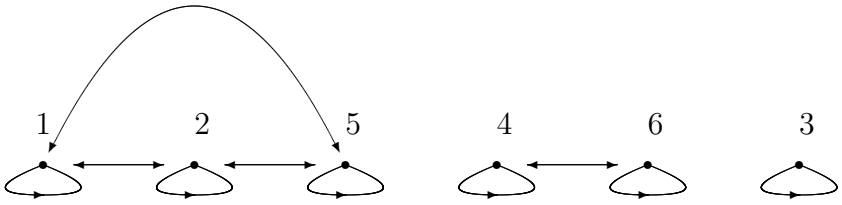
$$A_1 \cap A_2 = \emptyset, \quad A_1 \cap A_3 = \emptyset, \quad A_2 \cap A_3 = \emptyset, \quad A_1 \cup A_2 \cup A_3 = A.$$

Put a relation R on A by saying $(x, y) \in R$ if x and y are in the same subset A_1, A_2 or A_3. So

$$R = \{(1,1), (1,2), (1,5), (2,1), (2,2), (2,5), (3,3),$$

$$(4,4), (4,6), (5,1), (5,2), (5,5), (6,4), (6,6)\}$$

(This is similar to the last example but with more manageable numbers. You can think of A_1 as the students in the $CA1$ programme, A_2 as the students in the $BS1$ programme and A_3 as the students in the $AL1$ programme. The intersection and union properties correspond to the fact that each student is in precisely one programme.)
R is an equivalence relation: Each element is in the same subset as itself; if b is in the same subset as a then a is in the same subset as b; if b is in the same subset as a and c is in the same subset as b all three are in the same subset and c is in the same subset as a. Here is the digraph:



Here is the matrix:

	1	2	5	3	4	6
1	T	T	T	F	F	F
2	T	T	T	F	F	F
5	T	T	T	F	F	F
3	F	F	F	T	F	F
4	F	F	F	F	T	T
6	F	F	F	F	T	T

Definition: If A is a non-empty set, a partition of A is a collection of non-empty subsets, $A_1, A_2, \ldots, A_n$ satisfying
1. $A = A_1 \cup A_2 \cup \ldots \cup A_n$ and
2. $A_i \cap A_j = \emptyset$ whenever $i \neq j$.
The sets A_i are called the blocks of the partition.

Example: If A and B are sets, the collection $\{A \cap (\sim B), A \cap B, (\sim A) \cap B\}$ forms a partition of $A \cup B$.

Example: For $A = \mathbb{Z}$, $n = 2$, A_1 the set of odd integers and A_2 the set of even integers, we get a partition of $\mathbb{Z}$. $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \mathbb{Z}$.

Example: Let A be a set of first year DCU students taking CA, BS or AL, A_1 be the set of CA1 students in A, A_2 be the set of BS1 students in A and A_3 be the set of AL1 students in A. Then the collection of subsets A_1, A_2, A_3 defines a partition of A.

Proposition: Suppose $A_1, A_2, \ldots, A_n$ is a partition of a set A and a relation R is defined on A by $(x, y) \in R$ if and only if x and y belong to the same block of the partition. Then R is an equivalence relation.

Proof: Every element $a \in A$ is in the same block as itself, giving $(a, a) \in R$. This makes R reflexive. If $(a, b) \in R$, then a and b belong to the same block of the partition. However, this means b and a belong to the same block of the partition and $(b, a) \in R$. This makes R symmetric. If $(a, b) \in R$ and $(b, c) \in R$, then a and b belong to the same block of the partition and b and c belong to the same block of the partition. However, this means all three belong to the same block of the partition and $(a, c) \in R$. This makes R transitive.

Definition: If R is an equivalence relation on a set A and $a \in A$ we define the equivalence class of a to be the subset of A given by

$$E_a = \{b \in A \mid (a, b) \in R\}.$$

Example: Let A be a set of first year DCU students taking CA, BS or AL, A_1 be the set of CA1 students in A, A_2 be the set of BS1 students in A and A_3 be the set of AL1 students in A. Define the relation R by aRb if b is in the same programme as a. If a is a CA1 student then $E_a = A_1$.

Example: Let $A = Z$ and say aRb if $b - a$ is divisible by 3. Then

$$E_5 = \{\ldots, -4, -1, 2, 5, 8, 11, \ldots\}$$

$$E_9 = \{\ldots, -6, -3, 0, 3, 6, 9, \ldots\}$$

Example: Let A be the set of ratios p/q where p and q are integers and $q \neq 0$. Define the relation R on A by $(p/q)R(s/t)$ if $pt = qs$. Then

$$E_{3/2} = \{\ldots . - 6/(-4), -3/(-2), 3/2, 6/4, 9/6, \ldots\}$$

The equivalence classes are the usual rational numbers $\mathbb{Q}$.

Theorem: Let R be an equivalence relation on a non-empty set A. Then the collection of distinct equivalence classes form a partition of A.

Proof: For each $a \in A$ we have $a \in E_a$ since R is reflexive. This means the equivalence classes are nonempty and the union of the equivalence classes gives all of A. Next we show that $(a, b) \in R$ means $E_a = E_b$. (Recall that we show sets X and Y are equal by showing $X \subseteq Y$ and $Y \subseteq X$.) Suppose

$c \in E_a$. This means $(a, c) \in R$. We know $(a, b) \in R$, so that $(b, a) \in R$ by symmetry, and $(b, c) \in R$ by transitivity. So $c \in E_b$ and $E_a \subseteq E_b$. Similarly, $E_b \subseteq E_a$ and $E_a = E_b$. Finally suppose $E_a \cap E_b \neq \emptyset$. So there is a c in the intersection. This means $(a, c) \in R$ and $(b, c) \in R$. Again symmetry and transitivity give $(a, b) \in R$ and $E_a = E_b$. Thus the distinct equivalence classes are disjoint.

Definition: A partial order on a set A is a relation R which is reflexive, antisymmetric and transitive.

Note: Recall that
(i) R is reflexive if $(a, a) \in R$ for all $a \in A$,
(ii) R is antisymmetric if whenever we have $(a, b) \in R$ and $(b, a) \in R$ then $a = b$, and
(iii) R is transitive if whenever we have $(a, b) \in R$ and $(b, c) \in R$ we also have $(a, c) \in R$

Example: For $A = \mathbb{R}$, the 'less than or equal to' relation,

$$R = \{(x, y) \mid x \leq y\}$$

is a partial order. Think of $x \leq y$ if and only if $y - x \geq 0$. $x \leq x$ for each $x \in \mathbb{R}$ since $x = x$. Thus $\leq$ is reflexive. If $x \leq y$ and $y \leq x$ then $x = y$ so that $\leq$ is antisymmetric. If $x \leq y$ and $y \leq z$ then $x \leq z$ so that $\leq$ is transitive.

Example: For $A = P(B)$ (the power set or set of all subsets of a set B) the relation

$$R = \{(B_1, B_2) \mid B_1 \subseteq B_2\}$$

is a partial order. Certainly $B_1 \subseteq B_1$ so that $\subseteq$ is reflexive. If $B_1 \subseteq B_2$ and $B_2 \subseteq B_1$ then B_1 and B_2 have the same elements and $B_1 = B_2$, making $\subseteq$ antisymmetric. If $B_1 \subseteq B_2$ and $B_2 \subseteq B_3$ then each element of B_1 is an element of B_3 so that $B_1 \subseteq B_3$ and $\subseteq$ is transitive.

Example: For $A = \mathbb{N}$ the relation

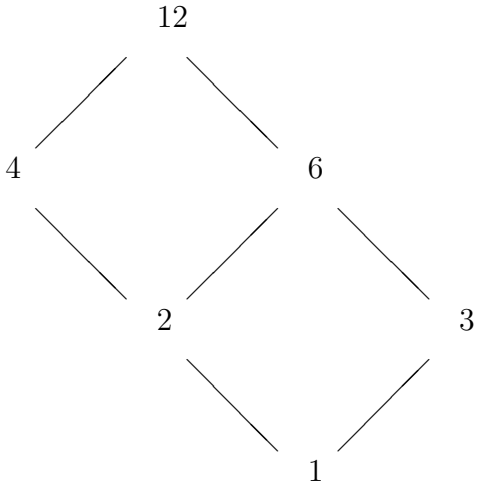
$$R = \{(a, b) \mid a \text{ is a divisor of } b\}$$

is a partial order. For $a \in \mathbb{N}$, $a = a(1)$ so that a is a divisor of a and R is reflexive. If $a, b \in \mathbb{N}$ satisfy aRb and bRa, then $b = ca$ and $a = db$. This gives $a = db = dca$ so that $dc = 1$. Since d and c are in $\mathbb{N}$ we have $c = d = 1$ and $b = a$. This makes R antisymmetric. Finally, suppose $a, b, c \in \mathbb{N}$ satisfy aRb and bRc. Then $b = pa$ and $c = qb$ so that $c = qb = qpa$. But this gives aRc and R is transitive.

Terminology: A set with a partial order on it is called a partially ordered set or poset. If R is a partial order on A and $(a, b) \in R$ with $a \neq b$ we say a

is a predecessor of b and b is a successor of a. If a is a predecessor of b and there is no successor of a which is a predecessor of b the we say that a is an immediate predecessor of b and write $a \prec b$. Instead of drawing a diagraph of R we often draw a subgraph called a Hasse diagram with an edge connecting a to b whenever $a \prec b$.

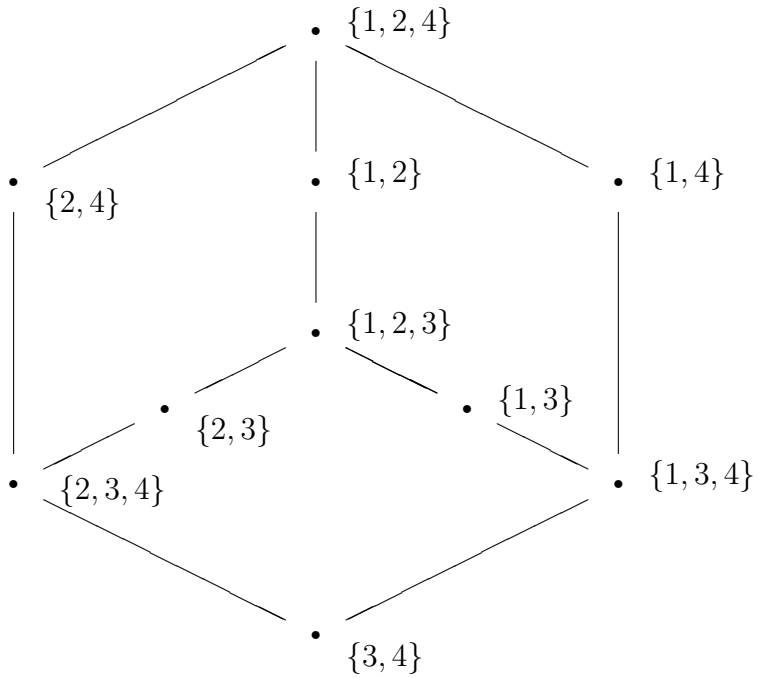
Example: A is the set of factors of the integer 12 with partial order given by 'is a divisor of'. Draw the Hasse diagram.



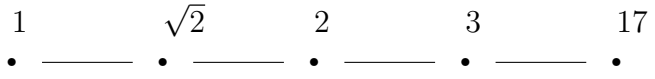
Example: A is the set of subsets of $\{1, 2, 3, 4\}$ given by

$$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{1,3,4\}$$

with partial order given by inclusion. Draw the Hasse diagram.



Example: A is the set of real numbers $\{1, 2, \sqrt{2}, 3, 17\}$ with partial order given by $x \leq y$. Draw the Hasse diagram.



Definition: A partial order R on a set A is called a total order if whenever a and b belong to A and $a \neq b$ then exactly one of $(a, b) \in R$ or $(b, a) \in R$ holds.

Example: $A = \mathbb{R}$ with the partial order $\leq$ is totally ordered. (If $x, y \in \mathbb{R}$ and $x \neq y$, then we have either $y - x > 0$ so that $x \leq y$ or we have $y - x < 0$ so that $y \leq x$.)

Example: If A is a totally ordered set then $A \times A$ can be ordered lexico-

graphically by

$$(a_1, a_2) \leq (a_3, a_4) \text{ if } (a_1 < a_3) \textbf{ or } (a_1 = a_3 \textbf{ and } a_2 \leq a_4).$$

This can be extended to several copies of A.

Example: If $A = \{a, b, \ldots, z\}$ and words are elements of $A \times A \times A \times \ldots$ then the lexicographic order is the one found in the dictionary where we have an end of word character which precedes a. Put these words in lexicographic order:
apple, aardvark, anthill, ant, antacid, antithesis

The correct order is

aardvark
ant
antacid
anthill
antithesis
apple

The words all begin with 'a' but the 'aa' precedes the 'an' which in turn precedes the 'ap'. The 'an' words all continue to 'ant' so are ordered by the next character, in this case an end of word character, an 'a', a 'h' and an 'i'.

Example: If $A = \{0, 1\}$ then A^4 is the set of binary strings of length 4. Using $0 < 1$ we can totally order A^4.

The correct order is

0000
0001
0010
0011
0100
0101
0110
0111
1000
1001
1010
1011
1100
1101
1110
1111

Definition: The inverse, R^{-1}, of a relation $R \subseteq A \times B$ between a set A and a set B is the relation between B and A given by

$$R^{-1} = \{(b, a) \in B \times A \mid (a, b) \in R\}.$$

Example: If A and B are sets of people and R is the relation 'is a parent of', then R^{-1} is the relation 'is a child of'.

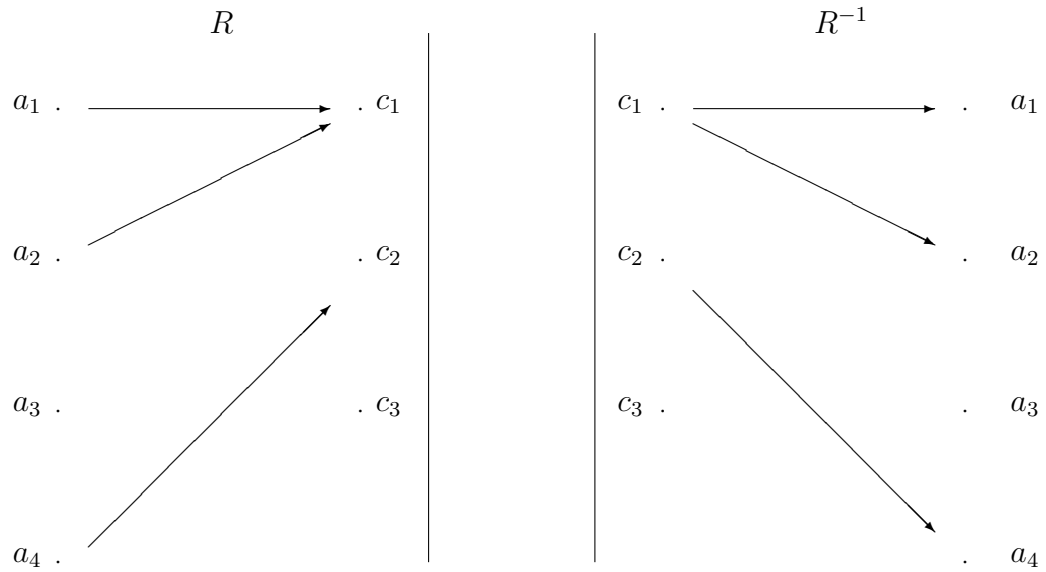
Example: If $A = B$ is a set of positive integers and R is the relation 'is a divisor of', then R^{-1} is the relation 'is a multiple of'.

