1. The Pigeonhole Principle

Notation. Throughout this course we write $\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$ for the set of natural numbers; note that we don't consider zero to be a natural number. In some books and courses, particular those closer to computer science, you will find that zero is also considered to be a natural number. There are several good arguments for and against including zero, and it's mostly a matter of preference, but be careful to avoid confusion.

The ceiling of a real number x, denoted $\lceil x \rceil$, is the smallest integer greater than or equal to x, and the floor of x, denoted $\lfloor x \rfloor$, is the largest integer less than or equal to x. So, for example, $\lceil \frac{1}{2} \rceil = 1$, $\lfloor \frac{1}{2} \rfloor = 0$, $\lceil \frac{7}{3} \rceil = 3$, $\lfloor \frac{7}{3} \rfloor = 2$, $\lceil -6.7 \rceil = -6$ and $\lfloor -6.7 \rfloor = -7$, but $\lceil x \rceil = \lfloor x \rfloor = x$ if x is an integer.

Theorem 1.1 (Pigeonhole Principle). *Let $n \in \mathbb{N}$. If at least $n+1$ pigeons are placed in n pigeonholes, then some pigeonhole contains at least two pigeons.*

This statement may seem obvious, especially if you consider a specific value of n (for example, the statement for $n = 2$ says that if 3 pigeons are placed in 2 pigeonholes, then one of the pigeonholes must contain at least two pigeons). However, it arises frequently as a useful tool for both mathematical and non-mathematical arguments, as in the following examples. We will derive Theorem 1.1 from a more general result shortly.

Example. In my drawer are 10 black socks and 8 white socks. How many socks must I take from the drawer (without looking) to guarantee that I have a matching pair among the socks I have taken out?

Solution. Three socks. To see this, think of each sock as a pigeon, and as you take them out the drawer, imagine that each white sock is placed in a 'white' pigeonhole and each black sock is placed in a 'black' pigeonhole. So socks in the same pigeonhole are the same colour. After three socks have been taken, there are three pigeons (socks) in two pigeonholes, so by the Pigeonhole Principle (with $n = 2$ pigeonholes) some pigeonhole contains more than one sock, so contains a matching pair. □

Example. Let $x_1, x_2, \dots, x_6$ be integers between 1 and 9. Prove that either two of these integers x_i are equal, or two of these integers x_i sum to 10.

Solution. Create 'pigeonholes' A, B, C, D and E. For each $1 \leq j \leq 6$, place the 'pigeon' x_j into a pigeonhole according to the following rule:

$$\text{Put } x_j \text{ in pigeonhole } \begin{cases} A & \text{if } x_j = 1 \text{ or } x_j = 9, \\ B & \text{if } x_j = 2 \text{ or } x_j = 8, \\ C & \text{if } x_j = 3 \text{ or } x_j = 7, \\ D & \text{if } x_j = 4 \text{ or } x_j = 6, \\ E & \text{if } x_j = 5. \end{cases}$$

Note that this rule places each of the 6 integers into one of the five pigeonholes. The Pigeonhole Principle with $n = 5$ therefore implies that some pigeonhole contains at least two of the integers. Choose two integers which lie in the same pigeonhole; then these integers are either equal or sum to 10 by definition of the pigeonholes (for example, if D contains two integers, then either they are equal or one is 4 and one is 6, giving a sum of 10). □

Example. Prove that for every $r \in \mathbb{N}$, every sequence of r integers contains a consecutive subsequence whose sum is a multiple of r.

Solution. Let $r \in \mathbb{N}$ and let $x_1, \dots, x_r$ be a sequence of r integers. Also let $S_k = \sum_{i=1}^{k} x_i$ be the partial sums for each $0 \leq k \leq r$ (so $S_0 = 0$, $S_1 = x_1$, $S_2 = x_1 + x_2$, $S_3 = x_1 + x_2 + x_3$ and so forth). We then have $r + 1$ partial sums $S_0, S_1, \dots, S_r$, but there are only r possible remainders for an integer when divided by r. So by the Pigeonhole Principle there must be two partial sums with the same remainder on division by r, S_j and S_k with $0 \leq j < k \leq r$. It follows that $S_k - S_j$ is divisible by r. Since $S_k - S_j = \sum_{i=1}^{k} x_i - \sum_{i=1}^{j} x_i = \sum_{i=j+1}^{k} x_i = x_{j+1} + x_{j+2} + \dots + x_k$, we have found a consecutive subsequence $x_{j+1}, x_{j+2}, \dots, x_k$ whose sum is a multiple of r.[1] □

As shown in these examples, the 'pigeons' may be any discrete[2] object, mathematical or otherwise, and the pigeonholes can be any collections to which pigeons may be assigned, provided that each pigeon is assigned to some pigeonhole[3].

The next result is a more general version of the Pigeonhole Principle.

Theorem 1.2 (Pigeonhole Principle – general form). *Let $n, k \in \mathbb{N}$. If at least n pigeons are placed in k pigeonholes, then some pigeonhole contains at least $\lceil \frac{n}{k} \rceil$ pigeons.*

Note that the (mean) average number of pigeons per pigeonhole is n/k. So this form of the Pigeonhole Principle is equivalent to saying that at least one member of any collection of integers is greater than or equal to the average of the collection, a fact you are probably familiar with. Also note that for every $n \in \mathbb{N}$ we have $\lceil \frac{n+1}{n} \rceil = 2$, so our original form of the Pigeonhole Principle, Theorem 1.1, is precisely the special case of Theorem 1.2 with $n + 1$ and n in place of n and k respectively.

Example. A hand in Bridge consists of 13 cards from a standard 52-card deck. Prove that every such hand must contain at least four cards of the same suit.

Solution. Divide the 13 cards of the hand into four piles, one for each suit (these piles are the 'pigeonholes', and the cards are the 'pigeons'). Since there are 13 cards and 4 piles, by the Pigeonhole Principle (applied with $n = 13$ and $k = 4$) some pile must contain at least $\lceil \frac{13}{4} \rceil = \lceil 3.25 \rceil = 4$ cards. □

Example. For any nine distinct points in a square of side length one, there are three points which form the vertices of a triangle whose longest side has length at most $\frac{1}{\sqrt{2}}$ (here we consider a straight line to be a 'flat' triangle).

Solution. Divide the square into four quarters, each of which is a square of side length $1/2$, in the natural way. So each of the nine points lies in some quarter (if a point is on a boundary than chose one of the corresponding quarters arbitrarily and say that it lies in that quarter). By the Pigeonhole Principle some quarter must then contain at least three of the points (since $\lceil \frac{9}{4} \rceil = 3$). Choose three points which lie in the same quarter and let T be the triangle with these three points as vertices. Since the maximum distance between points in the same quarter is $\sqrt{(\frac{1}{2})^2 + (\frac{1}{2})^2} = \frac{1}{\sqrt{2}}$, each side of T has length at most $\frac{1}{\sqrt{2}}$, as required. □

We now give a proof of the general form of the Pigeonhole Principle. This is a *proof by contradiction*, where we suppose that the statement in question is in fact false. If we can deduce a contradiction from this assumption, then (since we believe mathematics to be free of contradictions) we conclude that the statement therefore cannot be false, and so must be true. This is a very common method of proof.

Proof. Suppose for a contradiction that there exist natural numbers k and n for which the statement does not hold. This means that we can arrange at least n pigeons into k pigeonholes in such a way that there

[1] Note that in this example we did not say explicitly what are the pigeons (the partial sums) and pigeonholes (the remainders modulo r). This is fine provided it is unambiguously clear to the reader.

[2] That is, we can only apply the Pigeonhole Principle to *indivisible* objects (i.e. those which only occur in integer quantities). For example, it is not true that if three pints of milk are poured into two cups, some cup must contain at least two pints, because it is quite reasonable to have 1.5 pints in each.

[3] Formally, this says that the rule for assigning pigeons to pigeonholes is a *well-defined function*.

are fewer than n/k pigeons in each pigeonhole. Fix such an arrangement and let x_i be the number of pigeons in the pigeonhole i for each $1 \leq i \leq k$. We then have $x_i < n/k$ for each i, and so the total number of pigeons is

$$\sum_{i=1}^{k} x_i < k \cdot n/k = n,$$

contradicting the fact that there are at least n pigeons in total. □

Finally, we deduce the original form of the Pigeonhole Principle from the general form. The original form can also be proved directly by a similar argument as above.

Proof. Apply Theorem 1.2 with $n+1$ and n in place of n and k respectively; the theorem follows since $\lceil \frac{n+1}{n} \rceil = 2$ for every $n \in \mathbb{N}$. □

2. Set Sizes, Inclusion-Exclusion and Products

Our notion of ‘size’ for finite sets is intrinsically connected with the existence of bijections between sets. Indeed, the process of counting (“one", “two", “three", “four", ...) establishes a bijection between the set of objects we wish to count and the set $\{1, \dots, n\}$ for some $n \in \mathbb{N}$, from which we conclude that the set has size n. This motivates the next definition.

Definition (Set sizes)**.** For a non-negative integer n, we say that a set X has size n if there exists a bijection $f : \{1, 2, \dots, n\} \to X$. If such an integer n exists we say that X is finite, otherwise X is infinite. For finite sets X we write $|X|$ to denote the size of X; this is frequently also referred to as the order of X or the cardinality of X.

In particular, this definition justifies our ability to write a set X of size n as $\{x_1, \dots, x_n\}$ where $x_1, \dots, x_n$ are all distinct. Indeed, in doing so we implicitly choose a bijection $f : \{1, 2, \dots, n\} \to X$ (whose existence is guaranteed by the definition of size) and set $x_i := f(i)$ for each i.

Our next result describes size inequalities between sets in terms of the existence of injective functions between the sets in question.

Proposition 2.1. *Let X and Y be finite sets. Then $|X| \leq |Y|$ if and only if there exists an injection $f : X \to Y$.*

Proof. Let $m = |X|$ and $n = |Y|$, and write $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$. Suppose first that $|X| \leq |Y|$, so $m \leq n$. Then $f : X \to Y$ given by $f(x_i) = y_i$ for each $1 \leq i \leq m$ is an injection. Now suppose that $|X| > |Y|$, and consider an arbitrary function $g : X \to Y$. Then the Pigeonhole Principle (with the elements of X as pigeons and the elements of Y as pigeonholes) implies that g must map two distinct elements of X to the same element of Y, so g is not injective. Since g was arbitrary, this shows that there is no injection from X to Y, completing the proof. □

The next result is similar, but describes an equality between set sizes in terms of the existence of a bijection. This key result is known as the Pairing Principle, and will appear frequently in this course and beyond as a way of counting large finite sets, since it gives a natural way to show that two sets have the same size, namely to find a bijection between them.

Proposition 2.2 (Pairing Principle)**.** *Let X and Y be finite sets. Then $|X| = |Y|$ if and only if there exists a bijection $f : X \to Y$.*

Proof. Suppose first that $|X| = |Y|$, and let n denote this common size. Then by definition of size there exist bijections $g : \{1, 2, \dots, n\} \to X$ and $h : \{1, 2, \dots, n\} \to Y$. Since the inverse of a bijection is also a bijection, the inverse function $g^{-1} : X \to \{1, 2, \dots, n\}$ is a bijection, and so the composition $f = h \circ g^{-1} : X \to Y$ is a bijection, as required (since the composition of two bijections is also a bijection).

Now suppose instead that there exists a bijection $g : X \to Y$, and note that the inverse function $g^{-1} : Y \to X$ is also a bijection. Since every bijection is an injection, it follows by Proposition 2.1 that we have both $|X| \leq |Y|$ and $|Y| \leq |X|$, from which we conclude that $|X| = |Y|$. □

2.1 Sizes of unions

Our first application of the Pairing Principle is to prove the following sum rule, which specifies the size of the union $C \cup D$ of finite sets C and D which are *disjoint* (recall that this means that C and D have no elements in common). Note that the theorem does *not* hold for sets which are not disjoint.

Theorem 2.3 (Sum rule). *If C and D are disjoint finite sets then $|C \cup D| = |C| + |D|$.*

Proof. Let $m = |C|$ and $n = |D|$. Then by definition of size there exist bijections $f : \{1, 2, \dots, m\} \to C$ and $g : \{1, 2, \dots, n\} \to D$. We can use these to define a bijection[1][2] $h : \{1, 2, \dots, m + n\} \to C \cup D$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \leq m, \\ g(x - m) & \text{if } x > m. \end{cases}$$

The existence of this bijection proves that $|C \cup D| = m + n = |C| + |D|$. □

As an immediate corollary of the sum rule, we get the following complement rule, which relates the sizes of a set and its complement (for a given ground set, without which the notion of complement is not defined).

Corollary 2.4 (Complement rule). *Let U be a finite set. Then, taking complements with respect to the ground set U, for every set $A \subseteq U$ we have $|A^c| = |U| - |A|$.*

Proof. By definition A^c is disjoint from A, and $A \cup A^c = U$. So by the sum rule we have $|U| = |A \cup A^c| = |A| + |A^c|$, and rearranging gives the desired conclusion. □

If sets A and B are not disjoint then we cannot apply the sum rule to find $|A \cup B|$ from $|A|$ and $|B|$. Instead we can calculate this by the following two-set inclusion-exclusion formula, *provided* that we also know the size of $|A \cap B|$.