Note: The digraph of R^{-1} is obtained from the digraph of R by reversing the direction of each edge. The matrix of R^{-1} is obtained from the matrix of R by reflecting it along its diagonal.

Example: For $A = \{a_1, a_2, a_3, a_4\}$ and $C = \{c_1, c_2, c_3\}$ with

$$R = \{(a_1, c_1), (a_2, c_1), (a_4, c_2)\}$$

the digraphs of R and R^{-1} are



and the matrices are

R	c_1	c_2	c_3
a_1	T	F	F
a_2	T	F	F
a_3	F	F	F
a_4	F	T	F

R^{-1}	a_1	a_2	a_3	a_4
c_1	T	T	F	F
c_2	F	F	F	T
c_3	F	F	F	F

Note: Suppose R is a relation on a set A. If R is reflexive, then R^{-1} is also reflexive. (Interchanging the entries of (a, a) gives (a, a).) If R is symmetric, then R^{-1} is the same relation as R. (If $(a, b) \in R$ then $(b, a) \in R$ by symmetry. This means $(a, b) \in R^{-1}$. Similarly $R^{-1} \subseteq R$.) If R is transitive, then R^{-1} is also transitive. (If (a, b) and (b, c) are pairs in R^{-1}, then (b, a) and (c, b) are pairs in R. However R transitive means $(c, a) \in R$ so that $(a, c) \in R^{-1}$.)

Definition: If R is a relation between a set A and a set B and S is a relation between B and a set C then the composition of S with R, written $S \circ R$, is the relation between A and C given by

$$S \circ R = \{(a, c) \in A \times C \mid \text{ for some } b \in B,\ [((a, b) \in R) \textbf{ and } ((b, c) \in S)]\}.$$

Example: If A is a set of men, R is the relation 'is the father of' and S is the relation 'is a brother of', then $S \circ R$ is the relation 'is an uncle of' while $R^2 = R \circ R$ is the relation 'is a grandfather of'.

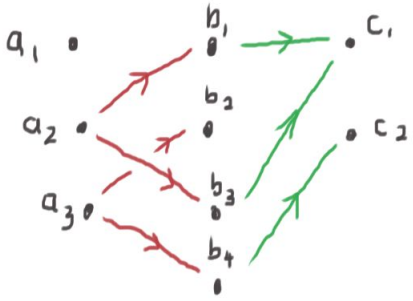
Note: The digraph of the composition of two relations can be read from the digraphs of the relations.

Example: Suppose $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2, b_3, b_4\}$ and $C = \{c_1, c_2\}$ and the relations R and S are given by

$$R = \{(a_2, b_1), (a_2, b_3), (a_3, b_2), (a_3, b_4)\} \subseteq A \times B$$

$$S = \{(b_1, c_1), (b_3, c_1), (b_4, c_2)\} \subseteq B \times C$$

draw the corresponding digraphs, deduce the digraph for $S \circ T$ and express $S \circ T$ as a set of ordered pairs.



For $a_1 \in A$ and $c_1 \in C$ we see no $(a_1, b) \in R$ so $(a_1, c_1) \notin S \circ R$. Similarly, $(a_1, c_2) \notin S \circ R$. For $a_2 \in A$ and $c_1 \in C$ we see $(a_2, b_1) \in R$ and $(b_1, c_1) \in S$, so $(a_2, c_1) \in S \circ R$. (We could also go via b_3.) For $a_2 \in A$ and $c_2 \in C$ we see $(a_2, b_1), (a_2, b_3) \in R$ but $(b_1, c_2) \notin S$ and $(b_3, c_2) \notin S$, so $(a_2, c_2) \notin S \circ R$. For $a_3 \in A$ and $c_1 \in C$ we see $(a_3, b_2), (a_3, b_4) \in R$ but $(b_2, c_1) \notin S$ and $(b_4, c_1) \notin S$, so $(a_3, c_1) \notin S \circ R$. For $a_3 \in A$ and $c_2 \in C$ we see $(a_3, b_4) \in R$ and $(b_4, c_2) \in S$, so $(a_3, c_2) \in S \circ R$. Thus

$$S \circ R = \{(a_2, c_1), (a_3, c_2)\}$$

Note: The matrix of the composition of two relations can be deduced from the matrices of the relations using a logical matrix product. (We will not pursue this.)

Chapter 4: Functions.

Recall the notion of function from school mathematics. Let X and Y be sets. A function from X to Y is a rule that assigns to each element of X exactly one element of Y. Usually we think of X and Y as being subsets of $\mathbb{R}$ and the function as being given by a formula.

Example : Consider the square root function. Even though we have a square root function on the calculator you have to careful with what it means. What is the square root of -1? The calculator gives an error message. If the square root of x is the number whose square is x why is the square root of 4 not -2?

Example : Let f be the function which assigns to the ID of each student registered for MS121 their score in test 1. There is no formula for f. We don't have anything like

$$\text{score} = \sqrt{\sin(\text{ID number}) - 5}.$$

Instead we have a table

ID number	Score
19237961	2
19247961	1
$\vdots$	$\vdots$
19949494	0

which is really just a set of pairs of the form (a, b) where A is the set of IDs of students registered for MS121 and B is the set $\{0, 1, 2, 3, 4\}$. So the function is a relation between A and B. We notice that for each $a \in A$, or each student ID, there has to be one and only one element of B, or score. It is precisely those relations with this further property which we call functions.

Definition: A function f from a set A to a set B is a relation between A and B which satisfies two properties:
(1) every element in A is related to some element in B, and
(2) no element in A is related to more than one element in B.

In other words, given any element $a \in A$, there is a unique element $b \in B$ with $(a, b) \in f$.

Note: The digraph of a relation which is a function has exactly one arrow leaving each point in the set A. The matrix of a relation which is a function has exactly one T in each row.

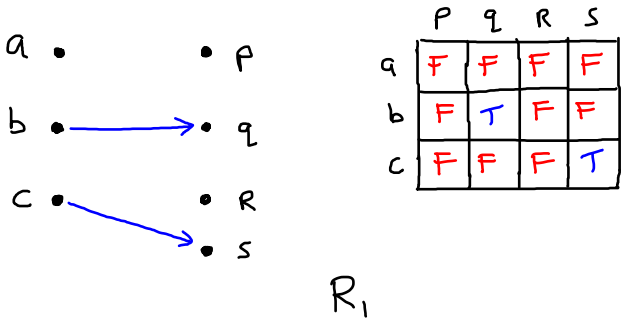
Example: $A = \{a, b, c\}$, $B = \{p, q, r, s\}$ with 3 relations

$$R_1 = \{(b, q), (c, s)\}$$

$$R_2 = \{(a, r), (b, q), (b, s), (c, p)\}$$

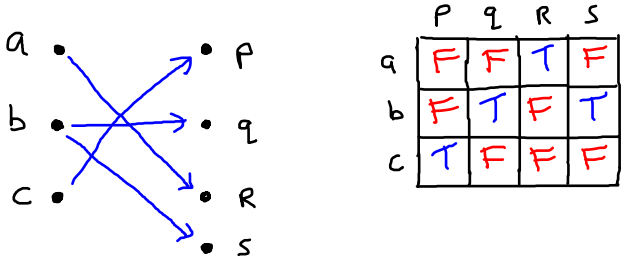
$$R_3 = \{(a, q), (b, q), (c, r)\}$$

Only R_3 is a function.



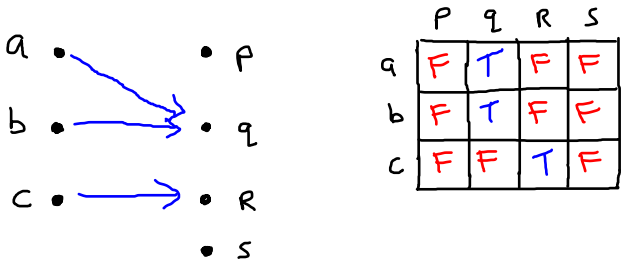
	p	q	R	s
a	F	F	F	F
b	F	T	F	F
c	F	F	F	T

R_1



	p	q	R	s
a	F	F	T	F
b	F	T	F	T
c	T	F	F	F

R_2



	p	q	r	s
a	F	T	F	F
b	F	T	F	F
c	F	F	T	F

R_3

Notation: If a relation f is a function we usually think of it as a rule which assigns to each $a \in A$ the unique element $b \in B$ with $(a, b) \in f$. We often write $b = f(a)$ and call b the image of a under f. The set A is called the domain of the function. The set B is called the codomain of the function. We will use the notation $f : A \to B : a \mapsto f(a)$ as shorthand for:
'f is a function with domain A and codomain B which takes a typical element a in A to the element in B given by $f(a)$.'

Example: If $A = \mathbb{R}$ and $B = \mathbb{R}$, the relation

$$R = \{(x, y) \mid y = \sin(x)\}$$

defines the function $f(x) = \sin(x)$.

Example: If $A = \{0, 1\}^4$ and $B = \{0, 1, 2, 3, 4\}$, the relation

$$S = \{(abcd, n) \mid n = \text{the number of 1's among } a, b, c, d\}$$

defines the function which measures the number of 1's in a binary string of length 4.

Example: If $A = \mathbb{Z}$ and $B = \{0, 1, 2\}$ we can define a function $f : A \to B$ with $f(n)$ equal to the remainder when n is divided by 3. There are similar functions where 3 is replaced by some other number. These are used to construct hashing functions. Sometimes this function is denoted by the % symbol as in $5\%3 = 2$ meaning that when 5 is divided by 3 the remainder is 2.

Example: The set of all functions from the finite set $A = \{a, b, c\}$ to $B = \{0, 1\}$ can be identified with the subsets of A by $f \leftrightarrow C$ where C is

the subset of A consisting of those x with $f(x) = 1$. So writing a function $f : A \to B$ as a binary string $f(a)f(b)f(c)$ we get

$$000 \leftrightarrow \emptyset, \quad 001 \leftrightarrow \{c\}, \quad 010 \leftrightarrow \{b\}, \quad 011 \leftrightarrow \{b, c\},$$

$$100 \leftrightarrow \{a\}, \quad 101 \leftrightarrow \{a, c\}, \quad 110 \leftrightarrow \{a, b\}, \quad 111 \leftrightarrow \{a, b, c\}.$$

Notation: The range of a function f is the set of all images of elements of A under f. That is,

$$\text{Range}(f) = \{b \in B \mid (a, b) \in f \text{ for some } a \in A\}.$$

Note: In terms of the digraph Range(f) is the subset of the codomain which are the endpoints of arrows. In terms of the matrix of the relation Range(f) is the subset of the codomain for which the corresponding columns have at least one T in them.

Examples: $f : \{a, b, c\} \to \{p, q, r, s\}$ given by $f(a) = q$, $f(b) = q$ and $f(c) = r$ has range $\{q, r\}$.
$g : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$ has range $\{x \in \mathbb{R} \mid x \geq 0\}$.
$h : \mathbb{Z} \to \mathbb{Z} : n \to r$ where r the remainder when n is divided by 3, has range $\{0, 1, 2\}$.

Note: When A or B or both are infinite it is not possible to draw a digraph of the relation but a variation on the matrix of the relation called the graph of f is useful when A and B are subsets of $\mathbb{R}$. The graph is given by

$$\text{graph}(f) = \{(x, y) \in \mathbb{R}^2 \mid y = f(x)\}$$

Example: $A = \{-2, -1, 0, 1, 2\}$, $B = \{-1, 0, 1, 2, 3, 4\}$, $f : A \to B : x \mapsto x^2$. Matrix is

f	-1	0	1	2	3	4
-2	F	F	F	F	F	T
-1	F	F	T	F	F	F
0	F	T	F	F	F	F
1	F	F	T	F	F	F
2	F	F	F	F	F	T

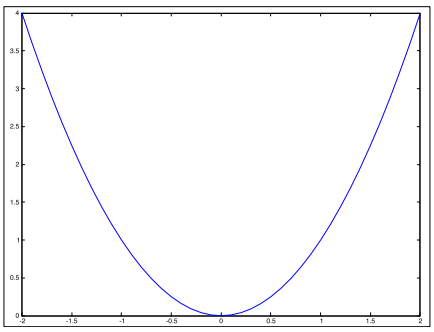
Replace F by blank space and T by $*$:

f	-1	0	1	2	3	4
-2						$*$
-1			$*$			
0		$*$				
1			$*$			
2						$*$

Usually we interchange the axes and get

f	-2	-1	0	1	1
4	$*$				$*$
3					
2					
1		$*$		$*$	
0			$*$		
-1					

When we enlarge A to $\{x \in \mathbb{R} \mid -2 \leq x \leq 2\}$ and B to $\{y \in \mathbb{R} \mid -1 \leq y \leq 4\}$ we get



Note: We recognise the graph of a function by the feature that a vertical line $x = a$ crosses the graph exactly once for each point a in the domain of f.

Example: The relation from $\mathbb{R}$ to $\mathbb{R}$ given by

$$R = \{(x, y) \mid x^2 = y^2\}$$

is not a function. The set of points in R is the pair of lines $y = x$ and $y = -x$ since

$$x^2 = y^2 \Leftrightarrow x^2 - y^2 = 0 \Leftrightarrow (x - y)(x + y) = 0 \Leftrightarrow (y = x) \textbf{ or } (y = -x)$$

A vertical line $x = a$ crosses this set at (a, a) and $(a, -a)$ which are different if $a \neq 0$.

Definition: We say two functions f and g are equal and write $f = g$ if f and g have the same domain and the same codomain and $f(x) = g(x)$ for each x in the domain.

Note: When f and g are obtained by complicated processes knowing that they are equal will be important.

Example: Sometimes we use equality to define new functions, as in

$$\tan(x) = \frac{\sin(x)}{\cos(x)}.$$

Example: Sometimes equality follows from some arithmetical fact, as in $f = g$ for $f(x) = (-2)(x - 2)$ and $g(x) = -2x + 4$.

Note: When looking at the digraph of a function we see a lack of symmetry between the domain and the codomain. We require exactly one arrow leaving each point in the domain but can have several or none ending at a point in the codomain.

Definition: A function f from a set A to a set B is called surjective or onto if $\text{Range}(f) = B$, that is, if $b \in B$ then $b = f(a)$ for at least one $a \in A$.

Note: The digraph of a surjective function will have at least one arrow ending at each element of the codomain.

Definition: A function f from a set A to a set B is called injective or one-to-one if no two elements in A have the same image in B, that is

$$[f(a_1) = f(a_2)] \Rightarrow [a_1 = a_2].$$

Note: The digraph of an injective function will have at most one arrow ending at each element of the codomain.

Note: Can think of one-to-one as 'not two-to-one'.

Definition: A function f from a set A to a set B which is both injective and surjective is called a bijective function or a one-to-one correspondence.

Example: $A = \{a, b, c\}$, $B = \{p, q, r, s\}$ with

$$R_3 = \{(a, q), (b, q), (c, r)\}$$

The corresponding function is neither surjective nor injective. There is no element of A mapping onto p, so not surjective. There are two elements a and b both mapping to q, so not injective.

Example: Suppose $A = \mathbb{Z}$, $B = \{0, 1, 2\}$ and $f : A \to B$ is defined by

$$f(n) = \text{the remainder when } n^2 \text{ is divided by 3.}$$

Is f injective? Is f surjective?

Compute some values and see if there is a pattern:
$n = 0$: Here $n^2 = 0 = 0(3) + 0$ so the remainder after dividing n^2 by 3 is 0.
$n = 1$: Here $n^2 = 1 = 0(3) + 1$ so the remainder after dividing n^2 by 3 is 1.
$n = 2$: Here $n^2 = 4 = 1(3) + 1$ so the remainder after dividing n^2 by 3 is 1.
$n = 3$: Here $n^2 = 9 = 3(3) + 0$ so the remainder after dividing n^2 by 3 is 0.

$n = -1$: Here $n^2 = 1 = 0(3) + 1$ so the remainder after dividing n^2 by 3 is 1.
$n = -2$: Here $n^2 = 4 = 1(3) + 1$ so the remainder after dividing n^2 by 3 is 1.
$n = -3$: Here $n^2 = 9 = 3(3) + 0$ so the remainder after dividing n^2 by 3 is 0.

It looks like $f(n)$ is always 0 or 1.

f is not injective since $f(3) = 0 = f(0)$. Two different integers are taken by f to the same point.

f is not surjective since $f(n) \neq 2$ for any integer n. To see this look at the remainder when n itself is divided by 3. Suppose $n = 3k + r$ where $r \in \{0, 1, 2\}$. Then

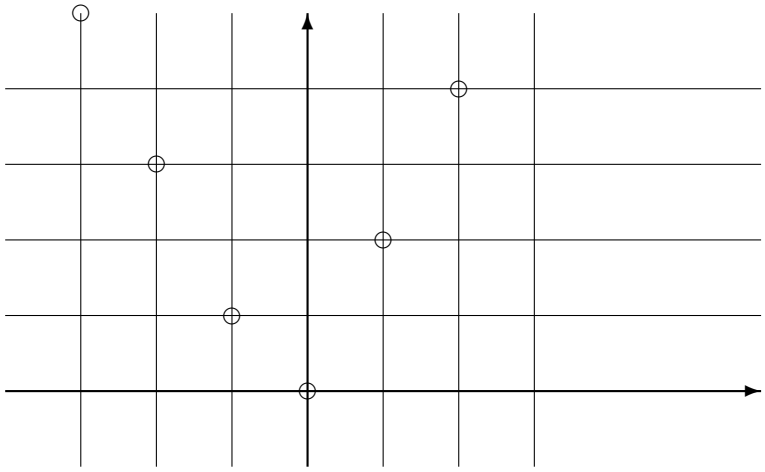
$$n^2 = (3k + r)^2 = 9k^2 + 6kr + r^2.$$

Since the first two terms are multiples of 3, the remainder when n^2 is divided by 3 is the same as the remainder when r^2 is divided by 3. But $r \in \{0, 1, 2\}$ so that $r^2 \in \{0, 1, 4\}$ and the remainder is either 0 or 1.

Example: The function $f : \mathbb{R} \to \mathbb{R} : x \mapsto 3x - 5$ is both injective and surjective and is hence a bijection. To show injectivity suppose $f(a) = f(b)$. Then $3a - 5 = 3b - 5$ and $a = b$. To show surjectivity suppose $y \in \mathbb{R}$ and look for $x \in \mathbb{R}$ with $f(x) = y$. Then $3x - 5 = y$ and we can always solve to get $x = (1/3)(y + 5)$.

Example: A function from $\mathbb{Z} \to \mathbb{Z}_{\geq 0}$ which is both surjective and injective!

$$f(n) = \begin{cases} 2n & \text{if } n \geq 0 \\ -2n - 1 & \text{if } n < 0 \end{cases}$$



We note that $f(n)$ is even if $n \geq 0$ while $f(n)$ is odd if $n < 0$. To establish surjectivity, suppose $m \in \mathbb{Z}_{\geq 0}$. If m is even $m = 2k = f(k)$. If m is odd, $m = 2l - 1 = f(-l)$. To establish injectivity, suppose $f(a) = f(b) = m$. If m is even, then a and b are non-negative with $2a = 2b = m$ so that $a = b$. If m is odd, then a and b are negative with $-2a - 1 = -2b - 1 = m$ so that $a = b$ also.

Definition: A function f from a set A to a set B is called invertible if the inverse relation $f^{-1} \subseteq B \times A$ is a function.

Example: $A = \{a, b, c\}$, $B = \{p, q, r\}$ with

$$f = \{(a, q), (b, r), (c, p)\}.$$

The inverse relation is

$$f^{-1} = \{(p, c), (q, a), (r, b)\}.$$

This is a function from B to A so f is invertible.

Example: Recall the function $f : \mathbb{Z} \to \mathbb{Z}_{\geq 0}$ given by

$$f(n) = \begin{cases} 2n & \text{if } n \geq 0 \\ -2n - 1 & \text{if } n < 0 \end{cases}$$

As a relation

$$f = \{\ldots, (-3, 5), (-2, 3), (-1, 1), (0, 0), (1, 2), (2, 4), (3, 6), \ldots\}.$$

The inverse relation is

$$f^{-1} = \{(0, 0), (1, -1), (2, 1), (3, -2), (4, 2), (5, -3) \ldots\}.$$

This inverse relation is a function so f is invertible.

Theorem: A function $f : A \to B$ is invertible if and only if f is bijective.

Proof: Suppose $f : A \to B$ is invertible. We will show f is bijective. Start with surjectivity. If $b \in B$ then $(b, a) \in f^{-1}$ for some $a \in A$ since f^{-1} is a function (property 1). Thus $(a, b) \in f$ and $f(a) = b$. Next consider injectivity. Suppose $f(a_1) = f(a_2) = b$. Then $(a_1, b) \in f$ and $(a_2, b) \in f$ so that $(b, a_1) \in f^{-1}$ and $(b, a_2) \in f^{-1}$. But f^{-1} is a function so that $a_1 = a_2$ (property 2).

For the converse, suppose $f : A \to B$ is bijective. We will show f is invertible by showing f^{-1} is a function. Let $b \in B$. Since f is surjective $(a, b) \in f$ for some $a \in A$. This gives $(b, a) \in f^{-1}$ and f^{-1} satisfies property 1 of a function. Next suppose $(b, a_1) \in f^{-1}$ and $(b, a_2) \in f^{-1}$. Then $(a_1, b) \in f$ and $(a_2, b) \in f$ so that $f(a_1) = b$ and $f(a_2) = b$. But f is injective, so $a_1 = a_2$ and f^{-1} satisfies property 2 of a function.

Example: For the function $f : \mathbb{Z} \to \mathbb{Z}_{\geq 0}$ given by

$$f(n) = \begin{cases} 2n & \text{if } n \geq 0 \\ -2n - 1 & \text{if } n < 0 \end{cases}$$

we can write a formula for the inverse function $f^{-1} : \mathbb{Z}_{\geq 0} \to \mathbb{Z}$

$$f^{-1}(n) = \begin{cases} n/2 & \text{if } n = 2k \\ -(n+1)/2 & \text{if } n = 2k+1 \end{cases}$$

Example: The function $f : \mathbb{R} \to \mathbb{R} : x \mapsto 3x - 5$ is both injective and surjective and is hence a bijection. By the theorem it is invertible. We can, in this case, construct the inverse function.

$$y = 3x - 5 \Leftrightarrow y + 5 = 3x \Leftrightarrow (y+5)/3 = x$$

so that the inverse function is $g : \mathbb{R} \to \mathbb{R} : y \mapsto (y+5)/3$.

Example: Let $A = B = \{0, 1, 2, 3, 4, 5, 6\}$ and define the function $f : A \to B$ by

$$f(n) = r \text{ where } 3n = 7(q) + r \text{ with } 0 \leq r < 7.$$

So r is the remainder we get when we divide $3n$ by 7. Thus $f(n) = 3n \text{mod} 7$ or $f(n) = 3n\%7$.
This f is both injective and surjective and is hence a bijection.

$$f(0) = 0, f(1) = 3, f(2) = 6, f(3) = 2, f(4) = 5, f(5) = 1, f(6) = 4.$$

By the theorem it is invertible. The inverse is $g : B \to A$ given by

$$g(0) = 0, g(3) = 1, g(6) = 2, g(2) = 3, g(5) = 4, g(1) = 5, g(4) = 6.$$

We can, in this case, check that the inverse has a similar formula

$$g(n) = s \text{ where } 5n = 7(t) + s \text{ with } 0 \leq s < 7.$$

Note: We have seen that a function is invertible if and only if it is bijective. How do we reconcile this with the fact that the function $\sin(x)$ is not injective but $\sin^{-1}(x)$ is defined? For that matter, $f(x) = x^2$ is not injective but its inverse $g(x) = \sqrt{x}$ exists.

The answer to these questions is in the definition of a function, which specifies a domain and a codomain. The function $f : \mathbb{R} \to \mathbb{R} : x \to x^2$ is neither injective (since $(-2)^2 = 4 = 2^2$) nor surjective (since $x^2 \neq -1$) but the function $h : [0, \infty) \to [0, \infty) : x \mapsto x^2$ is bijective and has inverse $k : [0, \infty) \to [0, \infty) : x \mapsto \sqrt{x}$. Here $[0, \infty)$ is the set of non-negative real numbers.

Example : Suppose $f(x) = (2x - 1)/(x + 2)$. What is the natural domain of f? What is the range of f? Show that $f(x)$ is bijective as a function from its natural domain to its range and compute the inverse function.

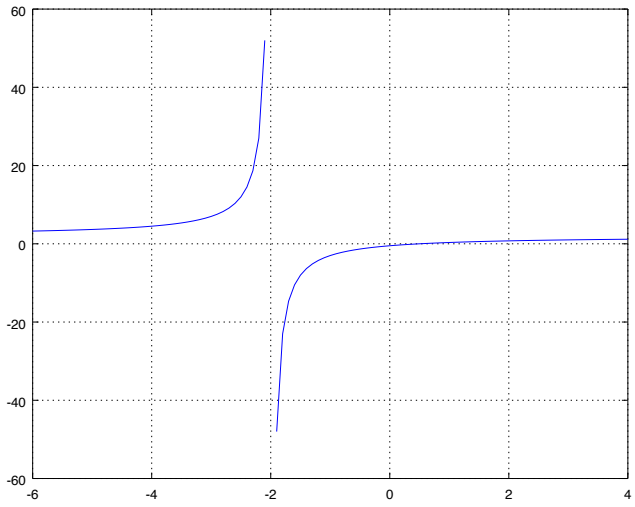
We cannot divide by zero so $x = -2$ is not in the domain. For every other real number x, $f(x)$ is real so the domain of f is $\mathbb{R} \setminus \{-2\}$. To work out the range suppose

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

$$\text{So} \quad yx + 2y = 2x - 1 \quad \text{and} \quad x(y - 2) = -1 - 2y \quad \text{which gives}$$

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

Thus the range is $\mathbb{R} \setminus \{2\}$. For any $y \neq 2$ we can find an x with $f(x) = y$, namely, $x = (-1 - 2y)/y - 2)$ and for $y = 2$ we cannot find any x with $f(x) = y$. Furthermore, f is bijective from $\mathbb{R} \setminus \{-2\}$ to $\mathbb{R} \setminus \{2\}$ and the inverse function is $g(y) = (-1 - 2y)/(y - 2)$.



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

Note: Recall that, if R is a relation between a set A and a set B and S is a relation between B and a set C then the composition of S with R, written $S \circ R$, is the relation between A and C given by

$$S \circ R = \{(a,c) \in A \times C \mid \text{ for some } b \in B,\ [((a,b) \in R) \textbf{ and } ((b,c) \in S)]\}.$$

This reduces to something much simpler in the case where R and S are functions.

Definition: If $f : A \to B$ and $g : B \to C$ are functions, then the composition of g with f, written $g \circ f$, is the function

$$g \circ f : A \to C : a \mapsto g(f(a)).$$

Proposition : The composition of functions is a function.

Proof : Suppose $f : A \to B$ and $g : B \to C$ are functions. Check the two properties for the relation $g \circ f$:
(1) If $a \in A$, then, since f is a function $(a,b) \in f$ for some unique $b \in B$. This b is denoted $f(a)$. Since g is a function $(b,c) \in g$ for some unique $c \in C$. This c is denoted $g(b)$. However, $c = g(b) = g(f(a)) = (g \circ f)(a)$ is an element of C with $(a,c) \in g \circ f$.
(2) Since b is uniquely determined by a and c is uniquely determined by b, c is uniquely determined by a.

Example : Suppose $A = \{a,b,c\}, B = \{p,q,r,s\}$ and $C = \{x,y\}$, with the two functions $f : A \to B$ and $g : B \to C$ defined by

$$f(a) = q, f(b) = r, f(c) = q \quad \text{and} \quad g(p) = y, g(q) = x, g(r) = y, g(s) = x$$

Then $g \circ f : A \to C$ is given by

$$(g \circ f)(a) = g(f(a))) = g(q) = x,$$

$$(g \circ f)(b) = g(f(b)) = g(r) = y,$$

$$(g \circ f)(c) = g(f(c)) = g(q) = x.$$

Example : Suppose $f : \mathbb{R} \to \mathbb{R} : x \mapsto x+1$ and $g : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$. Then

$$(g \circ f) : \mathbb{R} \to \mathbb{R}; x \mapsto g(f(x)) = g(x+1) = (x+1)^2 = x^2 + 2x + 1.$$

However the composition $f \circ g$ is also defined and

$$(f \circ g) : \mathbb{R} \to \mathbb{R}; x \mapsto f(g(x)) = f(x^2) = x^2 + 1.$$

Thus, in general, $g \circ f \neq f \circ g$.

Proposition: If $f : A \to B$, $g : B \to C$ and $h : C \to D$ are any mappings then

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

That is, composition of mappings is always associative.

Proof: We simply calculate the value of each composition on a typical element of the domain A. Let $a \in A$ and set $b = f(a)$, $c = g(b)$ and $d = h(c)$. So $c = (g \circ f)(a)$ and $d = (h \circ g)(b)$. This gives

$$\begin{aligned}
(h \circ (g \circ f))(a) &= h((g \circ f)(a)) \\
&= h(c) \\
&= d \\
&= (h \circ g)(b) \\
&= (h \circ g)(f(a)) \\
&= ((h \circ g) \circ f)(a)
\end{aligned}$$

If $f : A \to B$ is invertible, what is the composition $f^{-1} \circ f : A \to A$? It takes each point of A to itself.

Definition: If A is a non-empty set we define the identity function

$$\mathrm{id}_A : A \to A : a \mapsto a$$

that is, the function which takes each element to itself.

Proposition: (a) If $f : A \to B$ then

$$\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A.$$

(b) If, furthermore, f is invertible, then

$$f^{-1} \circ f = \mathrm{id}_A \text{ and } \quad f \circ f^{-1} = \mathrm{id}_B.$$

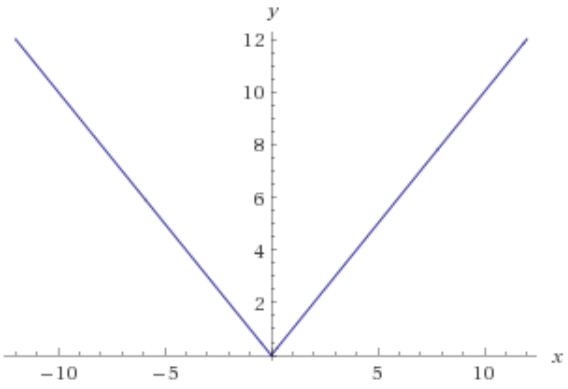
Definition: The absolute value function $|x|$ has domain $\mathbb{R}$ and codomain $\mathbb{R}$ and is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Examples : $|45| = 45$, $|-3| = 3$.

Note : $|x| = \sqrt{x^2}$.