Theorem 2.5 (Inclusion-exclusion for two sets). *Suppose that A and B are finite sets. Then*

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

Proof. Note that A and $B \setminus A$ are disjoint sets whose union is $A \cup (B \setminus A) = A \cup B$.[3] So by Theorem 2.3 applied with $C = A$ and $D = B \setminus A$ we have

$$|A \cup B| = |A \cup (B \setminus A)| = |A| + |B \setminus A|. \tag{1}$$

Next note that $B \cap A$ and $B \setminus A$ are disjoint sets whose union is $(B \cap A) \cup (B \setminus A) = B$. So by Theorem 2.3 applied with $C = B \cap A$ and $D = B \setminus A$ we have

$$|B| = |(B \cap A) \cup (B \setminus A)| = |B \cap A| + |B \setminus A|. \tag{2}$$

Combining (1) and (2) completes the proof. □

The use of the inclusion-exclusion formula is that it allows us to calculate information we don't know from information that we have available, as in the following example. Most commonly, as in the following example, we will know the sizes of the sets and their intersections and wish to calculate the size of the union, but other scenarios are possible, such as calculating $|B|$ from $|A|$, $|A \cap B|$ and $|A \cup B|$.

[1] This is where the assumption that C and D are disjoint is used; if C and D intersect then h would not be injective.

[2] Many arguments in this course proceed by claiming that a certain function is a bijection (or injection or surjection). In cases where it is not straightforward to prove this property a proof will be included, but in simpler cases, like here, this is omitted to avoid distraction from the main argument. In all such cases you should consider it an exercise to justify that the function is indeed a bijection/injection/surjection as claimed (and please ask for help if unsure).

[3] Similarly as in the previous comment, straightforward set equalities like this and the equality in the line following (1) are stated without proof to avoid distraction from the main argument, but you should check as an exercise that you can justify the statement by showing that every element of the LHS set is an element of the RHS set and vice versa.

Example. If Facebook tells us that I have 155 friends, you have 274 friends, and we have 25 mutual friends, how many friends do we have between us?

Solution. We can describe this scenario in set terms: let A be the set of my friends and B be the set of your friends. Then $|A| = 155$ and $|B| = 274$. Also, $A \cap B$ is the set of people who are my friend and your friend, that is, a mutual friend, so $|A \cap B| = 25$. The set of people who are my friend or your friend is $A \cup B$, so applying the inclusion-exclusion formula we find that the number of such people is

$$|A \cup B| = |A| + |B| - |A \cap B| = 155 + 274 - 25 = 404.$$

So we have 404 friends between us. □

There are similar inclusion-exclusion formulae for larger numbers of finite sets. We next give the form for three sets, in which to find the size of the union we sum the set sizes, then subtract the sizes of the pairwise intersections, then add the three-way intersection.

Theorem 2.6 (Inclusion-exclusion for three sets)**.** *Suppose that A, B and C are finite sets. Then*

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

To prove this theorem we apply Theorem 2.5 (the inclusion-exclusion formula for two sets) three times, and also make use of a distributivity law for sets, namely that for any sets A, B and C we have $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof. Let $Y = B \cup C$. Then Theorem 2.5 implies that

$$|A \cup B \cup C| = |A \cup Y| = |A| + |Y| - |A \cap Y| \tag{3}$$

However, the distributivity law above and then Theorem 2.5 imply that

$$\begin{aligned}|A \cap Y| &= |A \cap (B \cup C)| = |(A \cap B) \cup (A \cap C)| \\ &= |A \cap B| + |A \cap C| - |(A \cap B) \cap (A \cap C)| \\ &= |A \cap B| + |A \cap C| - |A \cap B \cap C| \end{aligned} \tag{4}$$

Also, Theorem 2.5 implies that

$$|Y| = |B \cup C| = |B| + |C| - |B \cap C| \tag{5}$$

Now the result follows by substituting (4) and (5) into (3). □

One significant application of this formula is in counting integers which are divisible by one of several specified integers, as in the following example.

Example. How many integers between 1 and 1000 are divisible by at least one of the integers 2, 3 and 5?

Solution. Define sets A, B and C as follows.

$$\begin{aligned} A &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 2\}, \\ B &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 3\}, \\ C &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 5\}. \end{aligned}$$

Then

$$\begin{aligned} A \cap B &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 6\}, \\ A \cap C &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 10\}, \\ B \cap C &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 15\}, \\ A \cap B \cap C &= \{n \in \mathbb{Z} : 1 \leq n \leq 1,000 \text{ and } n \text{ is divisible by } 30\}. \end{aligned}$$

For any integer r, the number of integers between 1 and 1,000 which are divisible by r is equal to $\lfloor \frac{1000}{r} \rfloor$. So $|A| = 500$, $|B| = 333$, $|C| = 200$, $|A \cap B| = 166$, $|A \cap C| = 100$, $|B \cap C| = 66$, $|A \cap B \cap C| = 33$.

The set of integers divisible by at least one of 2, 3 and 5 is $A \cup B \cup C$, so by the inclusion-exclusion formula the number of such integers is

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C| \\ &= 500 + 333 + 200 - 166 - 100 - 66 + 33 = 734. \end{aligned}$$

□

The versions of this formula for two and three sets give a hint of the general formula which applies for any number of sets: we add the sizes of the given sets, then subtract the sizes of their pairwise intersections. Then we add back the sizes of the three-way intersections, before subtracting the sizes of the four-way intersections, and so forth until all intersections have been included in the calculation.

Theorem 2.7 (General inclusion-exclusion formula). *Suppose that $A_1, A_2, \dots, A_r$ are finite sets. Then*

$$\begin{aligned} |A_1 \cup A_2 \cup \cdots \cup A_r| =& |A_1| + |A_2| + \cdots + |A_r| \\ & - (|A_1 \cap A_2| + |A_1 \cap A_3| + \cdots + |A_{r-1} \cap A_r|) \\ & + (|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + \dots) \\ & - (|A_1 \cap A_2 \cap A_3 \cap A_4| + \dots) \\ & \dots \\ & \pm |A_1 \cap A_2 \cap \cdots \cap A_r|, \end{aligned}$$

where the sum in the jth line of the right hand side consists of the sizes of all intersections of j sets, and the sign in the final line is '$+$' if r is odd and '$-$' if r is even. Alternatively, the statement can be written in the following more compact form

$$\left| \bigcup_{i=1}^{r} A_i \right| = \sum_{\substack{J \subseteq \{1,\dots,r\} \\ J \neq \emptyset}} (-1)^{|J|+1} \left| \bigcap_{i \in J} A_i \right|,$$

where the sum is taken over all non-empty subsets of $\{1, 2, \dots, r\}$.

Proof (exercise). Giving a proof of this theorem is an excellent exercise in mathematical communication, so only the following outline is provided here (a full model solution will be provided in due course).

You might like first to try to prove the theorem for four sets. Start by writing out exactly what this says for sets A, B, C and D, and then try to prove it by a similar argument to the proof for three sets, starting by setting $Y = B \cup C \cup D$. You will need to use the two-set inclusion-exclusion formula once and the three-set formula twice, as well as the distributivity law $A \cap (B \cup C \cup D) = (A \cap B) \cup (A \cap C) \cup (A \cap D)$.

Having understood the proof for three sets and for four sets, you might expect that a similar argument can be used to deduce the $r+1$-set version from the r-set version, and indeed this is the case. This allows the general form to be proved by induction on r, where the aforementioned deduction gives the inductive step; you also need the distributive law

$$(A_1 \cup \cdots \cup A_k) \cap A_{k+1} = (A_1 \cap A_{k+1}) \cup \cdots \cup (A_k \cap A_{k+1}).$$

You will find that the main difficulty is notational – there are a lot of terms to keep track of, and you need to do this in an understandable way. This is true whether you work with the general formula in its first form (with '. . . ' everywhere) or second form (with set notation), but I recommend the latter as being somewhat easier to manage. □

2.2 Sizes of products

For a non-negative integer r, an ordered r-tuple is a sequence of r objects, which we refer to as elements or coordinates, in a given order. We write $(x_1, x_2, \dots, x_r)$ for the ordered r-tuple which has elements $x_1, \dots, x_r$ in the order presented. Unlike for a set $\{x_1, \dots, x_r\}$ of r elements, in the ordered r-tuple $(x_1, x_2, \dots, x_r)$ the order of the elements matters, and repeated elements are allowed. So, for example, $(1, 2, 3) \neq (1, 3, 2)$, and $(1, 1, 1)$ is a valid ordered 3-tuple. We refer to ordered 2-tuples as ordered pairs, ordered 3-tuples as ordered triples, and so forth.

Definition (Cartesian product)**.** The Cartesian product[4] of two sets A and B is defined by

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

In other words, $A \times B$ is the set of ordered pairs whose first coordinate is a member of A and whose second coordinate is a member of B. For example, $\{1,2\} \times \{2,3,4\} = \{(1,2),(1,3),(1,4),(2,2),(2,3),(2,4)\}$. We also make a similar definition which applies to collections of more than two sets.

Definition (Cartesian product, r sets)**.** For each non-negative integer r, the Cartesian product of sets $A_1, \dots, A_r$, denoted $\bigtimes_{i=1}^r A_i$ or $A_1 \times \cdots \times A_r$, is defined by

$$\bigtimes_{i=1}^r A_i = A_1 \times \cdots \times A_r = \{(x_1, \dots, x_r)\colon x_j \in A_j \text{ for each } j \in \{1, 2, \dots, r\}\}$$

So $A_1 \times A_2 \times \cdots \times A_r$ is the set of all ordered r-tuples whose jth coordinate is a member of A_j for each j. As with multiplication, given a set A and a non-negative integer n we define

$$A^n = \overbrace{A \times A \times \cdots \times A}^{n \text{ copies of } A}.$$

This is consistent with the definitions of $\mathbb{R}^2$, $\mathbb{R}^3$ etc. with which you are already familiar.

One subtle consequence of the definition of Cartesian product is that, given sets A, B and C, the products $A \times B \times C$ and $(A \times B) \times C$ are not quite the same thing: the first is a set of ordered triples (a, b, c), whilst the second is a set of ordered pairs $((a, b), c)$ whose first coordinate is an ordered pair. However, there is a natural bijection between the two sets given by

$$(a, b, c) \longleftrightarrow ((a, b), c).$$

Using this correspondence, the sets $A \times B \times C$ and $(A \times B) \times C$ are essentially equivalent for most purposes (in particular, they have the same size by the pairing principle); an analogous bijection shows that the same is true of $A_1 \times A_2 \times \cdots \times A_{r-1} \times A_r$ and $(A_1 \times A_2 \times \cdots \times A_{r-1}) \times A_r$.

The product rule tells us the size of the Cartesian product of a collection of sets. Note that unlike for the sum rule, we do not require that the sets are disjoint.[5]

Theorem 2.8 (Product rule for two sets)**.** *If A and B are finite sets, then $|A \times B| = |A| \cdot |B|$.*

Proof. Let $m = |A|$ and $n = |B|$, and write $A = \{a_1, \dots, a_m\}$ and $B = \{b_1, \dots, b_n\}$. We can then list the elements of $A \times B$ as

$$\left\{ \begin{array}{cccc} (a_1, b_1), & (a_1, b_2), & \dots & (a_1, b_n), \\ (a_2, b_1), & (a_2, b_2), & \dots & (a_2, b_n), \\ \dots & \dots & \dots & \dots \\ (a_m, b_1), & (a_m, b_2), & \dots & (a_m, b_n) \end{array} \right\}$$

So altogether there are m rows and n columns in the table, so it has $m \cdot n = |A||B|$ entries; since these are all distinct, we have $|A \times B| = |A||B|$. □

Using another induction argument we now use the 2-set product rule to give a product rule for any number of finite sets.

[4] We will often say just "product" for short.

[5] You might object that the proof of this theorem relies on the unproven assumption that a $m \times n$ grid has mn elements. Whilst I'm sure you are intuitively happy with this fact, formally this is a reasonable objection. If this bothers you, then try to give instead a proof by induction on k of the following equivalent statement: for every non-negative integer k, if A and B are finite sets with $|A| = k$, then $|A \times B| = k \cdot |B|$.

Theorem 2.9 (Product rule for r sets). *For any $r \in \mathbb{N}$ and any finite sets $A_1, A_2, \dots, A_r$ we have*

$$|A_1 \times A_2 \times \cdots \times A_r| = |A_1||A_2| \dots |A_r|.$$

Proof. We proceed by induction on r. The base case $r = 1$ is a tautology; it states simply that $|A_1| = |A_1|$.

Now suppose that the statement holds for $r = k$ for some $k \in \mathbb{N}$, that is, that for any finite sets $A_1, A_2, \dots, A_k$ we have $|A_1 \times A_2 \times \cdots \times A_k| = |A_1||A_2| \dots |A_k|$. Then, given any finite sets $A_1, \dots, A_{k+1}$, we have

$$\begin{aligned} |A_1 \times A_2 \times \cdots \times A_{k+1}| &= |(A_1 \times A_2 \times \cdots \times A_k) \times A_{k+1}| \\ &= |A_1 \times A_2 \times \cdots \times A_k||A_{k+1}| \\ &= |A_1||A_2| \dots |A_{k+1}|, \end{aligned}$$

where the first equality holds due to the natural bijection discussed earlier, the second equality holds by Theorem 2.8 (the product rule for two sets) applied to the sets A_{k+1} and $A_1 \times \cdots \times A_k$, and the final equality holds by the inductive hypothesis (that is, our assumption that the statement holds for k). So the statement holds for $r = k + 1$ also.

Having proved that the statement holds for $r = 1$, and that if it holds for $r = k$ then it also holds for $r = k + 1$, we conclude that it holds for every $r \in \mathbb{N}$, as required. □

One significant application of the product rule is in counting the number of factors of an integer, as in the following example.

Example. How many positive integers are factors of 1200?