Note : $|x|$ measures the distance on the real line from x to 0. $|x - y|$ measures the distance on the real line from x to y.



$$f(x) = |x|$$

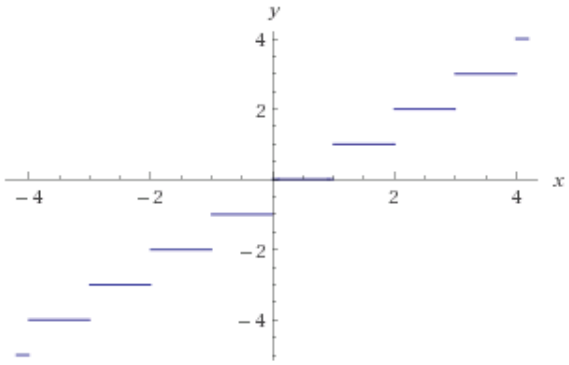
Example : Solve $|x+2|=3$. Since $|y|=3$ implies $y=3$ or $y=-3$, we get $x+2=3$ or $x+2=-3$. This is equivalent to $x=1$ or $x=-5$. This is what we expect: 1 and -5 are the two numbers which are distance 3 from -2. ($|x+2|=|x-(-2)|$.)

Definition: The floor function $\lfloor x \rfloor$ has domain $\mathbb{R}$ and codomain $\mathbb{Z}$ and is defined by

$$\lfloor x \rfloor = \text{largest integer} \leq x$$

Examples : $\lfloor 5.7 \rfloor = 5$, $\lfloor -3.4 \rfloor = -4$.

Note : $\lfloor x \rfloor$ rounds x down to the nearest integer. What does $(1/10)\lfloor 10x \rfloor$ do? (It rounds down to first decimal place.)



$$f(x) = \lfloor x \rfloor$$

Pigeonhole principle

It is intuitively clear that, when we try to put more than n objects (pigeons) into n containers (pigeonholes) then there will be at least one container with more than one object.

Proposition: Suppose A and B are finite sets and $f : A \to B$ is injective. Then $|B| \geq |A|$.

Proof: Let the elements of A be $a_1, a_2, \ldots, a_n$. Since f is injective the elements $f(a_1), f(a_2), \ldots, f(a_n)$ are distinct in B. So B contains at least $|A|$ elements.

Corollary: (Pigeonhole principle) Suppose A and B are finite sets with $|B| < |A|$ and $f : A \to B$ is a function. Then f cannot be injective.

Example: An extended family of 14 people are gathered for a function. Show that at least two have birthdays in the same month of the year.

Here the set A is the set of family members, the set B is the set of months and the function $f : A \to B$ takes each family member to his/her birth month. Since $|A| = 14 > 12 = |B|$, f cannot be injective and two members have the same birth month.

Example: How big does a crowd have to be to ensure that two people from it share a birthday?

Here the set A is the set of people in the crowd, the set B is the set of days in the year and the function $f : A \to B$ takes each person to his/her birth day. Since we want $|A| > 366 = |B|$ (possible leap year), f will not be injective and two people have the same birthday provided $|A| > 366$.

Example: How many different surnames must be in a telephone directory in order to ensure that at least two surnames have the same first and the same second letter?

Here the set A is the set of names in the directory, L is the set of letters of the alphabet, the set B is $L \times L$, the set of pairs of letters of the alphabet and the function $f : A \to B$ takes each surname to the pair (first letter, second letter). Since we want $|A| > 26^2 = |B|$, f will not be injective and two surnames have the same first and second letter provided $|A| > 676$.

Example: What is the smallest n which has the property that among any n numbers chosen from

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

there will be a pair which sums to 11?

This example is more subtle. Here the set A is the set of numbers in a set of size n, the set B is the set of pairs of numbers which sum to 11

$$B = \{(1, 10), (2, 9), (3, 8), (4, 7), (5, 6)\}$$

and the function $f : A \to B$ takes each number to the pair that contains it.

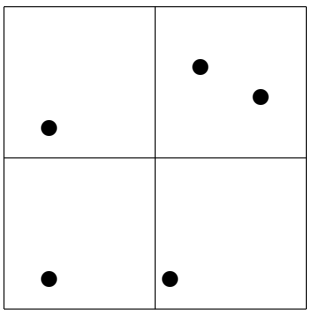
$$1 \mapsto (1, 10), 2 \mapsto (2, 9), 3 \mapsto (3, 8), 4 \mapsto (4, 7), 5 \mapsto (5, 6),$$

$$6 \mapsto (5, 6), 7 \mapsto (4, 7), 8 \mapsto (3, 8), 9 \mapsto (2, 9), 10 \mapsto (1, 10)$$

Since we want $|A| = n > 5 = |B|$, f will not be injective and two numbers will sum to 11 provided $n \geq 6$.

Example: Five points are chosen on the unit square. Show that two are within a distance of $\sqrt{2}/2$ of each other.

Split the square into four squares of sidelength $1/2$. Here the set A is the set of five points, the set B is the set of four small squares and the function $f : A \to B$ takes each point to one of the small squares containing it. (It may lie on the boundary of two or more small squares.) Since $|A| = 5 > 4 = |B|$, f will not be injective and two points lie in the same small square and are within a distance of $\sqrt{2}/2$ of each other.



Example: Each of the 9 squares on a 3×3 grid contains one of the integers $-1, 0$ or 1. Show that among the 8 resulting sums (three rows, three columns and two diagonals) there will always be two that add to the same number.

This example is again more subtle. Here the set A is the set of 8 triples from such a grid, the set B is the set of possible sums of three numbers from the set $\{-1, 0, 1\}$

$$B = \{-3, -2, -1, 0, 1, 2, 3\}$$

and the function $f : A \to B$ takes each triple to its sum. Since $|A| = 8 > 7 = |B|$, f will not be injective and two triples will sum to the same number.

Extension of pigeonhole principle: Suppose A and B are finite sets with $k|B| < |A|$ and $f : A \to B$ is a function. Then at least one element of B must be the image of at least $k + 1$ elements of A under f.

Example: How many different surnames must be in a telephone directory in order to ensure that at least four surnames have the same first and the same second letter?

Here we need at least $3(26)^2 + 1$ surnames.

Example: Show that, in any group of six people, there are either three who all know each other or three complete strangers.

Let x be any one of the people and set A equal to the set consisting of the other 5. Set $B = \{0, 1\}$ and define $f : A \to B$ by

$$f(y) = \begin{cases} 1 & \text{if } x \text{ knows } y \\ 0 & \text{if } x \text{ does not know } y \end{cases}$$

Since $|A| = 5 > 4 = 2|B|$ at least one of the numbers 0 or 1 arises three times as $f(y)$. So either x knows three of the others or x does not know three of the others. In the first case, suppose x knows y_1, y_2 and y_3. If any two of these know each other then we can add x to this pair to get a set of three who all know each other. If no two of these know each other then we have a set of three complete strangers. The second case is similar.

Note: If A and B are disjoint sets with n and m elements respectively then $A \cup B$ has $n + m$ elements. This follows from our rule

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

Saying A and B are disjoint means $A \cap B = \emptyset$ so that $|A \cap B| = 0$.

Addition Principle of Counting: If an event E_1 can occur in n different ways and an event E_2 can occur in m different ways and both events cannot occur at the same time then E_1 or E_2 can occur in $n + m$ different ways.

Rule of thumb: 'or' $\leftrightarrow$ addition

Example: If a phone shop stocks 5 models of phone of brand 1 and 6 models of phone of brand 2, then it stocks 11 different models of phone.

Example: There are 33 numbers between 1 and 100 inclusive which are multiples of three and 50 which are multiples of 2. There are not 83 numbers in this range which are multiples of 3 or multiples of 2. Here A, the set of multiples of 3, and B, the set of multiples of 2, are not disjoint since $A \cap B$ contains all the multiples of 6.

Note: Of course, this principle can be extended to more than 2 events: If event E_i can occur in n_i different ways for $i = 1, 2, \ldots, k$ and no two of the events can occur at the same time then E_1 or E_2 or $\ldots$ or E_k can occur in $n_1 + \ldots + n_k$ different ways.

Note: If E is the compound event that exactly one of E_1 or E_2 or $\ldots$ or E_k occurs, then E is partitioned by the sets $E_1, E_2, \ldots, E_k$.

Example: Binary strings of length at least 4 consist of strings of length 0, 1, 2, 3 or 4 and a string cannot have two different lengths. If B_i is the set of binary strings of length i then the total number is

$$|B_0| + |B_1| + |B_2| + |B_3| + |B_4| = 1 + 2 + 4 + 8 + 16 = 31.$$

Example: The number of integers between 100 and 299 which are divisible by 5 can be expressed as the disjoint union of 4 sets:

$$S_1 = \{\text{strings of form [1][k][0]}\}$$

$$S_2 = \{\text{strings of form [1][k][5]}\}$$
$$S_3 = \{\text{strings of form [2][k][0]}\}$$
$$S_4 = \{\text{strings of form [2][k][5]}\}$$

Each S_i has 10 elements (10 choices for k). Now the number we want is

$$n = |S_1| + |S_2| + |S_3| + |S_4| = 10 + 10 + 10 + 10 = 40.$$

Difference Rule: If A is a finite set and B is a subset of A then

$$|A - B| = |A| - |B|.$$

Proof: The sets $B, A - B$ partition A. So $|A| = |B| + |A - B|$.

Example: How many binary strings of length 5 are not divisible by 4?

A string $abcde$ is divisible by 4 if and only if $d = e = 0$. (Remember that $abcde = a2^4 + b2^3 + c2^2 + d2 + e = 2^2(a2^2 + b2 + c) + d2 + e$.) So a string of length 5 which is divisible by 4, $abc00$, is a string of length 3 followed by two 0's. The number of strings divisible by 4 is $2^3 = 8$ and the number not divisible by 4 is

$$2^5 - 2^3 = 32 - 8 = 24.$$

Here A is the set of strings of length 5. B is the set of strings of length 5 which are divisible by 4. We want $|A - B|$.

Note: If A and B are sets with n and m elements respectively then their Cartesian product $A \times B$ has nm elements.

Product Principle of Counting: If event E_1 can occur in n ways and event E_2 can occur in m ways, then the number of ways the events can occur in the order E_1 followed by E_2 is nm.

Rule of thumb: 'and' $\leftrightarrow$ multiplication

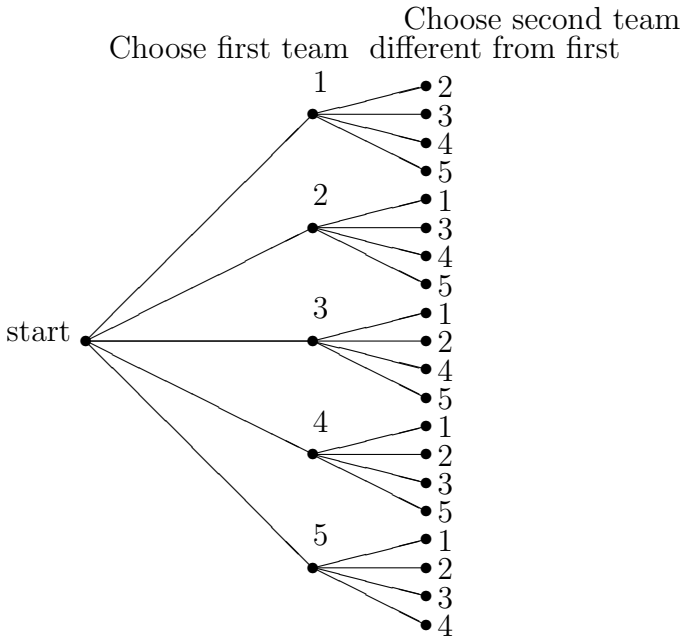
Note: Of course, this principle can be extended to more than 2 events: If event E_i can occur in n_i different ways for $i = 1, 2, \ldots, k$ then E_1 followed by E_2 followed by $\ldots$ followed by E_k can occur in $n_1 n_2 \ldots n_k$ different ways.

Example: How many license plates are there with three letters followed by three digits? The first digit cannot be zero. Answer: (26)(26)(26)(9)(10)(10). E_1 is choice of first letter, E_2 is choice of second letter, E_3 is choice of third letter, E_4 is choice of first non-zero digit, E_5 is choice of second digit, E_6 is choice of last digit.

Example: How many ways can the first three places be filled in a league table if the league has 20 teams? Answer: (20)(19)(18). E_1 is choice of first team, E_2 is choice of second team different from first, E_3 is choice of third team different from first and second.

Note: It may help to view each element as the result of applying a number of choices and visualising the process as a tree.

Example: If the league has 5 teams and we are concerned with the first two places then the tree is



Note: Just as there is a subtraction rule, we sometimes count using division.

Example: If the top three teams in a league of twenty go into a superleague, in how many ways can this happen? Answer: (20)(19)(18)/(3)(2)(1). We have 20 choices for first team, 19 for second and 18 for third. However each set of three teams $\{a, b, c\}$ has been counted 6 times as

$$abc, acb, bac, bca, cab, cba$$

with 3 choices for the first of the three, 2 choices for the second and 1 choice for the last of the three.

Note: We are seeing products of consecutive integers in decreasing order such as (20)(19)(18) arising in counting problems.

Definition: A permutation of a finite set of objects is an ordering of the objects in a row.

Example: Above we saw the 6 permutations of three objects

$$abc, acb, bac, bca, cab, cba.$$

Example: There are 24 permutations of four objects

$$abcd, abdc, acbd, acdb, adbc, adcb,$$

$$bacd, badc, bcad, bcda, bdac, bdca,$$

$$cabd, cadb, cbad, cbda, cdab, cdba,$$

$$dabc, dacb, dbac, dbca, dcab, dcba,$$

These are found by picking one of the elements to be first in 4 ways and following it with one of the 6 permutations of the remaining three. Note that the last row is obtained from the previous example by putting a d in front of the 6 permutations of $\{a, b, c\}$.

Definition: If n is a positive integer we define n factorial, written $n!$ to be the product of the first n positive integers.

$$n! = n(n-1)(n-2)\dots(3)(2)(1)$$

Note: By convention, we agree that $0! = 1$.

Note: n factorial can be defined inductively by

$$1! = 1 \quad \text{and} \quad n! = n[(n-1)!]$$

Example: The number of permutations of 5 objects is $5! = 120$.

Example: In how many ways can a group of five world leaders be arranged in a row for a photoshoot? In how many of these ways will president A be beside president B?

In the first case the answer is $5! = 120$. In the second case, we treat $\{A, B\}$ as a unit, permute the four sets $\{A, B\}, \{C\}, \{D\}, \{E\}$ in 4! ways and permute

the set $\{A, B\}$ in 2! ways to give a total, by the multiplication rule of $4!2! = 24(2) = 48$.

Example: In how many ways can a group of five world leaders be arranged around a circular table where two arrangements are the same if one is obtained from the other by rotation?

Here we take any circular arrangement and rotate it so that president A is in a fixed place, say the north end, on the table. Now arrange the others to their right in $4! = 24$ ways.

Example: The number of permutations of 3 objects out of 20 is (20)(19)(18), which we can write as (20)!/(17)!.

Definition: If n and r are positive integers with $r \leq n$ then we define the quantity ${}^nP_r$ by

$$ {}^nP_r = \frac{n!}{(n-r)!} $$

This is the number of r-permutations from a set of size n.

Note: In the case $r = n$ we get $n!$ if we agree $0! = 1$.

Example: How many 4-digit PINs are there without repeated digits?

Here the number is ${}^{10}P_4 = 5040$.

Example: In how many ways can we arrange 3 people from a group of 7 in a row? In how many of these arrangements will a particular person from the seven be the first person in the row?

In the first case we have just ${}^7P_3 = (7)(6)(5)$. In the second case, the first person is fixed and we just have a 2-permutation from 6 people to fill the second and third places, giving ${}^6P_2 = (6)(5)$. Note that this is 1/7 of the answer for the first part.

Example: The number of 3 element subsets of 20 objects is (20)(19)(18)/(3)(2)(1), which we can write as (20)!/(17)!(3)!.