Solution. The prime factorisation of 1200 is $2^4 \cdot 3 \cdot 5^2$, so the factors of 1200 are the integers of the form $2^a \cdot 3^b \cdot 5^c$ for $0 \leq a \leq 4$, $0, \leq b \leq 1$ and $0 \leq c \leq 2$ (this is a consequence of uniqueness of prime factorisation, which likewise implies that the integers of this form are all distinct). Let S be the set of positive factors of 1200, and let $A = \{0, 1, 2, 3, 4\}$, $B = \{0, 1\}$ and $C = \{0, 1, 2\}$ (so A, B and C are the possible values for a, b and c when forming a factor). Then the function $f : A \times B \times C \to S$ given by $f((a, b, c)) = 2^a 3^b 5^c$ is a bijection. By the pairing principle and product rule, it follows that the number of factors is $|S| = |A \times B \times C| = |A| \cdot |B| \cdot |C| = 5 \cdot 2 \cdot 3 = 30.$ □

Another important application is to count the number of subsets of a set of a given size, and the following definition is useful for describing this.

Definition (Power Set). The power set of a set A is defined by $\mathscr{P}(A) = \{X : X \subseteq A\}$.

So $\mathscr{P}(A)$ is a *set* whose *elements* are the *subsets* of A. For example, if $A = \{1, 2, 3\}$ then $\mathscr{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Note that for any set A, both $\emptyset$ and A are elements of the power set $\mathscr{P}(A)$.

Theorem 2.10. *If A is a set with $|A| = n$, then $|\mathscr{P}(A)| = 2^n$. Equivalently, a set with n elements has 2^n subsets.*

Proof. Write $A = \{x_1, \dots, x_n\}$. Observe that the elements of $\{0, 1\}^n$ are ordered n-tuples whose coordinates are each either 0 or 1. Define a function $f : \{0, 1\}^n \to \mathscr{P}(A)$ with rule $f((y_1, \dots, y_n)) = \{x_i : y_i = 1\}$. Then f is a bijection from $\{0, 1\}^n$ to $\mathscr{P}(A)$, so $|\mathscr{P}(A)| = |\{0, 1\}^n|$ by the pairing principle. By the product rule we have $|\{0, 1\}^n| = |\{0, 1\}|^n = 2^n$, completing the proof.[6] □

[6] The ordered n-tuple of zeros and ones corresponding to a subset $B \subseteq A$ in this proof is commonly called the characteristic vector of B. This is a useful way to represent subsets of a ground set which appears in many areas of mathematics.

3. Counting Choices

The next focus of the course is to develop a number of useful tools for counting large finite sets. The product rule, sum rule and inclusion-exclusion formulae are the starting point for our counting arguments, though we will often use these implicitly by 'counting choices'. Indeed, given a collection of sets, then the sum rule or inclusion-exclusion formulae tell us how many ways there are to choose one item from these sets (that is, a single item is taken from the union of all the sets, so only one item is chosen in total). The product rule tells us instead how many ways there are to choose one item from each set (so the number of items chosen in total is equal to the number of sets). The following example illustrates the difference between these situations.

Example. One society has 5 members, and another has 8 members; they have no members in common. How many ways are there to choose from the memberships

(i) a *single* representative for the societies (who can be from either)?
(ii) a representative from each society?

We first present a solution which makes use of the sum rule and product rule explicitly.

Solution. Let A be the set of members of the first society, and B the set of members of the second society. Then choosing a single representative for the societies means choosing an element of $A \cup B$, so the answer to (i) is $|A \cup B| = |A| + |B| = 5 + 8 = 13$ by the sum rule. Similarly, choosing a representative from each society means choosing a pair $(a, b) \in A \times B$, as a is then the representative for the first society and b is then the representative for the second. So the answer to (ii) is $|A \times B| = |A||B| = 5 \times 8 = 40$ by the product rule. □

Because these results and arguments are elementary to counting arguments, and used very frequently, you will become sufficiently familiar with them that it is not always necessary to state the use of the sum and product rule, or the sets involved. Instead we will usually present solutions in the manner of the following example, which presents the same argument as our previous solution, stated less formally.

Solution. For (i) there are 5 possible choices for the representative in the first society, and 8 possible choices for the representative in the second society, so in total there are 13 possible choices. On the other hand, for (ii) there are 5 choices for the first representative, and for each of these choices there are then 8 choices for the second representative, so in total there are $5 \times 8 = 40$ possibilities. □

However, it is vital when using this style to keep in mind that when 'counting choices' like this you are implicitly using the sum and/or product rules, and in particular to make sure you don't get confused as to whether to add or multiply.[1]

One major focus of this section is to establish expressions for how many possible ways there are to choose *several* elements from a single set. This depends on the rules for the choices; in particular:

- Are we allowed to choose the same element more than once? If so, the number of possibilities will be higher than if not. We say that "repetition is allowed" to mean that an element may be chosen multiple times, and that "repetition is forbidden" to mean that an element may be chosen at most once.
- Does the order in which the elements are chosen matter? That is, if we first choose a and then b, is this different from choosing first b and then a? If we consider these to be different, then there will be more possibilities than if we consider them to be the same.

[1] As an exercise, try writing the example from the previous section regarding the number of factors of 1200 in this style (by considering the choices for a, b and c).

Choosing uniformly at random. Several of our examples in this chapter will consider probabilities of events. These are naturally linked to counting results because, if a random experiment has a finite number of possible outcomes, and every outcome is equally likely, then the probability of an event E is

$$\mathbb{P}(E) = \frac{\text{number of outcomes for which } E \text{ occurs}}{\text{total number of outcomes}}.$$

We say that a selection is made uniformly at random if every outcome is equally likely (in which case we can apply the formula above). For example, rolling a fair standard 6-sided die selects an element of the set $\{1, 2, 3, 4, 5, 6\}$ uniformly at random.[2] Likewise, in the example above, if we choose uniformly at random a single representative for the societies, then each person has probability $\frac{1}{13}$ of being chosen.

Example. Let $X = \{1, 3, 5, 7, 9, 11, 13\}$. An element of X is selected uniformly at random; what is the probability that the chosen element is prime?

Solution. There are seven possible outcomes (the elements of X), of which 5 are prime, namely $3, 5, 7, 11, 13$. So the probability is $5/7$. □

Ordered choice

We first consider the situation when the order in which items are chosen *does* matter. In this case it is natural to consider the sequence formed by the options chosen at each step, and the first part of the next lemma reformulates the product rule to encapsulate our 'counting choices' approach in a way that is very useful for this scenario. The second part complements this by showing that if the individual choices are made uniformly at random, then the overall outcome is also chosen uniformly at random; this is the only probabilistic result we will use in this course (as opposed to using counting results to obtain probabilities). A full proof is included for completeness, but for the latter part this requires familiarity with conditional probabilities which are outside the scope of this course (this is, of course, non-examinable).

> **Lemma 3.1** (Counting Choices). *Let $r \in \mathbb{N}$, let $a_1, a_2, \dots, a_r$ be non-negative integers, and let X be a set. Suppose that we make r choices in turn, and that for each $1 \leq i \leq r$, regardless of what option was chosen at choices $1, 2, \dots, i-1$, the number of options for choice i is precisely a_i. Suppose moreover that for each $x \in X$ there is precisely one sequence of options which yields X. If for each $x \in X$ there is precisely one sequence of options which yields x, then $|X| = a_1 a_2 \dots a_r$.*
>
> *Moreover, if at each choice we choose an option uniformly at random regardless of the outcomes of previous choices[3], then the sequence of options chosen yields a uniformly-random element of X.*

Proof. Arbitrarily order the options available at each choice that could be made, and write $A_i := \{1, 2, \dots, a_i\}$ for each $1 \leq i \leq r$. We may then define a function

$$f : A_1 \times A_2 \times \dots \times A_r \to X$$

by defining $f(x_1, x_2, \dots, x_r)$ to be the element of X yielded by choosing option x_i at the ith choice, for each $1 \leq i \leq r$. The assumption that for each $x \in X$ there is precisely one sequence of options which yields X implies that this function is a bijection, so by the pairing principle and product rule we find that $|X| = |A_1 \times A_2 \times \dots \times A_r| = |A_1||A_2| \dots |A_r| = a_1 a_2 \dots a_r$.

For the 'moreover' part, for each $1 \leq i \leq r$ let $x_i \in A_i$ be the number of the option we select (randomly) at choice i. We will prove by induction on k that for each $1 \leq k \leq r$ the ordered k-tuple $(x_1, x_2, \dots, x_k)$ is a uniformly-random element of $A_1 \times A_2 \times \dots \times A_k$; this suffices to complete the proof as we may take $k = r$.

[2] All random selections considered in this course will be made uniformly at random; in other courses you will see examples of non-uniform random selections.

[3] The point of this remark is that it's not enough to say that each choice is made uniformly at random. For example, if $A_1 = A_2 = \{1, 2\}$, and I choose an element $a_1 \in A_1$ uniformly at random and then choose the *same* element a_1 from A_2, then I have chosen an element uniformly at random from A_2 also – since both elements are equally likely to be chosen – but the ordered pair formed is not chosen uniformly at random, since the only possible outcomes are $(1, 1)$ and $(2, 2)$.

Our base case $k = 1$ holds since our first choice is made uniformly at random. It remains to establish the inductive step, so suppose that for some $1 \leq j \leq r-1$ the ordered j-tuple $(x_1, x_2, \dots, x_j)$ is a uniformly-random element of $A_1 \times A_2 \times \cdots \times A_j$. Fix a $(j+1)$-tuple $(b_1, b_2 \dots, b_j, b_{j+1}) \in A_1 \times A_2 \times \cdots \times A_j \times A_{j+1}$, let Y denote the event that $x_i = b_i$ for every $1 \leq i \leq j+1$, and let Z denote the event that $x_i = b_i$ for every $1 \leq i \leq j$. Our inductive hypothesis them implies that $\mathbb{P}(Z) = 1/\prod_{i=1}^{j} |A_i| = 1/a_1 a_2 \dots a_j$. Moreover, the condition of the lemma tells us that $\mathbb{P}(Y|Z) = 1/|A_{j+1}| = 1/a_{j+1}$, since no matter what the outcome of Z, the element x_{j+1} is chosen uniformly at random from A_{j+1}. Finally by definition of conditional probability we have $\mathbb{P}(Y) = \mathbb{P}(Y|Z)\mathbb{P}(Z) = 1/a_1 a_2 \dots a_{j+1}$. Since $(b_1, b_2, \dots, b_{j+1})$ was arbitrary, we conclude that the ordered $(j+1)$-tuple $(x_1, x_2, \dots, x_j, x_{j+1})$ is a uniformly-random element of $A_1 \times A_2 \times \cdots \times A_j \times A_{j+1}$, completing the inductive step and therefore the proof. □

We illustrate this formulation of 'counting choices' with two different solutions for the following example.

Example. I deal four cards, in order, from a standard 52-card deck. How many outcomes are there in which all four cards have different suits?

Solution. We can form such an outcome by making the following choices, in sequence.

- First, choose any card (52 options).
- Next, choose any card not from the same suit as the first card (39 options).
- Next, choose any card not from the same suit as the first or second cards (26 options).
- Finally, choose any card not from the same suit as any of the three previous cards (13 options).

Each outcome in which all four cards have different suits is formed by precisely one sequence of choices, so the number of such outcomes is $52 \cdot 39 \cdot 26 \cdot 13 = 685,464$. □

Solution. We can form such an outcome by making the following choices, in sequence.

- First, choose the suit of the first card (4 options).
- Next, choose the suit of the second card (3 options).
- Next, choose the suit of the third card (2 options). Note that the suit of the fourth card is now determined also.
- Now choose the first card, which must be from the chosen suit (13 options).
- Likewise choose the second card (13 options).
- Likewise choose the third card (13 options).
- Finally choose the fourth card in the same way (13 options).

Each outcome in which all four cards have different suits is formed by precisely one sequence of choices, so the number of such outcomes is $4 \cdot 3 \cdot 2 \cdot 13^4 = 685,464$. □

Very often, as in the first few choices of the second solution, we find the number of options declining by one at each choice, in which case the following notation is very useful.[4]

Definition. For each integer $n \geq 0$, we define $\underline{n}$ factorial, denoted $n!$, by $n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$ (so in particular $0! = 1$).

For example, $0! = 1$, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, $6! = 720$, etc.

We can now state the number of ways to make r successive choices, in order, from a set S.

Theorem 3.2. *Let $r, n \geq 0$ be integers, and let S be a set of size n.*

(i) There are n^r possible ways to make r successive choices from S if the order of choices matters and repetition is allowed. Moreover, if each choice from S is made uniformly at random regardless of the outcomes of previous choices, then each of these outcomes has equal probability $1/n^r$.

[4] Be careful not to mix up the mathematical and punctuational uses of the '!' symbol! For example, if you solve a problem and write 'the answer is 10!', do you mean 10 or 10 factorial?

(ii) For $r \leq n$ there are $\frac{n!}{(n-r)!}$ possible ways to make r successive choices from S if the order of choices matters but repetition is forbidden. Moreover, if each choice from S is made uniformly at random from those elements of S which have not previously been chosen, then each of these outcomes has equal probability $\frac{(n-r)!}{n!}$.

(iii) For $r > n$ it is not possible to make r successive choices from S if repetition is forbidden.