Definition: If n and r are positive integers with $r \leq n$ then we define the binomial coefficient ${}^nC_r$ or $\left(\begin{array}{c} n \\ r \end{array} \right)$ by

$$ \left(\begin{array}{c} n \\ r \end{array} \right) = \frac{n!}{r!(n-r)!} $$

Example: For $n = 6$ these numbers are

$$\begin{aligned}
\binom{6}{0} &= \frac{6!}{0!(6-0)!} = \frac{6!}{(1)6!} = 1 \\
\binom{6}{1} &= \frac{6!}{1!(6-1)!} = \frac{6!}{1!5!} = 6 \\
\binom{6}{2} &= \frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = \frac{6\times 5}{2\times 1} = 15 \\
\binom{6}{3} &= \frac{6!}{3!(6-3)!} = \frac{6!}{3!3!} = \frac{6\times 5\times 4}{3\times 2\times 1} = 20 \\
\binom{6}{4} &= \frac{6!}{4!(6-4)!} = \frac{6!}{4!2!} = \frac{6\times 5}{2\times 1} = 15 \\
\binom{6}{5} &= \frac{6!}{5!(6-5)!} = \frac{6!}{5!1!} = 6 \\
\binom{6}{6} &= \frac{6!}{6!(6-6)!} = \frac{6!}{6!(1)} = 1
\end{aligned}$$

In each case we are cancelling the largest factorial on the bottom with the corresponding part of 6!.

Proposition: The number of r element subsets of a set with $n \geq r$ elements is $\left(\begin{array}{c} n \\ r \end{array} \right)$.

Proof: Each such subset arises when we pick a first element followed by a second element up to an rth element. The number of such choices is ${}^nP_r$. But this process counts each subset $r!$ times, one for each permutation of the subset.

Example: A committee of 4 is to be chosen from a group of 10 people. In how many ways can this be done? If there are 6 men and 4 women in the group, how many of the possible committees will have 2 men and 2 women? No women? Suppose two of the group refuse to serve on a committee together. How many committees are now possible?

The first number is $\left(\begin{array}{c} 10 \\ 4 \end{array} \right) = (10)(9)(8)(7)/(4)(3)(2) = (10)(3)(7) = 210.$ The second situation involves choosing 2 of the 6 men **and** 2 of the four women. The product principle applies to give the number as

$$\left(\begin{array}{c} 6 \\ 2 \end{array} \right) \left(\begin{array}{c} 4 \\ 2 \end{array} \right) = \frac{(6)(5)}{(2)(1)} \frac{(4)(3)}{(2)(1)} = (15)(6) = 90.$$

If there are to be no women on the committee we must choose 4 from the 6 men giving $\left(\begin{array}{c} 6 \\ 4 \end{array} \right) = (6)(5)(4)(3)/(4)(3)(2)(1) = 15.$ In the last part it is easier to count the number of committees with the disagreeable two and apply the subtraction principle to get

$$\text{number} = \left(\begin{array}{c} 10 \\ 4 \end{array} \right) - \left(\begin{array}{c} 2 \\ 2 \end{array} \right) \left(\begin{array}{c} 8 \\ 2 \end{array} \right) = 210 - 28 = 182.$$

Here the number of committees including the difficult pair are chosen by picking those 2 in 1 way and picking another 2 from the remaining 8.

Theorem: (Binomial Theorem) The coefficient of $a^r b^{n-r}$ in the expansion of $(a+b)^n$ is $\left(\begin{array}{c} n \\ r \end{array} \right)$.

Example: For $n = 6$ we get

$$(a+b)^6 = (1)a^6 + (6)a^5 b + (15)a^4 b^2 + (20)a^3 b^3 + (15)a^2 b^4 + (6)ab^5 + (1)b^6$$

For example the 15 terms involving a^4b^2 are

$$aaaabb, aaabab, aaabba, aabaab, aababa,$$

$$aabbaa, abaaab, abaaba, ababaa, abbaaa,$$

$$baaaab, baaaba, baabaa, babaaa, bbaaaa,$$

one for each choice of 4 places from 6 for the a's.

Proof: (of Binomial Theorem) Expand $(a+b)^n$ into 2^n terms where we keep track of the order the terms, that is, do not replace ba by ab. Now collect terms. The number contributing to $a^r b^{n-r}$ is equal to the number of ways of picking r places from n for the a's, and thus is equal to $\begin{pmatrix} n \\ r \end{pmatrix}$.

Example: The coefficient of x^5 in the expansion of $(1+x)^8$ is $\begin{pmatrix} 8 \\ 5 \end{pmatrix} = 56$.

Example: The coefficient of x^5 in the expansion of $(2+x)^8$ is $2^3 \begin{pmatrix} 8 \\ 5 \end{pmatrix} = 448$.

Pascal's Triangle. Pascal's Triangle is the name given to an arrangement of all the binomial coefficients in a triangular pattern with the numbers $\begin{pmatrix} n \\ r \end{pmatrix}$ centred on the nth row in order of increasing r from left to right.

$$\binom{0}{0}$$

$$\binom{1}{0}\ \binom{1}{1}$$

$$\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}$$

$$\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}$$

$$\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4}$$

$$\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$1$$

$$1 \quad 1$$

$$1 \quad 2 \quad 1$$

$$1 \quad 3 \quad 3 \quad 1$$

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$

$$\vdots \qquad \vdots \qquad \vdots$$

Proposition: The following identities hold

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \left(\begin{array}{c} n \\ n-r \end{array} \right) \quad , \quad \left(\begin{array}{c} n+1 \\ r \end{array} \right) = \left(\begin{array}{c} n \\ r \end{array} \right) + \left(\begin{array}{c} n \\ r-1 \end{array} \right)$$

Proof: For the first identity, using the interpretation of $\left(\begin{array}{c} n \\ r \end{array} \right)$ as the number of r element subsets of a set A of size n, note that each choice of an r element subset B of A automatically identifies a $(n-r)$-element subset $A - B$.

For the second identity, again using the interpretation of $\left(\begin{array}{c} n+1 \\ r \end{array} \right)$ as the number of r element subsets of a set A of size $n+1$, we can single out an element a_{n+1} of A and partition A into

$$A_1 = \{a_1, \ldots, a_n\}, \quad A_2 = \{a_{n+1}\}.$$

By the addition rule a subset of size r from A either contains a_{n+1} or not and by the multiplication rule the total number is

$$\left(\begin{array}{c} n \\ r-1 \end{array} \right) \left(\begin{array}{c} 1 \\ 1 \end{array} \right) + \left(\begin{array}{c} n \\ r \end{array} \right) \left(\begin{array}{c} 1 \\ 0 \end{array} \right).$$

(Pick $r-1$ from A_1 **and** 1 from A_2) **or** (Pick r from A_1 **and** 0 from A_2)

Note: We have seen that ${}^nP_r$ is the number of permutations of n objects taken r at a time, while ${}^nC_r$ is the number of subsets of size r from n. The difference between ${}^nP_r$ and ${}^nC_r$ is that order is important in the first case but not in the second. We have also seen that n^r is the number of ways of choosing r objects from n if order is important and each chosen object is replaced before the next choice is made. This leaves one other case, where we select r objects from n with replacement but order is not important.

Example: We have seen that there are $10^4 = 10,000$ four-digit strings, ${}^{10}P_4 = 5040$ of these have no repeated digits and there are ${}^{10}C_4 = 210$ 4-element subsets of a set of size 10. The final computation is the number of distributions of digits occurring among all four-digit strings. We could list the possibilities as

$$0000, 0001, \ldots, 9998, 9999$$

where occurrences of 0 are put on the left followed by occurrences of 1, etc. However, this will take a long time and there is an easier way.

Definition: An r-selection from n is an unordered selection of r objects from n with repetition allowed.

Theorem: The number of r-selections from n is

$$\left(\begin{array}{c} r+n-1 \\ n-1 \end{array} \right)$$

Proof: Start by ordering the types of elements from 1 to n. For each r-selection arrange the elements of the selection so that type 1 elements appear first, type 2 elements appear next, etc. Between each type of element in the selection put a separating marker of the form $\ldots xxx|yy\ldots$ including extra markers for types unrepresented in the selection:

$$\ldots xxx||zz\ldots$$

The result is a string of length $r+n-1$ since there are r elements and we need $n-1$ markers to separate the n types. Therefore an r-selection can be identified with a choice of $n-1$ places for the markers in a string of length $r+n-1$.

Example: Suppose we want to count the number of distributions of digits occurring among all four-digit strings. Here $n = 10$ and $r = 4$ and the number is

$$\left(\begin{array}{c} 4+10-1 \\ 10-1 \end{array} \right) = \left(\begin{array}{c} 13 \\ 9 \end{array} \right) = \left(\begin{array}{c} 13 \\ 4 \end{array} \right) = 715.$$

Example: If 5 cards chosen from a standard deck of 52, the number of different distributions of hearts $\heartsuit$, diamonds $\diamondsuit$, spades $\spadesuit$ and clubs $\clubsuit$ in such a hand is the number of 5-selections from 4 objects and is thus

$$\left(\begin{array}{c} 4+5-1 \\ 4-1 \end{array} \right) = \left(\begin{array}{c} 8 \\ 3 \end{array} \right) = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56.$$

Example: 3 dice are thrown. How many distributions of the numbers $1, 2, 3, 4, 5, 6$ are possible?

We are selecting 3 unordered things from 6 with repetition so the number is

$$\left(\begin{array}{c} 3+6-1 \\ 6-1 \end{array} \right) = \left(\begin{array}{c} 8 \\ 5 \end{array} \right) = \left(\begin{array}{c} 8 \\ 3 \end{array} \right) = 56.$$

Example: A program consists of three nested loops of form

while $1 \leq i \leq 4$
    while $1 \leq j \leq i$
        while $1 \leq k \leq j$
            Some commands
        $k \to k+1$
    $j \to j+1$
$i \to i+1$

How many times is the inner loop iterated? During each iteration the variables i, j and k take values in $\{1,2,3,4\}$ but can coincide. The triple (k,j,i) is a selection of 3 from $\{1,2,3,4\}$ with repetition allowed. The number is

$$\left(\begin{array}{c} 3+4-1 \\ 4-1 \end{array} \right) = \left(\begin{array}{c} 6 \\ 3 \end{array} \right) = 20.$$

Note: We summarise the formulae for the number of ways of choosing r objects from n, when order is important or not and repetition is allowed or not.

	ordered	unordered
repeated	n^r r-samples	$\left(\begin{array}{c} r+n-1 \\ n-1 \end{array} \right)$ r-selections
unrepeated	$\frac{n!}{(n-r)!}$ r-permutations	$\left(\begin{array}{c} n \\ r \end{array} \right)$ r-combinations

Multinomial coefficients: Our final concept in counting will be multinomial coefficients. We will introduce this idea using three examples, which all compute the same number.

Example: How many five-letter words can be formed using the letters of NAVAN?

NNAAV, NNAVA, NNVAA, NANAV, NANVA, NAANV, NAAVN, NAVNA, NAVAN, NVNAA, NVANA, NVAAN, ANNAV, ANNVA, ANANV, ANAVN, ANVNA, ANVAN, AANNV, AANVN, AAVNN, AVNNA, AVNAN, AVANN, VNNAA, VNANA, VNAAN, VANNA, VANAN, VAANN. (30 in total)

These can be counted by looking at possibilities for first letter, then second, etc.,

However it is easier to first distinguish the two N's and the two A's to give 5! words and then note that these words are in groups of four which all read the same if the two N's and the two A's are undistinguished.

Theorem: The number of arrangements of $n = n_1 + \ldots + n_r$ objects of which n_1 are identical, n_2 are identical, $\ldots$, n_r are identical is

$$\frac{n!}{n_1!n_2!\ldots n_r!}$$

Example: In how many ways can a group of 5 people be partitioned into 3 ordered sets of sizes 2, 2 and 1?

Break the problem into 3 stages and use the product principle. First choose 2 from 5 for the first subset, then 2 from 3 for the second and finally 1 from 1 for the last. The total is

$$\left(\begin{array}{c} 5 \\ 2 \end{array} \right) \left(\begin{array}{c} 3 \\ 2 \end{array} \right) \left(\begin{array}{c} 1 \\ 1 \end{array} \right) = \frac{5!}{2!3!} \frac{3!}{2!1!} \frac{1!}{1!0!} = \frac{5!}{2!2!1!}$$

Note the cancelling of the right hand term below the line with the term above the line of the next factor.

Note: The connection between this an the first problem is that each five-letter word formed using the letters of NAVAN is determined by a choice of 2 places from 5 for the N's, a choice of 2 places from the remaining 3 places for the A's leaving 1 place for the single V.

Theorem: The number of ways of partitioning a set of $n = n_1 + \ldots + n_r$ objects into ordered subsets of sizes $n_1, n_2, \ldots, n_r$ is

$$\frac{n!}{n_1!n_2!\ldots n_r!}$$

Example: What is the coefficient of x^2y^2z in $(x + y + z)^5$?

Expanding while keeping track of the order of letters will yield 3^5 terms. The ones contributing to x^2y^2z will have 2 x's, 2 y's and 1 z. The number will be the number of ways of partitioning 5 into ordered subsets of sizes 2, 2 and 1 and is thus

$$\frac{5!}{2!2!1!}$$

Theorem: If $n = n_1 + n_2 + \ldots + n_r$ then the coefficient of $x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$ in $(x_1 + x_2 + \ldots + x_r)^n$ is

$$\frac{n!}{n_1! n_2! \ldots n_r!}$$

Definition: Given positive integers $n_1, n_2, \ldots, n_r$ and $n = n_1 + n_2 + \ldots + n_r$, we define the corresponding multinomial coefficient to be

$$\left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_r \end{array} \right) = \frac{n!}{n_1! n_2! \ldots n_r!}$$

Note: When $r = 2$ we have the usual binomial coefficients.
If $1 \leq s < n$ then set $t = n - s$ and write

$$\left(\begin{array}{c} n \\ s \end{array} \right) = \frac{n!}{s!(n-s)!} = \frac{(s+t)!}{s!t!} = \left(\begin{array}{c} s+t \\ s, t \end{array} \right)$$

Note: A multinomial coefficient can be interpreted as a coefficient in an expansion, the number of ordered partitions of a finite set or the number of rearrangements of a string of letters with repetitions.

Example: How many different ways can a group of 12 students be arranged into three teams of 4 if

(a) the teams are ordered, and

(b) the teams are not ordered.

For (a) the number is

$$\left(\begin{array}{c} 12 \\ 4,4,4 \end{array} \right) = \frac{12!}{4!4!4!} = 34,650$$

For (b), we can permute the 3 teams and the number is

$$\frac{1}{3!}\left(\begin{array}{c} 12 \\ 4,4,4 \end{array} \right) = \frac{12!}{3!4!4!4!} = 5,775.$$

Example: In how many ways can 52 cards be dealt out evenly among 4 people?

Here the number is

$$\left(\begin{array}{c} 52 \\ 13,13,13,13 \end{array} \right) = \frac{52!}{13!13!13!13!} = 5.364 \times 10^{28}$$

Example: How many eleven-letter words can be formed using the letters of MISSISSIPPI?

Here the number is

$$\left(\begin{array}{c} 11 \\ 1,4,4,2 \end{array} \right) = \frac{11!}{1!4!4!2!} = 34,650$$

Past exam question on counting

QUESTION 3

(a) A committee of 3 is to be chosen from a group of 9 politicians.
(i) In how many ways can this be done?
(ii) Suppose two of the group belong to Party I, three of the group belong to Party II and four of the group belong to Party III. How many of the committees have at least two members from the same party? Explain your answer.

[8 marks]

(i) Here the answer is the number of ways of choosing 3 people from 9. We are picking without replacement and the order is unimportant. Thus the number is

$$\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 3 \times 4 \times 7 = 84.$$

(ii) Here we should use the subtraction rule and count the number with one from each party which is

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = 2 \times 3 \times 4 = 24$$

Thus the answer to (ii) is $84 - 24 = 60$.

Incidently, we can count the number in (ii) directly but it takes a lot longer. The number of 3-selections from 3

$$\begin{pmatrix} 3 + (3-1) \\ 3 - 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} = 10.$$

In stars and bars notation these are

$$***||\qquad **|*|\qquad **||*\qquad *|**|\qquad *|*|*$$

$$*||**\qquad |***|\qquad |**|*\qquad |*|**\qquad ||***$$

translating into 'three from Party I', 'two from Party I, one from Party II', etc. Of these, the first is excluded since Party I has only two members and

the fourth is excluded since that does not have at least two in the same party. The counts for the remaining eight possibilities sum to

$$\begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

$$+\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

$$+\begin{pmatrix} 2 \\ 0 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}\begin{pmatrix} 3 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

$$= (1 \times 3 \times 1) + (1 \times 1 \times 4) + (2 \times 3 \times 1) + (2 \times 1 \times 6)$$

$$+(1 \times 1 \times 1) + (1 \times 3 \times 4) + (1 \times 3 \times 6) + (1 \times 1 \times 4)$$

$$= 3 + 4 + 6 + 12 + 1 + 12 + 18 + 4 = 60.$$

Probability

Probability theory is the mathematical study of chance. As a branch of mathematics, but it is perhaps unique in that it deals with questions that we encounter in everyday life.

Examples: (i) What is the probability of winning the jackpot in the game of lotto?
(ii) In a class of 42 students, what is the probability that two or more have a common birthday?
(iii) What is the probability that a hand of poker contains 4 cards of the same type?

We will learn shortly how to solve these problems and many more related ones. It will also be useful to consider simpler problems involving tossing coins and throwing dice, if only to illustrate the definitions.

At the other end of the spectrum, probability theory is also capable of handling much more complicated questions, like calculating the probability that a corporate stock will fall below a certain price at any stage during the next year.

Because probability theory is dealing with relatively concrete problems, we can have an intuitive idea as to what the answer should be. This can be useful because it can help us find the answer. However it can also be misleading because our intuition is often wrong.

Example: (The Monty Hall problem) In a game show a contestant is told that a prize is behind one of three doors. After the contestant picks one door the host (Monty) opens another door revealing no prize. The contestant is invited to switch his/her choice to the remaining door or not. What should he/she do?

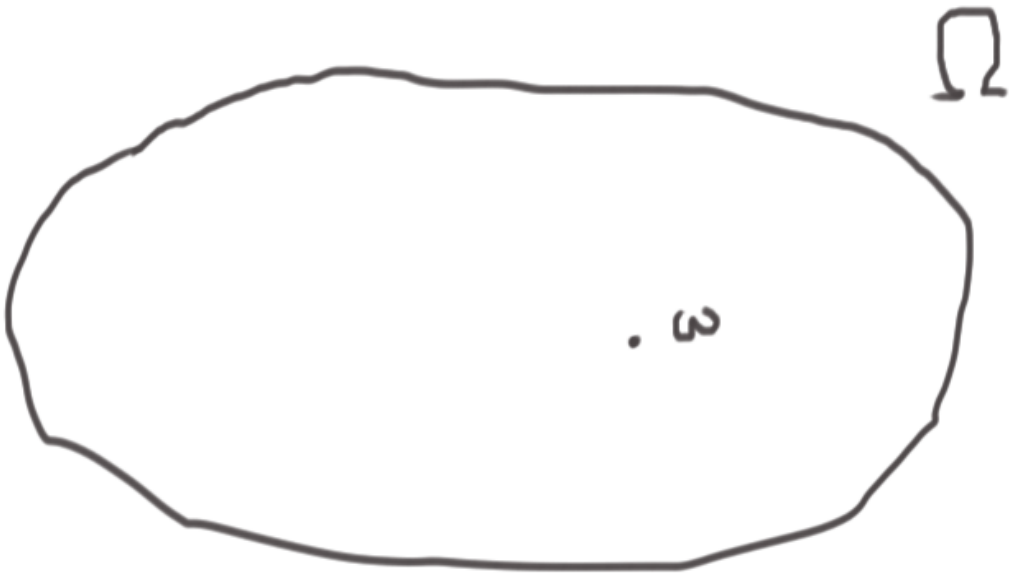
Note: Here are two views of the Monty Hall problem which may influence your intuition:
(a) The prize is either behind the door you picked or behind another door. You would expect the probability that it is behind your door to be 1/3 and therefore the probability that it is behind another door to be 2/3. However, if you change doors you will get the prize if it is behind another door.
(b) Imagine there were 100 doors. You pick a door, say 17, and Monty opens

all the other doors except one. He opens

$$1, 2, \ldots, 15, 16, 18, 19, \ldots, 60, 61, 63, 64, \ldots, 99, 100$$

Would you now stick with your door or switch to this very, very special door 62?

We think of an experiment involving randomness as an experiment with many possible outcomes. In order to build a mathematical model of the experiment we must first know the set of all possible outcomes of the experiment. This is called the *sample space* of the experiment and is denoted Ω (Capital omega). The individual outcomes, denoted ω (little omega) are the elements of the set Ω.



Example: The experiment involves tossing a coin. It can fall 'heads' or 'tails' so that the sample space is

$$\Omega = \{H, T\}.$$

Example: The experiment involves tossing a coin three times. Since each toss can fall 'heads' or 'tails' an outcome is a string of three symbols with each one a 'H' or a 'T'.

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Example: The experiment involves throwing a die. The upper side of the die can show any of the numbers $1, 2, 3, 4, 5, 6$ so that the sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Example: In the Lottery a machine draws successively 6 balls out of a pool of 42 numbered balls (disregard bonus ball). Although the balls are drawn in a specific order, the rules of the lottery are such that this order is irrelevant. So the outcome is a subset of size 6 taken from the set $\{1, 2, \ldots, 42\}$. The sample space is

$$\Omega = \{\{n_1, n_2, n_3, n_4, n_5, n_6\} \mid 1 \leq n_1 < n_2 < n_3 < n_4 < n_5 < n_6 \leq 42, n_j \in \mathbb{N}\}.$$

Example: A standard deck of cards has 52 cards, identified by their value

$$V = \{2, 3, 4, 5, \ldots, 9, 10, J, Q, K, A\}$$

and their suits

$$S = \{C(\clubsuit), D(\diamondsuit), H(\heartsuit), S(\spadesuit)\}$$

In the game of poker each player is dealt 5 cards, so the appropriate sample space is

$$\Omega = \{\{h_1, h_2, h_3, h_4, h_5\} \mid h_i \in V \times S\}.$$

Example: The Birthday problem. To analyse the probability of coincident birthdays in a group of n people, we conduct the following experiment. Order the people in some way, say alphabetically, and record the birthday of each person. An outcome is an ordered sequence

$$(d_1, d_2, \ldots, d_n) \quad \text{with} \quad d_j \in \{1, 2, \ldots, 365\}$$

(we disregard leap years). This is an n-sample from 365.

Sometimes, the question that we want to answer is 'what is the probability of a specific outcome?' (e.g., lotto: want probability of your grid being selected by the machine). But much more often, the relevant question is of the type 'what is the probability of the outcome being one of a specific set?'. Subsets of the sample space are called *events*.

Example: In the experiment where a coin is tossed three times with sample space

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

the event 'exactly two heads occur' is

$$A = \{HHT, HTH, THH\},$$

while the event 'tails come up for the first time at the second throw' is

$$B = \{HTH, HTT\}.$$

Example: In the lottery problem, the event 'all the winning numbers are even' is

$$A = \{\{n_1, n_2, \ldots, n_6\} \mid 2 \leq n_1 < n_2 < \ldots < n_6 \leq 42, n_j \in \{2, 4, 6, \ldots, 42\}\}.$$

Example: In the birthday problem, the event 'at least one common birthday' is

$$A = \{(d_1, d_2, \ldots, d_n) \mid d_i = d_j \quad \text{for some} \quad i \neq j\}.$$

Because events are subsets of the sample space Ω, we can use the tools of set theory to operate on them. The following summarises the probabilistic meanings of the various set operations:

Notation	set terminology	event terminology
$A \subset \Omega$	A is a subset	A is an event
$A \cup B$	the union of A and B	A or B
$A \cap B$	the intersection of A and B	A and B
$A \cap B = \emptyset$	A and B are disjoint	A and B are incompatible
$A \subseteq B$	A is included in B	A implies B
$\bar{A}$	the complement of A	the converse event of A
$\emptyset$	the empty set	the impossible event
Ω	the entire set	the certain event

Exercise: Review the set theoretic identities, see what they mean in probabilistic terms and convince yourself that they are true.

Probabilities.

We will be considering mostly experiments with a finite number of possible outcomes. For these the sample space is a finite set

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}.$$

As before we use the notation $|\Omega| = n$ to signify that Ω has n elements.

Note: Recall that such a sample space has 2^n events, i.e., 2^n distinct subsets, including $\emptyset$ and Ω itself.

To complete our model, it remains to construct a *probability measure*, i.e., a function $\mathbb{P}$ which associates to every event A a number $\mathbb{P}[A]$ with $0 \leq \mathbb{P}[A] \leq 1$. This is done as follows:
(a) To each outcome ω_j associate a number $p_j \geq 0$ in such a way that

$$p_1 + p_2 + \ldots + p_n = 1.$$

(Then clearly $p_j \leq 1$, for each j); the number p_j is called the *probability of the outcome* ω_j.
Next we define the probability of an event to be the sum of the probabilities of the outcomes contained in the event.

$$\mathbb{P}[A] = p_{j_1} + p_{j_2} + \ldots + p_{j_k}, \text{ where } A = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k}\}.$$

Note: We use the convention $\mathbb{P}[\emptyset] = 0$, and we note that by construction

$$\mathbb{P}[\Omega] = p_1 + p_2 + \ldots + p_n = 1.$$

Note: The theory does not tell you how to choose the numbers p_j; it is for you to decide on an appropriate choice of p_j's to model the specific experiment.

Example: The experiment is to toss a coin. Here $\Omega = \{H, T\}$. Since there are only two outcomes, we need only choose two non-negative numbers p_H and p_T with $p_H + p_T = 1$. If the coin is fair, we must have $p_H = p_T$ and hence $p_H = p_T = 1/2$. Here

$$\mathbb{P}[\Omega] = 1, \mathbb{P}[H] = 1/2, \mathbb{P}[T] = 1/2, \mathbb{P}[\emptyset] = 0.$$

Other choices model a biased coin (e.g., $p_H = 2/3$, $p_T = 1/3$.)

Example: The experiment is to toss a coin three times. Here

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Again a fair coin leads to giving equal probability $1/8$ to each outcome. We can compute the probability of any events associated to this sample space. If A is the event 'exactly two H's come up' then

$$A = \{HHT, HTH, THH\}$$

and we get

$$\mathbb{P}[A] = p_{HHT} + p_{HTH} + p_{THH} = 1/8 + 1/8 + 1/8 = 3/8.$$

If B is the event 'at least one H comes up' then

$$B = \{HHH, HHT, HTH, HTT, THH, THT, TTH\}$$

and we get

$$\begin{aligned} \mathbb{P}[B] &= p_{HHH} + p_{HHT} + p_{HTH} + p_{HTT} + p_{THH} + p_{THT} + p_{TTH} \\ &= 1/8 + 1/8 + 1/8 + 1/8 + 1/8 + 1/8 + 1/8 \\ &= 7/8. \end{aligned}$$

In general, $\mathbb{P}[C] = (1/8)|C|$, for this example.

Note: In those examples where the outcomes are all equally likely, their common value p must satisfy

$$1 = \mathbb{P}[\Omega] = p + p + p + \ldots + p = |\Omega| \cdot p$$

and p must be $1/|\Omega|$. In such a case, calculating the probability of an event amounts to counting the number of outcomes in it:

$$\mathbb{P}[A] = p_{j_1} + p_{j_2} + \ldots + p_{j_k} = |A| \cdot p = \frac{|A|}{|\Omega|}.$$

Example: What is the probability that a hand of poker gives a 'straight flush' (i.e., 5 cards of consecutive value, all of the same suit, such as $8\heartsuit, 9\heartsuit, 10\heartsuit, J\heartsuit, Q\heartsuit$)?

To find the answer, we come up with a procedure for constructing a generic straight flush:

Step 1	choose suit	4 choices
Step 2	Choose card of lowest value	10 choices

(Note: Ace can serve as highest card or as lowest card so that lowest card can belong to set $\{A, 2, 3, \ldots, 10\}$.)

This procedure determines the hand fully and uniquely; moreover all straight flushes can be constructed in this way. Therefore, there are 40 different hands of poker that give a straight flush and the probability is $40/\begin{pmatrix} 52 \\ 5 \end{pmatrix} = 1/(64974) = 0.00015$.

Example: Compute the probability of winning the jackpot in the lotto by playing a single grid.

An outcome of the lotto is a subset of 6 numbers from the set $\{1, 2, \ldots, 42\}$; there are

$$\left(\begin{array}{c} 42 \\ 6 \end{array} \right) = \frac{42!}{6!36!} = \frac{42 \times 41 \times 40 \times 39 \times 38 \times 37}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 5,245,786$$

such subsets. So the probability that a specific grid wins the jackpot is

$$\frac{1}{\left(\begin{array}{c} 42 \\ 6 \end{array} \right)} = \frac{1}{5,245,786} \simeq 0.19 \times 10^{-6}$$

Example: (where outcomes are not equally likely) Suppose a blue die and a red die are rolled together, and the numbers of dots that occur face up on each are added. What would reasonable probabilities be for the outcomes?

The sample space is $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and the probabilities are $p_2 = 1/36$, $p_3 = 2/36$, $p_4 = 3/36$, $p_5 = 4/36$, $p_6 = 5/36$, $p_7 = 6/36$, $p_8 = 5/36$, $p_9 = 4/36$, $p_{10} = 3/36$, $p_{11} = 2/36$, $p_{12} = 1/36$. We arrive at these by considering the related sample space

$$\Omega' = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$$

of possible pairs (i, j) where i is the number showing on the blue die and j is the number showing on the red die. If the dice are fair it is reasonable to assume each outcome in Ω' is equally likely and has probability $1/36$. Now for outcomes in the first sample space we add. For example, 6 can arise in 5 ways: $(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)$ so that we assign probability $5/36$.

Example: A sample space for the Monty Hall problem could be the set of 27 triples (A, B, C), where $A, B, C \in \{1, 2, 3\}$ and triple (A, B, C) is the outcome that
(a) The prize is behind door A
(b) You picked door B and
(c) The host opened door C, knowing that the prize was behind door A and that you had picked door B.

Now assign probabilities to each triple using these rules:
If C is either A or B (or both) the probability of the triple is 0 (Host will not

open the prize door, nor will he open the door you have chosen)
The probabilities for $(A, B, 1)$, $(A, B, 2)$ and $(A, B, 3)$ should add to $1/9$ (We expect all pairs (prize door, chosen door) to be equally likely)
The probabilities for (A, A, B) and (A, A, C) should be equal (If the host has a choice of two doors to open they should be equally likely)

triple	prob	triple	prob	triple	prob
$(1,1,1)$	0	$(2,1,1)$	0	$(3,1,1)$	0
$(1,1,2)$	$1/18$	$(2,1,2)$	0	$(3,1,2)$	$1/9$
$(1,1,3)$	$1/18$	$(2,1,3)$	$1/9$	$(3,1,3)$	0
$(1,2,1)$	0	$(2,2,1)$	$1/18$	$(3,2,1)$	$1/9$
$(1,2,2)$	0	$(2,2,2)$	0	$(3,2,2)$	0
$(1,2,3)$	$1/9$	$(2,2,3)$	$1/18$	$(3,2,3)$	0
$(1,3,1)$	0	$(2,3,1)$	$1/9$	$(3,3,1)$	$1/18$
$(1,3,2)$	$1/9$	$(2,3,2)$	0	$(3,3,2)$	$1/18$
$(1,3,3)$	0	$(2,3,3)$	0	$(3,3,3)$	0

Now, if E is the event that you win a prize by changing doors then E consists of those triples (A, B, C) with A and B distinct. This event has probability $2/3$.