Proof. Lemma 3.1 implies both (i) and (ii). Indeed, for (i) we are making r successive choices with precisely n options for each choice (since we may choose any element of S), so there are n^r possible outcomes. Similarly, for (ii) we are making r successive choices with precisely $n - i + 1$ options for the ith choice (since the options are all of the n elements of S except for the $i - 1$ which have previously been chosen, so the number of possible outcomes is $\prod_{i=1}^{r} n - i + 1 = \frac{n!}{(n-r)!}$. In both cases the 'moreover' statements follow likewise from the 'moreover' part of Lemma 3.1.

Finally, for (iii) observe that if $r > n$ then, by the pigeonhole principle, whenever we make r choices from the n elements of S there must be some element which is chosen at least twice, so it is not possible to choose r elements of S without repetition. □

Definition. A permutation of a set S of size n is an ordered n-tuple in which each element of S appears once.

In other words, the permutations of a set S are the ways to put the elements of S in order. For example, the permutations of $\{1, 2, 3\}$ are

$$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), \text{ and } (3, 2, 1).$$

So there are $6 = 3!$ permutations of a set of three elements.

In the algebra half of this course you will see a different definition, namely that a permutation of a set Ω is a bijection $f : \Omega \to \Omega$. The word permutation is widely used with both these meanings, but with minimal confusion since essentially they are the same. Indeed, let $S = \{x_1, x_2, \dots, x_n\}$ be a set of size n (the elements of which we have put in order by labelling them $x_1, \dots, x_n$). Then

(i) $(x_1, x_2, \dots, x_n)$ is a permutation of S (in the n-tuple sense),

(ii) for each permutation $f : S \to S$ (in the bijection sense), the ordered n-tuple $(f(x_1), f(x_2), \dots, f(x_n)$ is a permutation of S (in this n-tuple sense), and

(iii) for each permutation $(y_1, \dots, y_n)$ of S (in the n-tuple sense), the function $f : S \to S$ given by $f(x_i) = y_i$ is a permutation of S (in the bijection sense).

In other words, once we have fixed some initial order on a set S, the permutations in the n-tuple sense correspond exactly to permutations in the bijection sense (and vice versa).[5]. It is also generally clear from the context which meaning is intended, as in one case a permutation is being considered as an ordered sequence and in the other it is a function.

Corollary 3.3. *A set S of n elements has $n!$ permutations.*

Proof. We may form a permutation of S by choosing n elements from S in turn, with repetition forbidden (that is, we may not pick the same element more than once. By Theorem 3.2(ii), applied with $r = n$, the number of possible ways to do this is $\frac{n!}{(n-n)!} = n!$. □

For example, there are $n!$ ways for n people to line up in a queue, since each order is a permutation of the set of people. We conclude this section with three more examples of ordered choice.

Example. How many anagrams are there of the word MATHS? In how many of these is 'T' immediately before 'H'?

[5] This means that though the two definitions of permutation appear quite different, and refer to different kinds of object (ordered tuples and functions), they are capturing the same notion in essence. Many of the deep links between different areas of mathematics rely on this kind of fundamental connection. We will see another example later in the course, namely the connection between equivalence relations and partitions.

Solution. Each anagram of MATHS is a permutation of the set $\{\mathrm{M}, \mathrm{A}, \mathrm{T}, \mathrm{H}, \mathrm{S}\}$, so by Corollary 3.3 there are $5! = 120$ anagrams. For the second part, note that the anagrams of MATHS in which 'T' is immediately before 'H' are precisely the permutations of the set $\{\mathrm{M}, \mathrm{A}, \mathrm{TH}, \mathrm{S}\}$ (i.e. we treat 'TH' as a single letter), and Corollary 3.3 tells us there are $4! = 24$ of these. □

Example. I roll four standard dice. What is the probability that the numbers obtained are consecutive?

Solution. If we imagine rolling the dice in turn, then we are making four choices from $\{1, 2, 3, 4, 5, 6\}$ in which order matters and where repetition is allowed. So by Theorem 3.2(i) there are $6^4 = 1296$ possible outcomes, each of which is equally likely. The outcomes for which the numbers obtained are consecutive are the permutations of $\{1, 2, 3, 4\}$, $\{2, 3, 4, 5\}$ and $\{3, 4, 5, 6\}$. Each of these sets has $4! = 24$ permutations by Corollary 3.3, so in total there are 72 such outcomes. So the probability is $72/1296 = 1/18$. □

Unordered choice without repetition

The common property of the examples of the previous section was that we cared about the order in which we made choices. This will often not be the case: for example, in the national lottery we do not care about the order in which the balls are drawn, only which balls are drawn. This is the distinguishing property of *unordered choice*, for which the next definition is crucial.[6]

Definition. Let n and r be non-negative integers with $0 \leq r \leq n$. Then the binomial coefficient of n and r, also called n choose r, and denoted $\binom{n}{r}$ is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!}.$$

Note that

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

For example,

$$\binom{10}{2} = \frac{10 \times 9}{2 \times 1} = 45.$$

$$\left(\text{Note this is much simpler than } \binom{10}{2} = \frac{10!}{2!\,8!} = \frac{3628800}{2 \cdot 40320} = \frac{3628800}{80640} = 45.\right)$$

$$\binom{14}{7} = \frac{14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{14}{7 \times 2} \times \frac{10}{5} \times \frac{12}{6} \times \frac{8}{4} \times \frac{9}{3} \times 13 \times 11$$

$$= 1 \times 2 \times 2 \times 2 \times 3 \times 13 \times 11 = 3432.$$

$$\left(\begin{array}{l}\text{Note the rearrangement to make the calculation simpler, especially if you are} \\ \text{not using a calculator.}\end{array}\right)$$

$$\binom{n}{n} = \binom{n}{0} = \frac{n!}{n!\,0!} = 1.$$

$$\binom{n}{n-1} = \binom{n}{1} = \frac{n!}{(n-1)!\,1!} = n.$$

$$\binom{n}{n-2} = \binom{n}{2} = \frac{n \times (n-1)}{2 \times 1} = \frac{n(n-1)}{2}.$$

[6] Please use $\binom{n}{r}$ instead of ${}^nC_r$ or $C(n, r)$; the latter forms are not universally recognised.

Theorem 3.4. *Let S be a set of size n, and let r be an integer with $0 \leq r \leq n$. Then there are $\binom{n}{r}$ possible ways to make r successive choices from S if the order of choice is irrelevant and repetition is forbidden. Equivalently, the number of subsets $R \subseteq S$ of size r is $\binom{n}{r}$. Furthermore, if each choice from S is made uniformly at random from those elements of S which have not previously been chosen, then each of these outcomes has equal probability $1/\binom{n}{r}$.*

To see that the first two statements are equivalent, recall that a set may not have repeated elements, and the order of the elements is immaterial.

Proof. By Theorem 3.2 there are $\frac{n!}{(n-r)!}$ ordered sequences of r elements of S with no repeated element. Each such sequence $(x_1, \dots, x_r)$ is a permutation of a subset $\{x_1, \dots, x_r\}$ of S of size r. Since a set of size r has precisely $r!$ permutations by Corollary 3.3, there is an $r!$-to-one correspondence between ordered sequences of r elements of S with no repeated element and subsets of S of size r. We deduce[7] that the number of subsets of S of size r is

$$\frac{1}{r!}\frac{n!}{(n-r)!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

Furthermore, by Theorem 3.2, if each choice from S is made uniformly at random from those elements of S which have not previously been chosen, then each ordered sequence of r elements of S with no repeated element is equally likely to occur, and so (ignoring the order of the sequence) each subset of S of size r is equally likely to result. □

Example. How many 5-card poker hands are there? How many contain two aces, a king, and two other cards (i.e. not an ace or king)? What is the probability that a random 5-card hand has this form?

Solution. The order of cards is irrelevant, and repetition is impossible (you can't have more than one of the same card), so the answer to the first part is the number of ways to choose 5 cards out of 52 without repetition and ignoring order, that is,

$$\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2\,598\,960.$$

For the second part, there are $\binom{4}{2}$ ways to choose two of the four aces in the deck, and for each of these there are $\binom{4}{1}$ ways to choose one of the four kings in the deck. Then, whatever we have chosen so far, there are $\binom{44}{2}$ ways to choose two cards from the 44 cards which are not an ace or a king. So (by the product rule) in total the number of such hands is

$$\binom{4}{2}\binom{4}{1}\binom{44}{2} = 6 \times 4 \times 946 = 22\,704.$$

In particular, the probability that a random 5-card hand contains 2 aces, a king, and two other cards, is

$$\frac{22\,704}{2\,598\,960} \approx 0.008735.$$

□

Example. In the National Lottery we choose 6 numbers from the set $S = \{1, \dots, 59\}$. The lottery draw then selects 6 numbers (i.e. a subset of S of size 6) uniformly at random. What is the probability of winning the jackpot by matching all six numbers? What is the probability of winning £25 by matching three numbers?[8]

[7] I hope the claim used here is intuitive, that if $f : A \to B$ is an $r!$-to-one correspondence, then $|A| = r!|B|$, is intuitive. For a formal proof observe that we can form a bijection from A to $B^* = B \times \{1, 2, \dots, r!\}$ by, for each $b \in B$, mapping each of the $r!$ elements of $f^{-1}(b)$ to a distinct element (b, x) of B^*. The pairing principle and product rule then imply that $|A| = |B^*| = r!|B|$.

[8] In October 2015 the rules of the National Lottery were changed; prior to this the six numbers were selected from the set $\{1, \dots, 49\}$. As an exercise, calculate the effect this change had on the winning probabilities in this example. Also, after the 6 balls have been drawn a 'bonus ball' is drawn, and the second-highest prize is to match five of the original six balls and also the bonus ball. As an exercise, calculate the probability of this event.

Solution. The order in which the balls are selected is irrelevant, and a ball cannot be selected more than once. So the number of possibilities for the six numbers drawn is $\binom{59}{6}$, and each outcome is equally likely. The probability that the outcome is the six numbers we chose is therefore

$$\frac{1}{\binom{59}{6}} = \frac{1}{45\,057\,474}.$$

For the second part, we count how many possible choices of six balls there are which match precisely three of the six balls we selected. For this, note that there are $\binom{6}{3}$ ways to choose three of our selected six balls, and for each of these choices there are $\binom{53}{3}$ ways to choose three of the 53 balls we did not select. So there are $\binom{6}{3}\binom{53}{3}$ possible outcomes of the draw which match exactly three of our selected balls, so the probability that this event occurs is

$$\frac{\binom{6}{3}\binom{53}{3}}{\binom{59}{6}} = \frac{20 \times 23\,426}{45\,057\,474} = \frac{468\,520}{45\,057\,474} \approx 0.0104 \qquad \text{(around 1 in 96.17).}$$

□

Unordered choice with repetition

Finally, we consider the case where repetition is allowed but we don't care about the order of choices.[9]

> **Theorem 3.5.** *Let S be a set of size n. Then there are $\binom{n+r-1}{r}$ possible ways to make r successive choices from S if repetition is allowed but the order of choosing is irrelevant.*

Proof. Suppose that $S = \{s_1, s_2, \ldots, s_n\}$. Imagine $n + r - 1$ positions, represented by crosses. For example, for $n = 6$, $r = 4$ we have

$$\times \; \times \; \times \; \times \; \times \; \times \; \times \; \times \; \times$$

Choose $n - 1$ of these and circle them, for example.

$$\times \; \otimes \; \otimes \; \times \; \otimes \; \otimes \; \times \; \times \; \otimes$$

There are $\binom{n+r-1}{n-1} = \binom{n+r-1}{r}$ possibilities for which $n - 1$ crosses are circled. Crucially, there is a bijection[10] between the set of ways to choose $n - 1$ of the $n + r - 1$ crosses and the set of ways to make r choices from S allowing repetition but ignoring order. Specfically, the number of $\times$ before the first $\otimes$ tells you how many times to choose s_1, the number of $\times$ between the first and second $\otimes$ tells you how many times to choose s_2, the number of $\times$ between the second and third $\otimes$ tells you how many times to choose s_3, and so forth, until the number of $\times$ after the final $\otimes$ tells you how many times to choose s_n. For instance, in the example above we would choose s_1 once, s_3 once and s_5 twice. Similarly, if we had circled

$$\otimes \; \otimes \; \otimes \; \times \; \times \; \times \; \times \; \otimes \; \otimes$$

then we would choose s_4 four times and not choose any other member of S at all, and if we had circled

$$\times \; \times \; \otimes \; \otimes \; \times \; \otimes \; \times \; \otimes \; \otimes$$

[9] Be very careful when using Theorem 3.5 to calculate probabilities, as unlike in Theorem 3.2 and Theorem 3.4 these will often not be uniform. For example, if I roll two identical dice and ignore the order in which they are rolled, by Theorem 3.5 there are $\binom{6+2-1}{2} = 21$ distinct possible outcomes (try listing them to check this), but these are not equally likely. For instance, I am twice as likely to get a three and a four than I am to get two sixes. For this reason, when we repeat random experiments with repetition allowed, such as rolling dice or drawing balls from a bag with replacement, we almost always count ordered outcomes and use Theorem 3.2. On the other hand, as we have already seen, when we repeat random experiments with repetition forbidden, such as dealing cards from a deck or drawing balls without replacement, we can either count outcomes with order using Theorem 3.2 or count outcomes without order using Theorem 3.4; the latter is usually simpler.