Proposition: (The addition rule for probabilities) Suppose $A \cap B = \emptyset$. Then

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B].$$

Proof: If $A = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k}\}$ and $B = \{\omega_{r_1}, \omega_{r_2}, \ldots, \omega_{r_s}\}$, then $A \cup B = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_k}, \omega_{r_1}, \omega_{r_2}, \ldots, \omega_{r_s}\}$ and there is no repetition since $A \cap B = \emptyset$. From the definition

$$\begin{aligned} \mathbb{P}[A \cup B] &= p_{j_1} + \ldots p_{j_k} + p_{r_1} + \ldots + p_{r_s} \\ &= \mathbb{P}[A] + \mathbb{P}[B]. \end{aligned}$$

Corollary: (The complement rule) $\mathbb{P}[\bar{A}] = 1 - \mathbb{P}[A]$.

Proof: Note that $\Omega = A \cup \bar{A}$, $A \cap \bar{A} = \emptyset$ and $\mathbb{P}[\Omega] = 1$, so that

$$1 = \mathbb{P}[\Omega] = \mathbb{P}[A] + \mathbb{P}[\bar{A}]$$

Note: As in the case of the subtraction rule for counting, the complement rule is useful in cases where $\bar{A}$ is a simpler event than A. Then it is better to calculate $\mathbb{P}[\bar{A}]$ first in order to get $\mathbb{P}[A]$.

Example: For example, when a coin is tossed three times and where B is the event 'at least one H comes up', it is clear that $\bar{B}$ is the event 'no H comes up' and $\bar{B} = \{TTT\}$, giving $\mathbb{P}[\bar{B}] = 1/8$ and

$$\mathbb{P}[B] = 1 - \mathbb{P}[\bar{B}] = 1 - 1/8 = 7/8.$$

Example: What is the probability that a four digit PIN has at least one repeated digit?

Here $|\Omega| = 10^4$ (4 samples from 10). Suppose A is the event that there is at least one repeated digit. Then $\bar{A}$ is the event that no digits are repeated (permutations of 4 numbers from 10). Thus $|\bar{A}| =^{10} P_4$ and $\mathbb{P}[\bar{A}] =^{10} P_4/10^4 = 0.504$. Hence $\mathbb{P}[A] = 1 - 0.504 = 0.496$.

Example: The birthday problem. We want to calculate the probability that, in a class of n students, two or more have a common birthday. The sample space is

$$\Omega = \{(d_1, d_2, \ldots, d_n) \mid d_i \text{ integers}, 1 \leq d_i \leq 365\}$$

and the event 'at least one common birthday' is

$$A = \{(d_1, d_2, \ldots, d_n) \in \Omega \mid d_i = d_j \text{ for at least one pair } i \neq j\}$$

Assume that all birthdays are equally likely, so that all outcomes in Ω are equally likely. Then $\mathbb{P}[A] = |A|/|\Omega|$ and the problem reduces to finding the number of elements in Ω and the number of elements in A. We start with Ω. This is just the set of n-samples from 365. (Replacement is used and order is important.) $|\Omega| = (365)^n$ and the probability of any specific outcome is

$$p = \frac{1}{365^n}.$$

Next we look at A. Since this involves single repeated dates, dates repeated multiple times, separate dates repeated etc., we consider the converse event $\bar{A}$. All the outcomes in $\bar{A}$ have no repeated date. So we will compute $\mathbb{P}[\bar{A}]$ and then get $\mathbb{P}[A] = 1 - \mathbb{P}[\bar{A}]$. Now $\mathbb{P}[\bar{A}] = |\bar{A}|/|\Omega|$. However outcomes in $\bar{A}$ are simply permutations of n dates from 365. So

$$|\bar{A}| =^{365} P_n = 365 \times 364 \times \ldots \times (365 - n + 1)$$

and

$$\mathbb{P}[\bar{A}] = \frac{365 \times 364 \times \ldots \times (365 - n + 1)}{365^n} = \frac{365}{365}\frac{364}{365}\frac{363}{365} \cdots \frac{365 - n + 1}{365}.$$

Note: The value for n is obtained from the value for $n - 1$ by multiplying by the fraction $(365 - n + 1)/(365)$. This can be programmed easily enough. Here's a python version:

```
>>> a, b = 1, 365
>>> while b > 340:
...  print a
...  a, b = (a*b)/(365.0), b-1
...
1
1.0
0.997260273973
0.991795834115
0.983644087533
0.9728644263
0.959537516351
0.943764296904
0.925664707648
0.905376166111
0.883051822289
0.858858621678
0.832975211162
0.805589724768
0.776897487995
0.747098680236
0.716395994747
0.684992334703
0.653088582128
0.620881473968
0.588561616419
0.556311664835
0.524304692337
0.492702765676
0.461655742085
```

The variable a is the current probability while b is $365 - n$ so initialise a at 1 and b at 365. I just do 25 steps, hence the $b > 340$ condition. In the loop the probability has to be printed, b has to decrease by 1 and the probability has to be multiplied by $b/365$. The 365.0 forces the output to be a decimal. We see a $50 - 50$ chance of shared birthdays as soon as n reaches 22.

Conditional Probability: Conditional probability tries to answer the following question: how should we modify the probability associated to an event A if we happen to know that some other event B has occured? To motivate the definition we consider an example where we know all the probabilities involved.

Example: Suppose we are dealt two cards from a standard pack. Let A be the event that the second card is black and let B be the event that the first card is black. Since we are concerned with the order of the cards the sample space is the set of ordered samples of size 2 from 52 without replacement. We would agree that $\mathbb{P}[B] = 1/2$ (26 black cards among the 52) while the probability that A occurs given that B has already occured should be $25/51$ (25 black cards left among the 51). Note that $A \cap B$ is the event that both cards are black and hence

$$\mathbb{P}[A \cap B] = \frac{{}^{26}P_2}{{}^{52}P_2} = \frac{26 \times 25}{52 \times 51} = \mathbb{P}[B] \times \frac{25}{51}.$$

It seems that for this example, the probability that A occurs given that B has already occured is $\mathbb{P}[A \cap B]/\mathbb{P}[B]$.

Definition: For any fixed event B, such that $\mathbb{P}[B] > 0$, we define the conditional probability of A given B, which we denote $\mathbb{P}[A|B]$, to be

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Note: It makes sense that $\mathbb{P}[A|B]$ should be proportional to $\mathbb{P}[A \cap B]$ since the outcomes ω contributing to this probability must lie in B since we assume B has occured. Also the denominator is needed for normalisation so that $\mathbb{P}[B|B] = 1$.

Example: Two dice are tossed and the upwards numbers are added. What is the probability that the sum is less than 6, given that the sum is even?

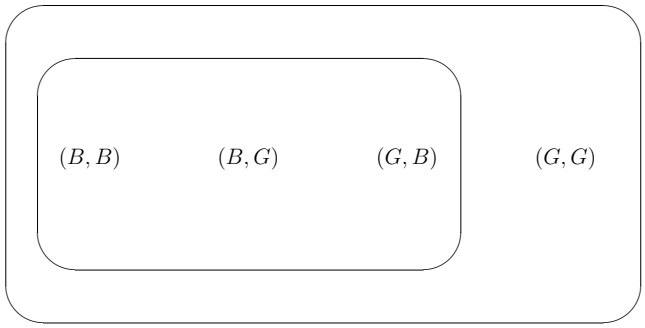
The sample space is $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and the probabilities are $P(2) = 1/36$, $P(3) = 2/36$, $P(4) = 3/36$, $P(5) = 4/36$, $P(6) = 5/36$, $P(7) = 6/36$, $P(8) = 5/36$, $P(9) = 4/36$, $P(10) = 3/36$, $P(11) = 2/36$, $P(12) = 1/36$. If A is the event the sum is even and B is the event that the sum is less than 6 then

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = (1/9)/(1/2) = 2/9.$$

Example: Imagine a couple with two children, each of whom is equally likely to be a boy or a girl. Now suppose you are given the information that one is a boy. What is the probability that the other child is a boy?

Here the sample space is the set of pairs $\{(B,B),(B,G),(G,B),(G,G)\}$ with each pair equally likely. The event A consists of the first 3 outcomes and has $\mathbb{P}[A] = 3/4$. The event B consists of $\{(B,B)\}$ and has $\mathbb{P}[B] = 1/4$. The event $A \cap B$ is the same event as B. Thus

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = (1/4)/(3/4) = 1/3.$$



Definition: Recall that we say that a collection of subsets

$$B_1, B_2, \ldots, B_k$$

forms a partition of the sample space Ω if

$$B_1 \cup B_2 \cup \ldots \cup B_k = \Omega \quad \text{and} \quad B_i \cap B_j = \emptyset \quad \text{for} \quad i \neq j$$

Example: If A is any event then the collection $\{A, \overline{A}\}$ always gives a partition.

Proposition: (The law of total probability) Let $B_1, B_2, \ldots, B_k$ be a partition of Ω into events of positive probability. Then, for every $A \subset \Omega$

$$\mathbb{P}[A] = \mathbb{P}[A|B_1]\mathbb{P}[B_1] + \mathbb{P}[A|B_2]\mathbb{P}[B_2] + \ldots + \mathbb{P}[A|B_k]\mathbb{P}[B_k]$$

Proof: Since the B_j form a partition of Ω we can write A as a disjoint union

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)$$

so that

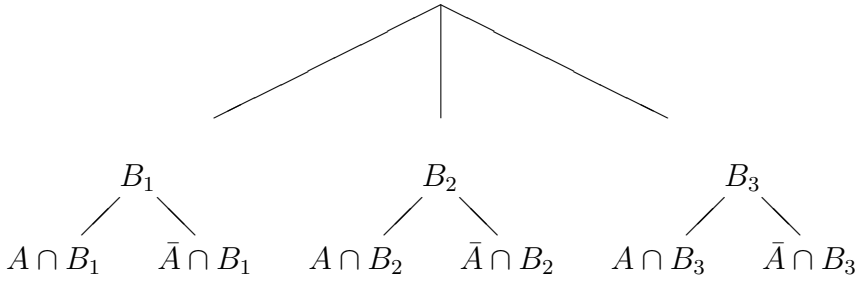
$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}[A \cap B_1] + \mathbb{P}[A \cap B_2] + \ldots + \mathbb{P}[A \cap B_k] \\ &= \mathbb{P}[A|B_1]\mathbb{P}[B_1] + \mathbb{P}[A|B_2]\mathbb{P}[B_2] + \ldots + \mathbb{P}[A|B_k]\mathbb{P}[B_k]\end{aligned}$$

where we use the fact

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \quad \Rightarrow \quad \mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$$

Note: The law of total probability is useful for breaking up a complicated problem into manageable pieces.

Note: Sometimes it helps to view the law in terms of a tree diagram:



Example: A bag contains two coins, a fair one and a biased one. The biased one falls heads with probability 2/3. A coin is selected at random and tossed. What is the probability that it falls heads?

Let F be the event that the fair coin is selected, B the event that the biased coin is selected and H the event that heads show. Since there are only two coins, $\{B, F\}$ form a partition of the sample space. Hence

$$\mathbb{P}[H] = \mathbb{P}[H|F]\mathbb{P}[F] + \mathbb{P}[H|B]\mathbb{P}[B] = (1/2)(1/2) + (2/3)(1/2) = 7/12$$

Note: It might be helpful to think of this example as follows. Suppose we do this experiment 12 times. We expect the fair coin to be selected half of the time, so we expect the fair coin to occur 6 times. When we toss the fair coin we expect to get heads half of the time. So three of the outcomes should be the fair coin showing heads. Similarly, three of the outcomes should be the fair coin showing tails. Likewise we expect the biased coin to be selected half of the time, so we expect the biased coin to occur 6 times. When we

toss the biased coin we expect to get heads two thirds of the time. So four of the outcomes should be the biased coin showing heads. Similarly, two of the outcomes should be the biased coin showing tails. Our 12 outcomes should be

$$\{(F,H),(F,H),(F,H),(F,T),(F,T),(F,T),$$
$$(B,H),(B,H),(B,H),(B,H),(B,T),(B,T)\}.$$

Sure enough, heads show 7 times out of the 12.

Note: When considering $\mathbb{P}[A|B]$ it is natural to think of B as the 'cause' and of A as the 'effect'. With this interpretation the following result allows us to derive the probability that a given effect is due to a certain cause.

Proposition: (Bayes' formula) Let $B_1, B_2, \ldots, B_k$ be a partition of Ω into events of positive probability. Then, for every $A \subset \Omega$

$$\mathbb{P}[B_j|A] = \frac{\mathbb{P}[A|B_j]\mathbb{P}[B_j]}{\mathbb{P}[A|B_1]\mathbb{P}[B_1] + \mathbb{P}[A|B_2]\mathbb{P}[B_2] + \ldots + \mathbb{P}[A|B_k]\mathbb{P}[B_k]}$$

Proof: Using the definition of conditional probability and the law of total probability we get

$$\begin{aligned}
\mathbb{P}[B_j|A] &= \frac{\mathbb{P}[B_j \cap A]}{\mathbb{P}[A]} \\
&= \frac{\mathbb{P}[A|B_j]\mathbb{P}[B_j]}{\mathbb{P}[A]} \\
&= \frac{\mathbb{P}[A|B_j]\mathbb{P}[B_j]}{\mathbb{P}[A|B_1]\mathbb{P}[B_1] + \mathbb{P}[A|B_2]\mathbb{P}[B_2] + \ldots + \mathbb{P}[A|B_k]\mathbb{P}[B_k]}
\end{aligned}$$

Example: A bag contains two coins, a fair one and a biased one. The biased one falls heads with probability 2/3. A coin is selected at random and tossed. Suppose now the selected coin falls H. What is the probability that the fair coin was selected?

Using the same notation as before, we know that

$$\mathbb{P}[H|F] = 1/2, \quad \mathbb{P}[H|B] = 2/3, \quad \mathbb{P}[F] = \mathbb{P}[B] = 1/2.$$

So we can use Bayes' formula to give

$$\mathbb{P}[F|H] = \frac{\mathbb{P}[H|F]\mathbb{P}[F]}{\mathbb{P}[H|F]\mathbb{P}[F] + \mathbb{P}[H|B]\mathbb{P}[B]}$$

$$
\begin{aligned}
&= \frac{(1/2)(1/2)}{(1/2)(1/2)+(2/3)(1/2)} \\
&= \frac{3}{7}.
\end{aligned}
$$

Note: When we last looked at this example we argued that doing the experiment 12 times should yield

$$\{(F,H),(F,H),(F,H),(F,T),(F,T),(F,T),$$

$$(B,H),(B,H),(B,H),(B,H),(B,T),(B,T)\}.$$

We see that of the 7 times heads show, 3 of them occur when we have selected the fair coin.

Example: A company produces components in two factories. It is know that a component from Factory A has a probability of $1/100$ of being faulty while a component from Factory B has a probability of $1/200$ of being faulty. It is also known that a finished component has a probability of $1/3$ of being from Factory A and a probability of $2/3$ of being from Factory B. Use Baye's Theorem to compute the probability that a finished component which is faulty was produced in Factory A.

Using the obvious names for the events we have

$$\mathbb{P}[F|A] = 1/100, \quad \mathbb{P}[F|B] = 1/200, \quad \mathbb{P}[A] = 1/3, \quad \mathbb{P}[B] = 2/3.$$

So we can use Bayes' formula to give

$$
\begin{aligned}
\mathbb{P}[A|F] &= \frac{\mathbb{P}[F|A]\mathbb{P}[A]}{\mathbb{P}[F|A]\mathbb{P}[A]+\mathbb{P}[F|B]\mathbb{P}[B]} \\
&= \frac{(1/100)(1/3)}{(1/100)(1/3)+(1/200)(2/3)} \\
&= \frac{1}{2}.
\end{aligned}
$$

Example: (MS121 May 2017) In a referendum 55% of the voters vote 'yes' and 45% vote 'no'. A TV station conducts an exit poll. All of those who respond say truthfully how they voted. Of those who vote 'yes' 70% respond and of those who vote 'no' 90% respond. Use Bayes' formula to compute the percentage of 'yes' votes the poll will predict.

For a given voter let R be the event that they respond, Y be the event that they vote 'yes' and N be the event that they vote 'no'. We know that $\mathbb{P}[R \mid Y] = 70/100$, $\mathbb{P}[R \mid N] = 90/100$, $\mathbb{P}[Y] = 55/100$ and $\mathbb{P}[N] = 45/100$, while $\{Y, N\}$ partitions the sample space. We seek $\mathbb{P}[Y \mid R]$. Apply Bayes' Theorem.

$$\begin{aligned}
\mathbb{P}[Y \mid R] &= \frac{\mathbb{P}[Y \cap R]}{\mathbb{P}[R]} \\
&= \frac{\mathbb{P}[R \mid Y]\mathbb{P}[Y]}{\mathbb{P}[R \mid Y]\mathbb{P}[Y] + \mathbb{P}[R \mid N]\mathbb{P}[N]} \\
&= \frac{(70/100)(55/100)}{(70/100)(55/100) + (90/100)(45/100)} \\
&= 77/158.
\end{aligned}$$

Example: (Monty Hall Problem) In a game show a contestant is told that a prize is behind one of three doors. After the contestant picks one door the host (Monty) opens another revealing no prize. The contestant is invited to switch his/her choice to the remaining door or not. What should he/she do?

Solution: Suppose the choice of door has occurred. So we now name the doors with the contestant choosing door 1. Let P_i be the event that the prize is behind door i, and H_i the event that the host opens door i. Conditional probabilities are computed after the contestant has chosen door 1. So $\mathbb{P}[H_1] = 0$ (host will not open the prize door) and $\mathbb{P}[H_i|P_i] = 0$ (host will not open the chosen door). Also $\mathbb{P}[H_i|P_j] = 1/2$ for $j = 1$ (host is equally likely to open either of the other doors if contestant has chosen the prize door) and $\mathbb{P}[H_i|P_j] = 1$ for $j \neq 1$ (host can only open the one other door if contestant has not chosen the prize door).

$$\begin{aligned}
\mathbb{P}[P_1|H_2] &= \frac{\mathbb{P}[P_1 \cap H_2]}{\mathbb{P}[H_2]} \\
&= \frac{\mathbb{P}[H_2|P_1]\mathbb{P}[P_1]}{\mathbb{P}[H_2|P_1]\mathbb{P}[P_1] + \mathbb{P}[H_2|P_2]\mathbb{P}[P_2] + \mathbb{P}[H_2|P_3]\mathbb{P}[P_3]}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1/2)(1/3)}{(1/2)(1/3)+(0)(1/3)+(1)(1/3)} \\
&= 1/3.
\end{aligned}
$$

Just to be sure let's check the other possibility.

$$
\begin{aligned}
\mathbb{P}[P_3|H_2] &= \frac{\mathbb{P}[P_3 \cap H_2]}{\mathbb{P}[H_2]} \\
&= \frac{\mathbb{P}[H_2|P_3]\mathbb{P}[P_3]}{\mathbb{P}[H_2|P_1]\mathbb{P}[P_1]+\mathbb{P}[H_2|P_2]\mathbb{P}[P_2]+\mathbb{P}[H_2|P_3]\mathbb{P}[P_3]} \\
&= \frac{(1)(1/3)}{(1/2)(1/3)+(0)(1/3)+(1)(1/3)} \\
&= 2/3.
\end{aligned}
$$

So be sure to change your choice if you are a contestant in such a show!

Note: It can happen that knowing the event B has occured does not change our view on the probability of A happening. This is the case if $\mathbb{P}[A|B] = \mathbb{P}[A]$ or

$$\mathbb{P}[A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Definition: Two events A and B are said to be independent for the probability measure $\mathbb{P}$ if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$$

Example: A card is selected at random from a standard deck. If A is the event that the card is an ace and S is the event that the card is a spade then A and S are independent.

Clearly, if all the cards are equally likely to be picked

$$\mathbb{P}[A] = 1/13, \quad \mathbb{P}[S] = 1/4, \quad \mathbb{P}[A \cap S] = 1/52$$

Example: A coin is tossed three times. If A is the event that H comes up at least twice and B is the event that the first time H occurs is on the second toss then A and B are independent.

The sample space has eight equally likely outcomes and the events of interest are

$$A = \{HHT, HTH, THH, HHH\}, B = \{THH, THT\}, A \cap B = \{THH\}$$

The probabilities are

$$\mathbb{P}[A] = 1/2, \quad \mathbb{P}[B] = 1/4, \quad \mathbb{P}[A \cap B] = 1/8$$

So A and B are independent.

Example: Two cards are chosen from a deck of 52 without replacement. If A is the event that the first card is black and B is the event that the second card is black then A and B are not independent.

$$\mathbb{P}[A] = 1/2, \mathbb{P}[B] = 1/2, \mathbb{P}[A \cap B] = \begin{pmatrix} 26 \\ 2 \end{pmatrix} / \begin{pmatrix} 52 \\ 2 \end{pmatrix} = \frac{25}{102} \neq \frac{1}{4}$$

Note: Do not confuse independent events ($\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$) with disjoint events ($A \cap B = \emptyset$). The two are very different:
Being disjoint is an absolute property, not relative to any probability measure.
If two events are disjoint, knowing that one event has occurred gives you a lot of information about the other event (namely that it cannot occur).
If $A \cap B = \emptyset$ then $\mathbb{P}[A \cap B] = 0$ and independence can only occur if one of A or B has probability zero.

Note: The idea of independence can be extended to collections of n events. A collection of events is said to be mutually independent for the probability measure $\mathbb{P}$ if, for any subcollection of the events, the probability of their intersection is equal to the product of their probabilities.

Definition: A Bernoulli trial is an experiment which can have only two outcomes, often described as 'success' and 'failure'; the standard notation for the probability of success is p so that the probability of failure is $1-p$. A sequence of n Bernoulli trials is an experiment in which n trials take place in succession. The trials are conducted under identical conditions (so that the probability of success in any trial remains p throughout) and independently of each other.

Example: The prototype of a sequence of Bernoulli trials is the experiment 'toss a coin n times '; whether you call H a success or T a success is up to you.

Example: Four coins are tossed and the number of heads showing is noted. The sample space is $\{0, 1, 2, 3, 4\}$ and the probabilities are computed using the

sample space of all strings of length 4 in H and T. The event that no heads show is the subset consisting of the single string TTTT with probability (1/2)(1/2)(1/2)(1/2) since the probability of T is 1/2 in each toss. The event that one head shows is the subset $\{HTTT, THTT, TTHT, TTTH\}$ with probability 4(1/2)(1/2)(1/2)(1/2) since the probability of T is 1/2 in each toss and the probability of H on each toss is 1/2. The event that two heads show is the subset

$$\{HHTT, HTHT, HTTH, THHT, THTH, TTHH\}$$

with probability 6(1/2)(1/2)(1/2)(1/2). The number in this subset is $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ corresponding to the number of ways of choosing the two positions for the H's in the string of length 4. We complete the other probabilities similarly.

Sample point	0	1	2	3	4
Probability	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Example: A die is rolled 6 times and the number of 1's showing is noted. The sample space is $\{0, 1, 2, 3, 4, 5, 6\}$ and the probabilities are computed using the sample space of all strings of length 6 in S and F, where S (success) is an occurence of 1 while F (failure) is an occurence of any of $\{2, 3, 4, 5, 6\}$. Thus $p = 1/6$ and $1 - p = 5/6$. The event that no 1's show is the subset consisting of the single string FFFFFF with probability $(5/6)^6$ since the probability of F is 5/6 for each roll. The event that one 1 shows is the subset $\{SFFFFF, FSFFFF, \ldots, FFFFFS\}$ with probability $6(1/6)(5/6)^5$ since the probability of S is 1/6 in each roll and the probability of F on each roll is 5/6. The event that two 1's show is the subset

$$\{SSFFFF, SFSFFF, \ldots, FFFFSS\}$$

with probability $(15)(1/6)^2(5/6)^4$. The number in this subset is $\begin{pmatrix} 6 \\ 2 \end{pmatrix}$ corresponding to the number of ways of choosing the two positions for the S's in the string of length 6. We complete the other probabilities similarly.

Proposition: The probability of exactly j successes in n Bernoulli trials is

$$\begin{pmatrix} n \\ j \end{pmatrix} p^j (1-p)^{n-j} \quad \text{for} \quad 0 \leq j \leq n$$

where p is the probability of success in any individual trial.

Proposition: The probability of exactly j successes in n Bernoulli trials is

$$\left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} \quad \text{for} \quad 0 \leq j \leq n$$

where p is the probability of success in any individual trial.

Proof: The sample space Ω is the set of all sequences of length n in the symbols S (success) and F (failure). There are 2^n such sequences, but they are not equally likely; for instance $SS \ldots S$ has probability p^n while $FFS \ldots S$ has probability $p^2(1-p)^{n-2}$. However all sequences with exactly j successes have the same probability, $p^j(1-p)^{n-j}$, irrespective of the location of the j successes in the sequence. So the probability of the event A_j: exactly j successes is $p^j(1-p)^{n-j}|A_j|$.

To count the number of outcomes in A_j we note that such outcomes are completely specified by the location of the j successful trials; for instance

$$SSFSFFFSFFSF$$

has five successes out of twelve, located at trials $1, 2, 4, 8, 11$. But choosing the location of j successes amounts to choosing j numbers out of n, a problem we have seen before. Hence

$$|A_j| = \left(\begin{array}{c} n \\ j \end{array} \right),$$

proving the result.

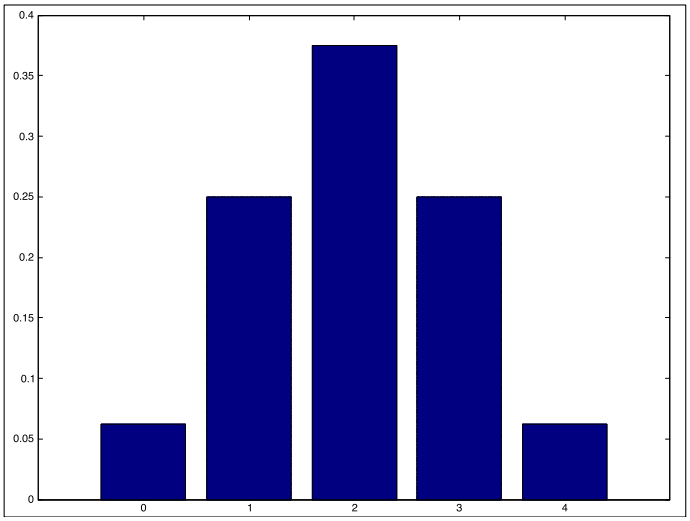
Definition: The assignment of probabilities

$$P(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

to the sample points $\{0, 1, 2, 3, \ldots, n\}$ is called the binomial distribution.

Example: Four coins are tossed and the number of heads showing is noted. The sample space is $\{0, 1, 2, 3, 4\}$ and the probabilities are

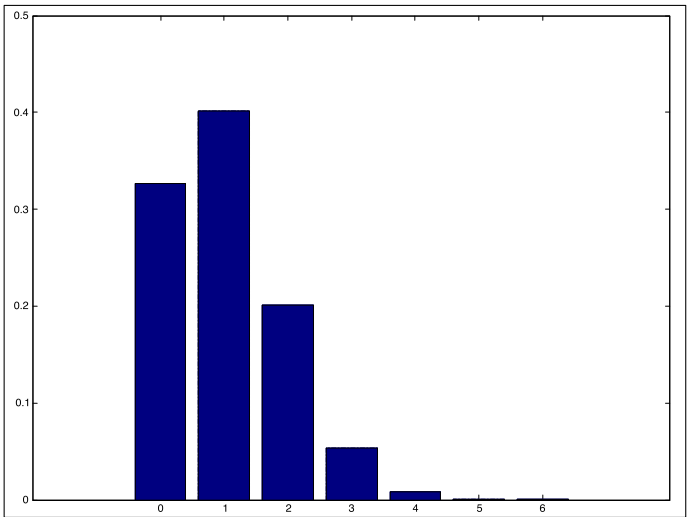
Sample point	0	1	2	3	4
Probability	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$



Here $p = q = 1/2$ and $n = 4$.

Example: Six dice are tossed and the number of 1's showing is noted. The sample space is $\{0, 1, 2, 3, 4, 5, 6\}$ and the probabilities are

Sample point	0	1	2	3	4	5	6
Probability	$\frac{15625}{46656}$	$\frac{18750}{46656}$	$\frac{9375}{46656}$	$\frac{2500}{46656}$	$\frac{375}{46656}$	$\frac{30}{46656}$	$\frac{30}{46656}$



Here $p = 1/6$, $q = 5/6$ and $n = 6$.

Example: A company produces microchips and one microchip in 1000 is defective. What is the probability that a sample of 200 microchips contains

at most two defective chips?

Setting $p = 1/1000 = 0.001$ and $n = 200$ the desired probability is

$$\binom{n}{0} p^0(1-p)^{n-0} + \binom{n}{1} p^1(1-p)^{n-1} + \binom{n}{2} p^2(1-p)^{n-2}$$

since the sample will have at most 2 defective items if it has 0, 1 or 2 defective items. Computing this quantity gives approximately 0.99887.

Question 3 from last year's repeat paper:

QUESTION 3

(a) A committee of 4 is to be chosen from a group of 10 people.
(i) In how many ways can this be done?
(ii) Suppose 5 of the group are women and 5 are men. How many of the committees have at least one woman and at least one man? Explain your answer.

[8 marks]

(b) A fair die (6-sided) is rolled 30 times. Use the Binomial distribution to calculate the probability that the number of times side 1 shows is two or less.

[9 marks]

(b) The probability that side 1 shows on any particular roll of the die is 1/6. The probability that side 1 does not show on any particular roll is 5/6. The probability that side 1 shows twice or less in 30 rolls of the die is

$$\mathbb{P}(0; 30, 1/6) + \mathbb{P}(1; 30, 1/6) + \mathbb{P}(2; 30, 1/6)$$

which we compute using $\mathbb{P}(k; n, p) = \binom{n}{k} p^k(1-p)^{n-k}$ to be

$$\begin{aligned} & \binom{30}{0}(1/6)^0(5/6)^{30} + \binom{30}{1}(1/6)^1(5/6)^{29} + \binom{30}{2}(1/6)^2(5/6)^{28} \\ = \;& (1)(1)(5/6)^{30} + (30)(1/6)(5/6)^{29} + (15)(29)(1/6)^2(5/6)^{28} \\ \approx \;& 0.103 \end{aligned}$$

(a) (i) Choose 4 from 10 in $\begin{pmatrix} 10 \\ 4 \end{pmatrix} = 210$ ways.

(ii) The number of committees which do not have at least one woman and at least one man is the number with no women or no men. This number is obtained by choosing all men or all women giving

$$\begin{pmatrix} 5 \\ 4 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 5 \\ 4 \end{pmatrix} = 5 + 5 = 10.$$

Thus the number we want is obtained using the subtraction law to be $210 - 10 = 200$.