[10] As an exercise, justify that this is indeed a bijection. That is, that for different choices of which crosses to circle we obtain different choices of members of S, and every possible way to choose r elements of s with repetition but ignoring order is mapped to by some choice of which crosses to circle.

then we would choose s_1 twice, s_3 once and s_4 once. On the other hand, choosing s_1 once, s_3 once and s_6 twice would be obtained by circling

$$\times \otimes \quad \otimes \quad \times \otimes \quad \otimes \quad \otimes \times \ \times.$$

□

Corollary 3.6. *For all $r, n \in \mathbb{N}$, the number of non-negative integer solutions of $x_1 + x_2 + \cdots + x_n = r$ is $\binom{n+r-1}{r}$.*

Proof. There is a bijection between the set of non-negative integer solutions of the equation and the set of ways to choose r elements of $\{1, 2, \dots, n\}$, allowing repetition but ignoring order. Indeed, the solution $(x_1, \dots, x_n)$ corresponds to the outcome in which for each $1 \leq i \leq n$ the element i is chosen x_i times. So by Theorem 3.5 the number of solutions is $\binom{n+r-1}{r}$. □

One reason this corollary is important is that it counts the number of ways of distributing r identical objects among n people, as in the next example.

Example. How many possible ways are there to share 7 pound coins among 5 people?

Solution. 330. Indeed, there are $\binom{5+7-1}{7} = \binom{11}{7} = 330$ solutions $(x_1, \dots, x_5)$ in non-negative integers to $x_1 + \cdots + x_5 = 7$, and these solutions correspond bijectively to distributions of 7 pound coins among 5 people, with x_i being the number of coins given to person i. □

Instead of using Theorem 3.5, it is also possible to prove Corollary 3.6 directly by a similar argument as used in the proof of Theorem 3.5: consider $n + r - 1$ positions, filled with the symbol $\times$. Then there is a bijection between the set of ways to choose $n - 1$ of the $\times$ to circle and the set of solutions to the equation: the value of x_1 is the number of $\times$ before the first $\otimes$, the value of x_2 is the number of $\times$ between the first and second $\otimes$, and so forth.[11]

We can solve similar problems with different restrictions on $x_1, x_2, \dots, x_n$ similarly, as in the following example.

Example. How many solutions are there in *positive* integers to $x_1 + x_2 + x_3 = 101$?

Solution. Let X be the set of positive integer solutions (x_1, x_2, x_3) of $x_1 + x_2 + x_3 = 101$ and let Y be the set of non-negative integer solutions (y_1, y_2, y_3) of $y_1 + y_2 + y_3 = 98$. Then the substitution $y_1 = x_1 - 1$, $y_2 = x_2 - 1$ and $y_3 = x_3 - 1$ gives a bijection between X and Y, since $x_1 + x_2 + x_3 = 101$ if and only if $y_1 + y_2 + y_3 = 98$, and y_i is a non-negative integer if and only if x_i is a positive integer. We conclude that the number of solutions is $|X| = |Y| = \binom{100}{98} = \binom{100}{2} = 4950$ by Corollary 3.6. □

To summarise the key results of this section, the following table displays the number of different possible ways of making r successive choices from a set of size n.

	with order	ignoring order
repetition allowed	n^r	$\binom{n+r-1}{r}$
repetition forbidden	$\frac{n!}{(n-r)!}$	$\binom{n}{r}$

More sophisticated examples

In this section we consider four examples of how our counting results from the previous sections may be combined to answer more complex questions, illustrating a number of important ideas and approaches. Our first example involves ‘mixed choice’, which is a situation when order ‘partly matters’. Indeed, in this example the order of different letters matters (‘MISSISSIPPI’ is a different anagram from ‘IMSSISSIPPI’), however, swapping the positions of two of the ‘S’s, say, yields the same anagram.

[11] Exercise: fill in the details in this argument.

Example. How many anagrams are there of 'MISSISSIPPI'?

Solution. There are $11!$ possible ways to put the elements of $\{\mathrm{M}, \mathrm{I}_1, \mathrm{S}_1, \mathrm{S}_2, \mathrm{I}_2, \mathrm{S}_3, \mathrm{S}_4, \mathrm{I}_3, \mathrm{P}_1, \mathrm{P}_2, \mathrm{I}_4\}$ in order, since the set has $11!$ permutations by Corollary 7.2.Ignoring the subscripts, these orders give all the anagrams of 'MISSISSIPPI'. However, each anagram of 'MISSISSIPPI' is formed $4!4!2!$ times in this manner: there are $4!$ possible arrangements of $\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3$ and I_4, for each of these there are $4!$ possible arrangements of $\mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3$ and S_4, and for each of these there are $2! = 2$ possible arrangements of P_1 and P_2. So in total the number of anagrams of 'MISSISSIPPI' is $\frac{11!}{4!4!2!} = 34\,650$.

An alternative approach is to consider a sequence of 11 blank spaces, into which we will place the letters of MISSISSIPPI. First, we decide where we are going to put the four 'S's. There are 11 spaces to choose from, so there are $\binom{11}{4}$ possibilities to choose from (since we are choosing four of the eleven empty spaces). For any of these choices, there there are then seven empty spaces remaining, so there are $\binom{7}{4}$ possibilities for how to place the four 'I's among these. There are then three empty spaces, so $\binom{3}{2}$ choices for how to place the two 'P's. Finally, there is now only one empty space, so the remaining letter (the 'M') must be placed here - there is no choice. We conclude that the number of anagrams of 'MISSISSIPPI' is

$$\binom{11}{4}\binom{7}{4}\binom{3}{2} = \frac{11 \times 10 \times 9 \times 8}{4!} \cdot \frac{7 \times 6 \times 5 \times 4}{4!} \cdot \frac{3 \times 2}{2!} = \frac{11!}{4!4!2!} = 34\,650,$$

as before. □

Note that in the first method we considered the problem with order (i.e. treating all characters as different), then 'divided out the overcounting'; that is, we divided to reflect the fact that we have counted order where we didn't want to (e.g. the order of the four 'M's). On the other hand, in the second method we viewed the mixed choice as a series of separate unordered choices. In general both of these methods will work for mixed choice problems, but one may be significantly simpler than the other, depending on the problem.

Example. A hat contains the names of 16 teams, 8 from League A and 8 from League B. The names are drawn out one-by-one, without replacement. The first team drawn then plays against the second team drawn, the third team drawn plays against the fourth team drawn, and so forth (so there are 8 matches in total). What is the probability that every match includes one team from each league?

Solution. There are 16! possible orders in which the teams can be drawn. We now count the number of orders which have the property that every match includes one team from each league. For each of the 8 matches there are two possibilities: either the team from League A was drawn first, or the team from League B was drawn first. Having fixed this for each match, there are 8! orders in which the teams from League 8 could be drawn (one to each match), and 8! orders in which the teams from League B could be drawn, giving a total of $2^8(8!)^2$ orders. So the probability is

$$\frac{2^8(8!)^2}{16!} = \frac{2^8(8!)}{9 \times 10 \times 11 \times 12 \times 13 \times 14 \times 15 \times 16} = \frac{2^7}{5 \times 9 \times 11 \times 13} = \frac{128}{6435} \approx 0.0199.$$

□

This example illustrates the potential for making use of Corollary 7.2 – the number of permutations of a set – for a range of counting problems. Observe also how we formed an outcome of the draw in which each tie has one team from each league by a sequence of choices, first picking which team in each tie is from which league, and then choosing, for each league, which team to assign to each tie.

Example. I roll seven standard dice. What is the probability that I get at least three sixes?

Solution. If we imagine rolling the dice in order, there are 6^7 possible outcomes, each of which is equally likely. We now count the outcomes in which I get at most two sixes (that is, the outcomes which don't satisfy our condition). There are 5^7 outcomes in which I roll no sixes, since in this case there are five possibilities (1, 2, 3, 4 or 5) for each die. There are 7×5^6 outcomes in which I roll one six, since there are 7 possibilities for which die shows six and 5 possibilities for each of the other six dice, and likewise there are $\binom{7}{2} \times 5^5$ outcomes in which I roll two sixes, since there are $\binom{7}{2}$ ways to choose the two dice which show six and 5 possibilities for each of the other five dice. So there are $5^7 + 7 \times 5^6 + \binom{7}{2} \times 5^5 = 81 \times 5^5$ outcomes in which I do not roll at least three sixes, and therefore there are $6^7 - (81 \times 5^5)$ outcomes in which I do roll at least three sixes. The probability of this event is therefore

$$\frac{6^7 - (81 \times 5^5)}{6^7} = \frac{26811}{279936} \approx 0.095775.$$

□

Observe that in this example, even though the question doesn't specify order on the rolls of the dice, we introduced an order on the rolls (by imagining them being rolled in some order). This is necessary for the fact that each outcome is equally likely – as discussed in the previous section, if we choose successive elements from a set uniformly at random, allowing repetition but ignoring order, then the outcome is not uniformly random among all possible outcomes. Because of this, introducing order as in this example is a standard approach.

Note also the use of the complement: rather than count the outcomes in which we got at least three sixes, we counted the outcomes in which we got at most two sixes, and subtracted from the total number of outcomes. Formally, if U is the set of all outcomes, and A is the set of outcomes in which we got at least three sixes, we counted $|A^c|$ (where A^c is the complement of A with respect to the ground set U), and then used the fact that $|A| = |U| - |A^c|$; this latter fact is just a rearrangement of $|A| + |A^c| = |U|$, which holds by the Sum Rule.

It is also worth noting how we handled the inequality in the question: our counting results don't adapt well to this type of inequality, so instead we considered each case satisfying the inequality separately. That is, we could have counted separately the outcomes with three, four, five, six and seven sixes, but instead we used the complement and counted separately the outcomes with zero, one or two sixes.

Once again we see the approach of forming an outcome by a sequence of choices. For example, when counting the outcomes where we rolled two sixes, we first chose which two rolls were the ones which showed six, and then chose the values shown by the other five dice.

Example. I deal a hand of six cards from a standard 52-card deck.

(i) What is the probability that I get two cards of one suit and four cards of another suit?
(ii) What is the probability that I get three cards of one suit and three cards of another suit?
(iii) What is the probability that I get exactly two aces and exactly three diamonds?

Solution. First note that there are $\binom{52}{6}$ possible 6-card hands, each of which is equally likely to be dealt. For (i) we count the number of hands consisting of two cards of one suit and four of another. There are 4 possibilities for the suit from which we have four cards, and then $\binom{13}{4}$ possibilities for what these four cards are. Having chosen these there are then 3 possibilities for the suit of the remaining two cards, and then $\binom{13}{2}$ possibilities for what these two cards are. So there are $4 \cdot \binom{13}{4} \cdot 3 \cdot \binom{13}{2}$ such hands, and we conclude that the probability is

$$\frac{4 \cdot \binom{13}{4} \cdot 3 \cdot \binom{13}{2}}{\binom{52}{6}} = \frac{669240}{20358520} \approx 0.0328727.$$

For (ii) we similarly count the number of hands consisting of three cards of one suit and three of another. There are $\binom{4}{2}$ ways to choose which two suits the cards come from. Having chosen these, there are $\binom{13}{3}$ ways to choose three cards from the first suit, and then $\binom{13}{3}$ ways to choose three cards from the second suit. So there are $\binom{4}{2} \cdot \binom{13}{3} \cdot \binom{13}{3}$ such hands, and we conclude that the probability is

$$\frac{\binom{4}{2} \cdot \binom{13}{3} \cdot \binom{13}{3}}{\binom{52}{6}} = \frac{490776}{20358520} \approx 0.0241067.$$

For (iii) we consider two cases. If I do not get the ace of diamonds, then I need to get two of the remaining three aces (there are $\binom{3}{2}$ possibilities for how this can be done), three of the remaining 12 diamonds (there are $\binom{12}{3}$ possibilities for how this can be done) and one of the 36 cards which is not a ace or a diamond (there are 36 possibilities for this card). So there are $36\binom{3}{2}\binom{12}{3}$ hands not containing the ace of diamonds with exactly two aces and exactly three diamonds. On the other hand, if I do get the ace of diamonds, then I need to get one of the remaining three aces (3 possibilities), two of the remaining 12 diamonds ($\binom{12}{2}$ possibilities) and two of the other 36 cards ($\binom{36}{2}$ possibilities). So there are $3\binom{12}{2}\binom{36}{2}$ hards containing the ace of diamonds with exactly two aces and exactly three diamonds. Overall this gives a probability of

$$\frac{36\binom{3}{2}\binom{12}{3} + 3\binom{12}{2}\binom{36}{2}}{\binom{52}{6}} = \frac{148500}{20358520} \approx 0.0072942.$$

□

Note carefully the difference between the solutions of the parts (i) and (ii) of this example. The fundamental reason is that in (i) the suits chosen play different roles (for example, having two spades and four diamonds is different from having two diamonds and four spades) whilst in (ii) they play the same role (for example, having three diamonds and three spades is the same as having three spades and three diamonds). As well as this, parts (i) and (ii) again illustrate the power of viewing an outcome as being formed by a sequence of choices to enable us to count them by counting the possible outcomes of these choices.

Part (iii) illustrates another important method by the way we handled the fact that the set of diamonds and the set of aces intersect. This makes a direct approach difficult as, for example, if we first choose the diamonds and then the aces, then the number of aces that we need to choose depends on which diamonds we chose. We avoided this difficulty by considering separate cases for which cards from the intersection are chosen; in this example this meant considering separately the hands containing the ace of diamonds and the hands not containing the ace of diamonds.

4. The Binomial Theorem and its Consequences

In this section we will prove the Binomial Theorem and explore some of its corollaries. We begin with the following lemma, which gives an important identity for binomial coefficients.

Lemma 4.1. *For all integers r and n with $0 \leq r < n$ we have*

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}.$$

We give two alternative proofs of this lemma, the first by algebraic manipulation and the second by a combinatorial argument. This choice of approaches is common when proving results of this type.

Proof. (Algebraic manipulation.) One approach is to write $\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$, and express the other terms similarly. We then have

$$\begin{aligned}\binom{n}{r} + \binom{n}{r+1} &= \frac{n!}{r!\,(n-r)!} + \frac{n!}{(r+1)!\,(n-(r+1))!}\\ &= \frac{n!}{(r+1)!}\left(\frac{(r+1)}{(n-r)!} + \frac{1}{(n-r-1)!}\right)\\ &= \frac{n!}{(r+1)!\,(n-r)!}\big((r+1)+(n-r)\big)\\ &= \frac{n!(n+1)}{(r+1)!\,((n+1)-(r+1))!} = \binom{n+1}{r+1}.\end{aligned}$$

□

The following, more elegant approach instead uses the fact that a set of size n has $\binom{n}{r}$ subsets of size r.

Proof. (Combinatorial argument.) Let N denote the set $\{1, \ldots, n\}$. Then

$$\begin{aligned}\binom{n+1}{r+1} &= |\{S \subseteq N \cup \{0\} : |S| = r+1\}|\\ &= |\{S \subseteq N \cup \{0\} : |S| = r+1 \text{ and } 0 \in S\}|\\ &\quad + |\{S \subseteq N \cup \{0\} : |S| = r+1 \text{ and } 0 \notin S\}|\\ &= |\{T \subseteq N : |T| = r\}|\\ &\quad + |\{T \subseteq N : |T| = r+1\}|\\ &= \binom{n}{r} + \binom{n}{r+1}.\end{aligned}$$

Here the second equality holds by the sum rule since each subset S either contains 0 or doesn't contain 0, whilst the third inequality holds since the second term in each sum counts the same set, whilst the sets in the first term of each sum correspond bijectively by taking $T = S \setminus \{0\}$. □

One consequence of Lemma 4.1 is that the binomial coefficients $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ are the numbers of the nth row of Pascal's triangle, since they are generated by the same rules. We next state and prove the Binomial Theorem.

Theorem 4.2 (Binomial theorem). *For every integer $n \geq 0$ and all $a, b \in \mathbb{R}$, we have*

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

This theorem justifies why the numbers $\binom{n}{r}$ are called binomial coefficients: they are the coefficients in the binomial expansion $(a+b)^n$. As with the previous lemma, there are two principle ways to prove this theorem; the first is a nice 'combinatorial' argument using the fact that $\binom{n}{r}$ is the number of ways to choose r elements from a set of size n, ignoring order and without repetition, whilst the second proceeds by induction using Lemma 4.1 for the inductive step.

Proof. (Combinatorial argument.) If we consider multiplying out $(a+b)^n$, it is easy to see that we must have

$$(a+b)^n = \sum_{i=0}^{n} t_i a^i b^{n-i}$$

for some $t_0, t_1, \ldots, t_n \in \mathbb{N}$. (Indeed, when multiplying out, from each of the n brackets we take either the 'a' or the 'b' term, so their exponents always add up to n). There are n brackets we multiply out and the term $a^i b^{n-i}$ arises precisely when we choose the 'a' term from exactly i of the brackets (and thus the 'b' term from all the other $n-i$ brackets). There are $\binom{n}{i}$ ways of doing this. So we must have $t_i = \binom{n}{i}$. □

Proof. (By induction.) For each integer $n \geq 0$, let P_n denote the statement that for all $a, b \in \mathbb{R}$, we have

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

Then P_0 holds since for any $a, b \in \mathbb{R}$ we have $(a+b)^0 = 1 = a^0 = b^0 = \binom{n}{0}$. Now suppose that P_k holds for some integer $k \geq 0$. Then

$$\begin{aligned}
(a+b)^{k+1} &= (a+b)(a+b)^k \\
&= (a+b)\left(\sum_{i=0}^{k} \binom{k}{i} a^i b^{k-i}\right) \\
&= \sum_{i=0}^{k} \binom{k}{i} a^{i+1} b^{k-i} + \sum_{i=0}^{k} \binom{k}{i} a^i b^{k+1-i} \\
&= \sum_{j=1}^{k+1} \binom{k}{j-1} a^j b^{k+1-j} + \sum_{i=0}^{k} \binom{k}{i} a^i b^{k+1-i} \\
&= a^{k+1} + b^{k+1} + \sum_{i=1}^{k} \left(\binom{k}{i-1} + \binom{k}{i}\right) a^i b^{k+1-i} \\
&= \binom{k+1}{k+1} a^{k+1} + \binom{k+1}{0} b^{k+1} + \sum_{i=1}^{k} \binom{k+1}{i} a^i b^{k+1-i} \\
&= \sum_{i=0}^{k+1} \binom{k+1}{i} a^i b^{k+1-i},
\end{aligned}$$

where the fourth equality uses the substitution $j = i+1$, and the penultimate equality holds by Lemma 4.1. So P_{k+1} holds. So by the principle of mathematical induction, P_n holds for every integer $n \geq 0$. □

Corollary 4.3. *For every integer $n \geq 0$ we have*

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

Proof.

$$2^n = (1+1)^n = \sum_{i=0}^{n} \binom{n}{i} 1^i 1^{n-i} = \sum_{i=0}^{n} \binom{n}{i},$$

where the middle equality uses the Binomial Theorem with $a = b = 1$. □

Note that Corollary 4.3 gives another proof of Theorem 2.10, which stated that for every set X we have $|\mathscr{P}(X)| = 2^{|X|}$, or in other words, that a set X of size n has 2^n subsets. Indeed, if we let N denote the total number of subsets of X, and N_i denote the number of subsets of X which have size i, then

$$N = \sum_{i=0}^{n} N_i = \sum_{i=0}^{n} \binom{n}{i} = 2^n,$$

where the final equality holds by Corollary 4.3.

Corollary 4.4. *For every integer $n \geq 1$ we have*

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots \pm \binom{n}{n} = 0,$$

where the $\pm$ at the end is $+$ if n is even and $-$ if n is odd.

Proof.

$$\begin{aligned} 0 = 0^n = (-1+1)^n &= \sum_{i=0}^{n} \binom{n}{i} (-1)^i 1^{n-i} \\ &= \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots \pm \binom{n}{n}. \end{aligned}$$

□

Note that Corollary 4.4 does *not* hold for $n = 0$, as in this case we just have the term $\binom{n}{0} = (-1+1)^0 = 0^0 = 1$.

Corollary 4.5. *Let S be a set of size $n \geq 1$. Then S has 2^{n-1} subsets of even size and 2^{n-1} subsets of odd size.*

Proof. Corollary 4.4 says that any non-empty set S has the same number of subsets of even size as it does of odd size. Indeed if S has size n then the number of subsets of odd size is $\sum \binom{n}{r}$ where the sum ranges over all *odd* integers r between 0 and n, and the number of subsets of even size is $\sum \binom{n}{r}$ where the sum ranges over all *even* integers r between 0 and n; Corollary 4.4 says that these two sums are equal. Combining this with Theorem 2.10 (which says that S has 2^n subsets) gives the result. □

Again, this result does *not* hold for $n = 0$, since the empty set has one subset of even size (itself) and no subset of odd size.

5. Sizes of Infinite Sets

In this section we consider how our notion of set size can be extended to infinite sets. Recall that for finite sets A and B we have $|A| \leq |B|$ if and only if there is an injection from A to B, and $|A| = |B|$ if and only if there is a bijection from A to B. To extend the notion of 'size' to infinite sets we follow the same rules.

> **Definition.** The cardinality[1] of a set X, denoted $|X|$, is an attribute associated with X such that, for any sets X and Y,
>
> (a) $|X| \leq |Y|$ if and only if there exists an injection $f : X \to Y$, and
> (b) $|X| = |Y|$ if and only if there exists a bijection $f : X \to Y$.

In particular, the cardinality of any finite set is its size in the sense which you are familiar with already, and so every non-negative integer is a cardinality of a set. For finite sets you are very familiar with the order of possible cardinalities (i.e. the standard order of non-negative integers). This extends to an order on all possible cardinalities of sets. By this[2] we mean that

(i) For all sets X, Y and Z, if $|X| \leq |Y|$ and $|Y| \leq |Z|$ then $|X| \leq |Z|$. In other words, if there exists an injection $f : X \to Y$ and an injection $g : Y \to Z$ then there exists an injection from X to Z, and this is true since $g \circ f$ is the required injection.

(ii) For all sets X and Y, if $|X| \leq |Y|$ and $|Y| \leq |X|$ then $|X| = |Y|$. In other words, if there exist injections from X to Y and from Y to X then there exists a bijection from X to Y. This is asserted by the Schröder-Bernstein theorem.

(iii) For all sets X and Y we must have $|X| \leq |Y|$ or $|Y| \leq |X|$. In other words there exists an injection from X to Y or an injection from Y to X. This is asserted by Zermelo's theorem.

Both the Schröder-Bernstein theorem and Zermelo's theorem are beyond the scope of this course, although neither is particularly long or technical, and I think both should be understandable by keen and interested students, so I recommend them to you for further study if you are interested. The important thing to take from the above is that cardinalities of sets follow the rules of order presented in (i), (ii) and (iii).

Example. $|\mathbb{N}| = |\mathbb{Z}|$, since we can define a bijection $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(n) := \begin{cases} \frac{n-1}{2} & \text{if } n \text{ is odd,} \\ -\frac{n}{2} & \text{if } n \text{ is even.} \end{cases}$$

[1] The word 'size' is often used in place of cardinality for infinite sets (just as for finite sets), and you should feel free to do this, but be careful as many infinite sets have alternative notions of size. For example, the size of the interval $[0, 5]$ in the real numbers is 5 under the standard measure, but it does not have cardinality 5 as it is infinite.

[2] Formally, $\leq$ is a *total order*, also called simply an *order*, of a set X if the following properties hold:

(i) For every $x, y, z \in X$, if $x \leq y$ and $y \leq z$ then $x \leq z$.

(ii) For every $x, y \in X$, if $x \leq y$ and $y \leq x$ then $x = y$.

(iii) For every $x, y \in X$, either $x \leq y$ or $y \leq x$.

It's easy to check that this holds for the standard orders on $\mathbb{Z}$ and $\mathbb{R}$, for example. Observe that the properties (i), (ii) and (iii) in the main text are exactly the properties (i), (ii) and (iii) here (respectively), restated in the context of set sizes. As well as total orders of sets, there is an extensive theory of *partial orders*, which satisfy (i) and (ii) but not (iii), meaning that you can have incomparable elements x, y with $x \not\leq y$ and $y \not\leq x$. The contents of this footnote are not within the scope of this course for the purposes of assessment.

Example. $|\mathbb{Z}| = |2\mathbb{Z}|$ (the latter denotes the set of even integers), since we can define a bijection $f : \mathbb{Z} \to 2\mathbb{Z}$ by $f(n) = 2n$.

Observe from these examples that, unlike for finite sets, a proper subset of an infinite set X can have the same cardinality as X. On the other hand, the next lemma shows that a subset of X cannot have larger cardinality than X.

Lemma 5.1. *For any sets X and Y with $X \subseteq Y$ we have $|X| \leq |Y|$.*

Proof. The function $i : X \to Y$ given by $i(x) = x$ for $x \in X$ is an injection. □

Our definition of cardinality refers to injections and bijections, but the following lemma allows us to give another formulation in terms of surjections (Corollary 5.3), which is often useful. To understand the separate case of X being empty in the statement, note that for any set Y there is precisely one injection $f : \emptyset \to Y$, namely the empty function which doesn't map anything anywhere, but if Y is non-empty then there are no functions $g : Y \to \emptyset$ at all, since there is nowhere to map the elements of Y.

Lemma 5.2. *Let X and Y be sets. Then there exists an injection $f : X \to Y$ if and only if either X is empty or there exists a surjection $g : Y \to X$.*

Proof. First suppose that there exists an injection $f : X \to Y$. Then we want to show that either X is empty or there exists a surjection $g : Y \to X$, that is, that if X is non-empty then there exists a surjection $g : Y \to X$. So assume X is non-empty, and choose some $x^* \in X$. Also let $Y' \subseteq Y$ be the image of f. Then the function $h : X \to Y'$ given by $h(x) = f(x)$ for $x \in X$ is a bijection (h is injective because f is, and surjective by definition of image). So h has an inverse function $h^{-1} : Y' \to X$. We then define a function $g : Y \to X$ by

$$g(y) = \begin{cases} h^{-1}(y) & \text{if } y \in Y', \\ x^* & \text{if } y \in Y \setminus Y'. \end{cases}$$

Then g is a surjection from Y to X, as required (to see this, observe that since h^{-1} was a bijection, for every $x \in X$ there exists some $y \in Y'$ with $h^{-1}(y) = x$, and this y has $g(y) = x$).

Next observe that, as noted above, if X is empty then the 'empty function' $o : X \to Y$, which doesn't map anything anywhere), is an injection from X to Y.

Finally, suppose that there exists a surjection $g : Y \to X$. Then for each $x \in X$ there exists $y_x \in Y$ with $g(y_x) = x$. Choose[3] some such y_x for each $x \in X$, and define a function $f : X \to Y$ by $f(x) = y_x$ for $x \in X$. Then f is an injection from X to Y, since if $f(x) = f(x')$ then $y_x = y_{x'}$, so $x = g(y_x) = g(y_{x'}) = x'$, so there exists an injection $f : X \to Y$, as required. □

Corollary 5.3. *Let X and Y be sets. Then $|X| \leq |Y|$ if and only if X is empty or there exists a surjection from Y to X.*

Proof. This follows immediately by combining Lemma 5.2 with the definition of cardinality. □

One particularly important cardinality is the cardinality of $\mathbb{N}$, which is denoted[4] $\aleph_0$. This is the smallest cardinality that an infinite set can have, as asserted by the following lemma.

Lemma 5.4. *Let X be a set. Then either X is finite or $|\mathbb{N}| \leq |X|$.*

[3] To choose these elements we are implicitly using the so-called Axiom of Choice. The intricacies of this are beyond the scope of this course, but it is strongly recommended as fascinating further reading for interested students.

[4] The symbol here is 'aleph', the first letter of the Hebrew alphabet, with a zero as a subscript. $\aleph_0$ is said 'aleph-nought' or 'aleph-null'. However, it is rare that you need to write this symbol; typically one would just say that a set is countably infinite.

Proof. Consider the following iterative process to construct a function f. At step t, if X is empty then stop. Otherwise, choose $x \in X$, remove x from X, set $f(t) = x$, and proceed to step $t+1$. There are two possibilities: either the process will stop at some point, or it will continue forever. If the process stops at step t, then we have formed a function $f : \{1, \dots, t-1\} \to X$ which is a bijection, so X is finite (and in particular has size $t-1$). On the other hand, if the process continues forever then by induction every element of $\mathbb{N}$ is mapped to an element of X, and so the process forms an injection $f : \mathbb{N} \to X$, which proves that $|\mathbb{N}| \leq |X|$. □

We say that a set is countably infinite if there exists a bijection $f : \mathbb{N} \to X$. In other words, a set X is countably infinite if and only X has the same cardinality as $\mathbb{N}$. A set is countable if it is either finite or countably infinite; otherwise, it is uncountable. It may help to think of countably infinite sets as being 'small' infinite sets, and uncountable sets as being 'large' infinite sets. Indeed, countable sets are those which we can 'count' using the natural numbers. This is because a set X is countably infinite if and only if there is a bijection $f : \mathbb{N} \to X$, that is, if we can give *every* member of X a *unique* 'label' from $\mathbb{N}$.[5] Our earlier examples prove that $\mathbb{Z}$ and $2\mathbb{Z}$ are both countably infinite. Perhaps surprisingly, the same is in fact true of $\mathbb{N} \times \mathbb{N}$ and $\mathbb{Q}$.

Lemma 5.5. *$\mathbb{N} \times \mathbb{N}$ is countably infinite.*

Proof. The function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(a,b) = \binom{a+b}{2} - b + 1$ is a bijection, so $\mathbb{N} \times \mathbb{N}$ is countably infinite by the definition of countably infinite (the proof that f is a bijection is left as an exercise). □

Proof. The function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ given by $f(n) = (n, 1)$ is an injection, so $|\mathbb{N}| \leq |\mathbb{N} \times \mathbb{N}|$. On the other hand, by uniqueness of prime factorisation the function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $g((a,b)) = 2^a 3^b$ is an injection, so $|\mathbb{N} \times \mathbb{N}| \leq |\mathbb{N}|$. We conclude that $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$, that is, that $\mathbb{N}$ is countably infinite. □

Proof. Observe that we can write out the elements of $\mathbb{N} \times \mathbb{N}$ on an infinite grid as illustrated on the left hand side below.

$$\begin{array}{|cccccc}
\hline
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & \cdots \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & \cdots \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & \cdots \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & \cdots \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
\qquad
\begin{array}{|cccccc}
\hline
1 & 3 & 6 & 10 & 15 & \cdots \\
2 & 5 & 9 & 14 & 20 & \cdots \\
4 & 8 & 13 & 19 & 26 & \cdots \\
7 & 12 & 18 & 25 & 33 & \cdots \\
11 & 17 & 24 & 32 & 41 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

The right hand side shows how the elements of this grid may be labelled with elements of $\mathbb{N}$, by proceeding up each diagonal in turn from bottom left to top right. This creates a bijection from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$, namely the function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ in which each $n \in \mathbb{N}$ in the right hand grid maps to the pair (a,b) in the corresponding position in the left hand grid (so, for instance, $f(10) = (1,4)$ and $f(8) = (3,2)$). Indeed, f is well-defined since every natural number appears once in the right hand grid and so is mapped to exactly one pair in the left-hand grid; f is injective and surjective since each pair appears precisely once in the left-hand grid so is mapped to by exactly one element of $\mathbb{N}$. The existence of this bijection proves that $\mathbb{N} \times \mathbb{N}$ is countably infinite. □

Lemma 5.6. *$\mathbb{Q}$ is countably infinite.*

The proof of this lemma will appear as an assessed exercise. By this point you may be wondering if all infinite sets in fact have the same cardinality, but this is not true; there do exist uncountable sets, one of which is $\mathbb{R}$.

[5] Another way of putting this is that X is countably infinite if and only if every member of X appears exactly once in the second row of an infinite table of the following form (since, given a bijection $f : \mathbb{N} \to X$, taking $x_n = f(n)$ gives such a table).

$$\begin{array}{c|cccccccc}
\mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \dots \\
\hline
X & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & \dots
\end{array}$$

Lemma 5.7. *$\mathbb{R}$ is uncountable.*

Proof. Since $\mathbb{R}$ is non-empty, by Corollary 5.3 it suffices to show that there is no surjection $f : \mathbb{N} \to \mathbb{R}$. To do this, consider any function $f : \mathbb{N} \to \mathbb{R}$. Then we can draw up a table giving the decimal expansion of $f(n) \in \mathbb{R}$ for each $n \in \mathbb{N}$, as in the left hand diagram below (for clarity, the right hand diagram shows the table for the specific function $f(n) = 13 + \sqrt{n}$. So, for instance, in this case we have $a^1 = 14, b^2_1 = 4$ and $b^3_3 = 2$.)

n	$f(n)$
1	$a^1.\mathbf{b^1_1}b^1_2b^1_3b^1_4b^1_5\ldots$
2	$a^2.b^2_1\mathbf{b^2_2}b^2_3b^2_4b^2_5\ldots$
3	$a^3.b^3_1b^3_2\mathbf{b^3_3}b^3_4b^3_5\ldots$
4	$a^4.b^4_1b^4_2b^4_3\mathbf{b^4_4}b^4_5\ldots$
5	$a^5.b^5_1b^5_2b^5_3b^5_4\mathbf{b^5_5}\ldots$
$\vdots$	$\vdots$

n	$13 + \sqrt{n}$
1	$14.\mathbf{0}0000\ldots$
2	$14.4\mathbf{1}421\ldots$
3	$14.73\mathbf{2}05\ldots$
4	$15.000\mathbf{0}0\ldots$
5	$15.2360\mathbf{6}\ldots$
$\vdots$	$\vdots$

Note that the digits b^n_n form a diagonal in the grid (these are the digits in bold font). For each natural number n, define

$$c_n := \begin{cases} b^n_n + 1 & \text{if } b^n_n \leq 5 \\ b^n_n - 1 & \text{if } b^n_n > 5. \end{cases}$$

So for any n, c_n is between 1 and 8 and is not equal to b^n_n. Let x be the real number whose decimal expansion is

$$x = 0.c_1c_2c_3c_4c_5\ldots$$

(So, for instance, in the right-hand example above we would have $x = 0.12315\ldots$). Then, crucially, there is no $m \in \mathbb{N}$ with $f(m) = x$. Indeed,

- $f(1)$ differs from x in the first decimal place (as $c_1 \neq b^1_1$),
- $f(2)$ differs from x in the second decimal place (as $c_2 \neq b^2_2$),
- $f(3)$ differs from x in the third decimal place (as $c_3 \neq b^3_3$),
- and so forth, that is, $f(m)$ differs from x in the mth decimal place (as $c_m \neq b^m_m$).[6]

This proves that f is not surjective. But since $f : \mathbb{N} \to \mathbb{R}$ was arbitrary, this proves that there is no surjection from $\mathbb{N}$ to $\mathbb{R}$, as required.[7] □

Lemma 5.7 shows that the cardinality of $\mathbb{R}$ is different from $\mathbb{N}$; we denote the cardinality of $\mathbb{R}$ by $2^{\aleph_0}$. By Lemma 5.1 this cardinality is larger than $\aleph_0$ since $\mathbb{N} \subseteq \mathbb{R}$. Collectively, the set sizes we have seen so far are every non-negative integer, $\aleph_0$ and $2^{\aleph_0}$, but we can find still bigger cardinalities; indeed, the next lemma implies that there is no largest cardinality, since for every set X the power set of X has larger cardinality than X.

Lemma 5.8. *For every set X, the cardinality of X is strictly less than the cardinality of $\mathscr{P}(X)$.*

Proof. The function $f : X \to \mathscr{P}(X)$ given by $f(x) = \{x\}$ is an injection (to check this, observe that if for some $x, y \in X$ we have $f(x) = f(y)$, then $\{x\} = \{y\}$, so $x = y$). It follows that the cardinality of X is at most that of $\mathscr{P}(X)$, so it remains to show that the cardinalities of X and Y are not equal, or in other words that there is no bijection from X to $\mathscr{P}(X)$.

For this, suppose for a contradiction that $f : X \to \mathscr{P}(X)$ is a bijection. Define a subset $Z \subseteq X$ by

$$Z = \{x \in X : x \notin f(x)\}.$$

[6] Actually, the fact that $x, y \in \mathbb{R}$ differ in some decimal place is not enough to prove that $x \neq y$ (since, for example, 0.39999. . . = 0.40000. . .). However, together with the fact that x cannot have 0 or 9 in its decimal expansion it does follow that $x \neq y$ (alternatively, $x \neq y$ follows from the fact that x and y differ by at least two in each decimal place).

[7] This argument, given by Georg Cantor in 1891, is known as Cantor's diagonalisation argument. Similar arguments, using a diagonal in an appropriate grid to construct an element which is not mapped to, are also known as diagonalisation arguments.

That is, Z consists of all elements of X which are not contained in their image. Let $z \in X$ be such that $f(z) = Z$ (this exists since we assumed that f is a bijection). If $z \in Z$ then $z \in f(z)$, so by definition of Z we have $z \notin Z$. On the other hand, if $z \notin Z$ then $z \notin f(z)$, so by definition of Z we have $z \in Z$. So $z \in Z$ if and only if $z \notin Z$, a contradiction. We conclude that no such bijection can exist. □

Another natural question to ask is whether there is a set whose cardinality is between those of $\mathbb{R}$ and $\mathbb{N}$.

Question 5.9 (Continuum Hypothesis)**.** *Is there a set whose cardinality is larger than $\aleph_0$ and smaller than $2^{\aleph_0}$? That is, is there a set which is 'bigger' than $\mathbb{N}$ and 'smaller' than $\mathbb{R}$? (The 'hypothesis' is that the answer is no).*

Gödel and Cohen proved that this question *cannot be answered* within ZF, the most commonly-used foundation of set theory. That is, they proved that you cannot prove that the answer is yes, and you cannot prove that the answer is no.

Russell's Paradox: There is an inherent problem in allowing sets to be defined simply by giving a rule, for example as in the following definitions:

$$\mathbb{N} = \{n : n \text{ is a natural number}\}, \text{ and}$$

$$\mathcal{U} = \{A : A \text{ is a set}\}.$$

Note that $\mathbb{N} \notin \mathbb{N}$, since $\mathbb{N}$ is a set, not a number. On the other hand, $\mathcal{U} \in \mathcal{U}$, since $\mathcal{U}$ is a set. In the same way, we might consider

$$\mathcal{W} = \{A : A \text{ is a set and } A \notin A\}.$$

So, for example, $\mathbb{N} \in \mathcal{W}$ and $\mathcal{U} \notin \mathcal{W}$. However, a problem arises if we consider whether $\mathcal{W} \in \mathcal{W}$. Indeed, if $\mathcal{W} \in \mathcal{W}$, then by definition of $\mathcal{W}$ we have $\mathcal{W} \notin \mathcal{W}$. On the other hand, if $\mathcal{W} \notin \mathcal{W}$, then since $\mathcal{W}$ is a set, then by definition of $\mathcal{W}$ we must have $\mathcal{W} \in \mathcal{W}$. So

$$\mathcal{W} \in \mathcal{W} \text{ if and only if } \mathcal{W} \notin \mathcal{W}. \qquad \text{(!!!)}$$

This contradiction is known as the Barber paradox[8] or Russell's paradox.[9] It illustrated that to avoid inherent contradictions in mathematics we need to be more careful about what constitutes a set, which led to the development of axiomatic set theory (which we saw earlier in the course). In particular, within the standard axioms of set theory you cannot define a set simply by giving a rule as in the examples above, unless you do this by restricting from a larger set, in which case the Axiom of Specification essentially states that if X is a set then, for any property,

$$\{x \in X : x \text{ satisfies this property}\}$$

is also a set.

[8] A barber has a sign in his window, stating 'I shave all those [and only those] who do not shave themselves'. Does the barber shave himself? Closely related, the Liar paradox is to consider the statement 'this statement is false'. Is the statement true or false? How about: 'Both this sentence and the next sentence are false. The Riemann Hypothesis is true'? Is this true or false, and what does this say about the Riemann Hypothesis?

[9] After Bertrand Russell, who discovered the paradox in 1901.

6. Introduction to Graph Theory

The formal definition of a graph is the following.

Definition. A graph $G = (V, E)$ consists of a finite[1] set of vertices[2] V and a set E of edges, where each edge is an unordered pair $\{u, v\}$ of distinct vertices $u, v \in V$.[3]

However, it may help you to think in terms of the following, more informal definition[4]: a graph consists of a set of points called vertices, some of which may be linked by lines called edges, subject to the following rules:

(i) every edge links two distinct vertices, and
(ii) any pair of distinct vertices is linked by at most one edge.

An example is shown in Figure 1.

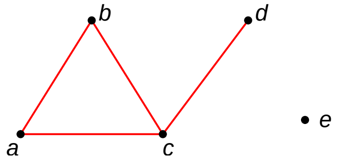

Figure 1: A graph with vertex set $V = \{a, b, c, d, e\}$ and edge set $E = \{\{a, b\}, \{b, c\}, \{a, c\}, \{c, d\}\}$.

Graphs are useful whenever we want to model *connections* between objects, either physical (e.g. cables in a computer network) or abstract (e.g. whether two people know one another).[5]

It follows from the definition of a graph that, formally, two graphs are equal if and only if they have the same vertices and the same edges. However, most of the time we only care about the connections, not the actual names of the vertices. For this reason we make the following definition: two graphs G and H are isomorphic if there exists a bijection $\psi : V(G) \to V(H)$ such that for every $x, y \in V(G)$ we have $\{\psi(x), \psi(y)\} \in E(H)$ if and only if $\{x, y\} \in E(G)$. In other words, G and H are isomorphic if we can transform G to H simply by relabelling vertices (in terms of the bijection ψ, the label x is replaced by $\psi(x)$). Usually in graph theory we consider isomorphic graphs to be the same, and ignore vertex labels; this is what is meant by the term ‘up to isomorphism’. To emphasise whether we consider labels to be important or not we sometimes refer to labelled graphs or unlabelled graphs.

Example. How many labelled graphs are there with vertex set $\{1, 2, 3, 4\}$? Up to isomorphism, how many graphs are there on four vertices? (In other words, how many unlabelled graphs are there with four vertices?)

[1] One can also consider infinite graphs, but we won’t do so in this course.

[2] Note that ‘vertices’ is the plural of ‘vertex’ (similar to ‘matrix’ and ‘matrices’). ‘Vertice’ and ‘vertexes’ are both incorrect!

[3] This definition of a graph is sometimes called a *simple graph*. Writers using this term would typically say that a graph may contain loops (an edge from a vertex to itself) and multiple edges between a pair of vertices. However, please note carefully that we forbid these (we would refer to a graph with these allowed as a multigraph, but these will not play a role in this course).

[4] Make sure you understand why these two definitions are equivalent.

[5] A variety of other types of graphs are also useful: in a directed graph each edge has a direction, from one vertex to another vertex. In a weighted graph each edge has a corresponding weight, which might indicate, for example, the capacity of the connection, and in a hypergraph we permit edges consisting of more than two vertices. However, we do not have time to study any of these objects in this course.

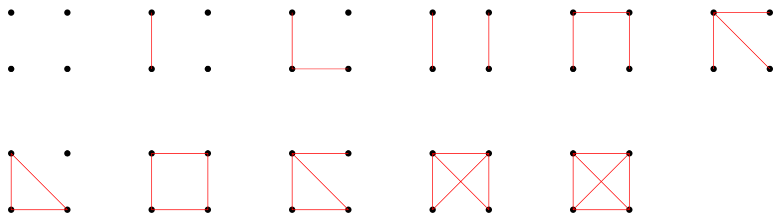

Figure 2: The 11 distinct graphs on four vertices up to isomorphism.

Solution. There are six possible edges in a graph with vertex set $\{1,2,3,4\}$, namely $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$ and $\{3,4\}$. We can form a labelled graph with vertex set $\{1,2,3,4\}$ by choosing for each of these pairs whether that pair is an edge or not an edge. Since we are making 6 choices, each with two options, there are $2^6 = 64$ possible graphs which could result.

Up to isomorphism, there are 11 graphs on four vertices, as shown in Figure 2. □

None of the graph properties we consider in this course depend on the vertex labelling, so we will almost always work with graphs up to isomorphism and not care about the specific labels of vertices. Due to this we adopt the following standard convention: *unless stated otherwise we consider two graphs to be the same if they are isomorphic.*

We now give definitions of many key terms relating to graphs.

Definition. The order of a graph G is the number of vertices of G, and the size of G is the number of edges of G. We often write uv as shorthand for an edge $\{u,v\}$, but you should remember that it is necessary that u,v are different and the edge uv is the same as the edge vu because these are unordered pairs. We say that an edge $e = uv$ is incident to the vertices u and v. If uv is an edge, we say that the vertices u and v are adjacent, that u is a neighbour of v and that v is a neighbour of u. For any vertex v of a graph G, the neighbourhood $N(v)$ of v is the set of neighbours of v, and we say that v is isolated if it has no neighbours. The degree of a vertex v in a graph G is $\deg(v) = |N(v)|$, that is, the number of neighbours of v (note that this is equal to the number of edges incident to v).

For example, in the graph illustrated in Figure 1 the neighbourhood of a is $N(a) = \{b,c\}$, the neighbourhood of d is $N(d) = \{c\}$, and the neighbourhood of e is $N(e) = \emptyset$, that is, vertex e is an isolated vertex. The vertex degrees are $\deg(a) = 2$, $\deg(b) = 2$, $\deg(c) = 3$, $\deg(d) = 1$ and $\deg(e) = 0$.

Sometimes we will want to talk about more than one graph at the same time, in which case we modify the above notation slightly to avoid ambiguity, writing $V(G)$ for the vertex set of a graph G, $E(G)$ for the edge set of G, and $N_G(v)$ and $\deg_G(v)$ for the neighbourhood and degree of v in G respectively.

Definition. Given a graph G, the complement of G is the graph $\overline{G}$ with vertex set $V(\overline{G}) = V(G)$ and whose edge set is $E(\overline{G}) = \{\{u,v\} : u,v \in V(G), u \neq v, \text{ and } \{u,v\} \notin E(G)\}$.

In other words, G and $\overline{G}$ have the same vertex set, and for any pair of distinct vertices u and v, there is an edge between u and v in $\overline{G}$ if and only if there is not an edge between u and v in G.

Observe that if G is a graph with n vertices, then the degree of each vertex of G is an integer between 0 and $n-1$ (since there are $n-1$ vertices other than v). The next lemma relates the sum of all vertex degrees to the number of edges.

Lemma 6.1 (The handshaking lemma). *Every graph $G = (V,E)$ satisfies $\sum_{v\in V} \deg(v) = 2|E|$. That is, the sum of all vertex degrees is twice the number of edges.*

Proof. Imagine that for each vertex we put a pebble on each edge to which it is incident. Then the number of pebbles placed for a vertex v is $\deg(v)$, so the total number of pebbles is equal to $\sum_{v\in V} \deg(v)$. However, there will be precisely two pebbles on each edge (one placed from each endvertex). So the total number of pebbles is also equal to $2|E|$, so we must have $2|E| = \sum_{v\in V} \deg(v)$.[6] □

[6] This kind of argument, where the same quantity (here the total number of pebbles) is counted in two different ways, is called a *double-counting argument*.

Corollary 6.2. *Every graph contains an even number of vertices with odd degree.*

Proof. Suppose for a contradiction that the number of vertices with odd degree is odd. Then $\sum_{v \in V} \deg(v)$ is odd. But $|E|$ must be an integer, so $2|E|$ is even, giving a contradiction to the handshaking lemma. We therefore conclude that there cannot be an odd number of vertices with odd degree. □

Often it is convenient to represent the degrees of vertices in a graph by a sequence, in accordance with the following definition.

Definition. The degree sequence of a graph G is the sequence of all degrees of vertices in G,

$$(\deg(v_1), \deg(v_2), \deg(v_3), \dots, \deg(v_n))$$

where $V = \{v_1, v_2, v_3, \dots, v_n\}$ and these degrees are given in *ascending* order, that is, so that $\deg(v_1) \leq \deg(v_2) \leq \deg(v_3) \leq \cdots \leq \deg(v_n)$.

Note, however, that the degree sequence of a graph G does *not* uniquely identify G up to isomorphism. In other words, there exist non-isomorphic graphs with the same degree sequence.[7] Also, not all ascending sequences of integers are degree sequences of graphs. For example, $(3, 3, 3, 3, 3)$ cannot be the degree sequence of a graph by Corollary 6.2, since the sum of the degrees in the sequence is odd, and many other sequences can also be seen to be impossible by similar arguments.

We conclude this section by defining a few more important terms relating to degrees in graphs.

Definition. The minimum degree of a graph G, denoted $\delta(G)$, is the smallest degree of a vertex of G. Similarly, the maximum degree of a graph G, denoted $\Delta(G)$, is the largest degree of a vertex of G. A graph G is regular if every vertex of G has the same degree, that is, if $\deg(u) = \deg(v)$ for all $u, v \in V$. Finally, we say that G is k-regular to mean that every vertex has degree k.

Subgraphs, Paths, Cycles and Cliques

The following definition gives some commonly-encountered graphs, which are drawn in Figure 3. Note in particular that $C_3 = K_3$; this graph is also known as the triangle.

Definition. The complete graph, or clique, on n vertices, denoted K_n, has vertex set $V(K_n) = \{1, 2, \dots, n\}$ and edge set $E(K_n) = \{\{i, j\} : 1 \leq i < j \leq n\}$. That is, K_n has n vertices and an edge between every pair of vertices, giving $\binom{n}{2}$ edges in total.

The path of length n, denoted P_n, has vertex set $V(P_n) = \{1, 2, \dots, n+1\}$ and edge set $E(P_n) = \{\{i, i+1\} : 1 \leq i \leq n\}$. So P_n has $n+1$ vertices and n edges which are arranged one after the next in a 'linear' fashion.

For $n \geq 3$, the cycle of length n, denoted C_n, has vertex set $V(C_n) = \{1, 2, \dots, n\}$ and edge set $E(C_n) = \{\{1, 2\}, \{2, 3\}, \dots, \{n-1, n\}, \{n, 1\}\}$. So C_n has n vertices and n edges, which are arranged one after the next in a 'cyclic' fashion.

One question of particular interest to us is when copies of these (or other) graphs can be found in other, larger graphs, and if so, how many such copies there are. For example, we might ask whether or not a graph contains a path from one vertex to another. The following definition formalises what we mean by 'contains' in this context: we say that a graph G contains another graph H if G is a *subgraph* of H.

Definition. A graph H is a subgraph of a graph G if we can obtain H by deleting vertices and edges of G. Equivalently H is a subgraph of G if H is a graph with $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$.[8] H is a spanning subgraph of G if additionally $V(H) = V(G)$, that is, if only edges were deleted. A copy of H in G is a specific instance of H as a subgraph of G, that is, specific subsets $V' \subseteq V(G)$ and $E' \subseteq E(G)$ such that $H = (V', E')$.

[7] Exercise: give an example.

[8] Remember, however, that unless stated otherwise we consider isomorphic graphs to be the same graph.

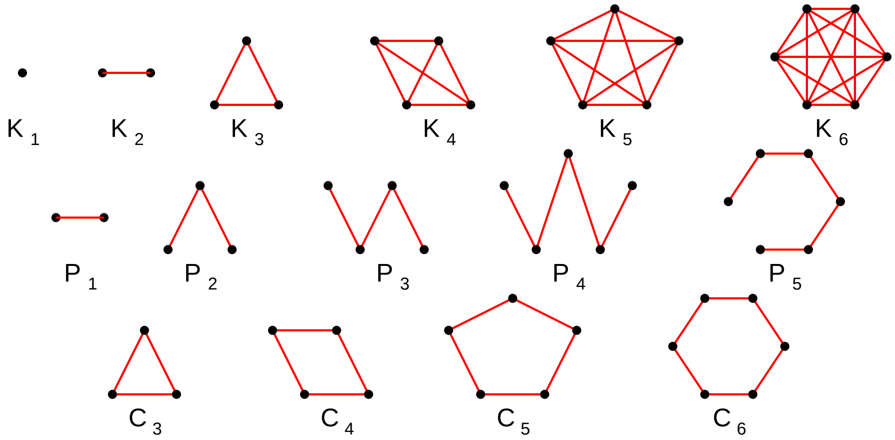

Figure 3: Some examples of the graphs defined above.

Example. How many copies of P_2 are there in K_n?

Solution. Note that on any three distinct vertices of K_n there are three different copies of P_2, as shown in Figure 4. Indeed, these copies each have the same vertex set $\{a, b, c\}$, but they are different since they have different edge sets, namely $\{ab, ac\}, \{ac, bc\}$ and $\{ab, bc\}$. So one way to calculate the number of copies of P_2 in K_n is to take the number of ways to choose three distinct vertices of K_n, namely $\binom{n}{3}$, and to multiply by 3, since each choice of three vertices supports three copies of P_2. So in total there are $3\binom{n}{3}$ copies of P_2 in K_n.

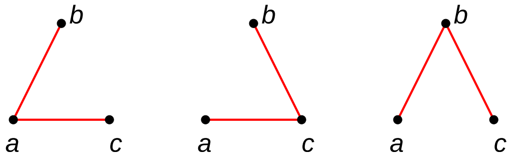

Figure 4: The three different copies of P_2 on vertices a, b, c of K_n.

Another approach is to note that any ordered triple (x, y, z) of distinct vertices of K_n gives a copy of P_2, whose vertices are x, y and z and whose edges are xy and yz. Recall that the number of ways to choose three vertices out of n, with order but without repetition, is $\frac{n!}{(n-3)!}$. However, this counts each copy of P_2 *twice*, since the triples (x, y, z) and (z, y, x) give rise to the same copy of P_2. So in total there are $\frac{1}{2}\frac{n!}{(n-3)!}$ copies of P_2 in K_3 (note that this expression is equal to the one obtained above).[9] □

Similar arguments can be used to count the number of copies of other small graphs in standard larger graphs.

[9] Compare these two approaches to the approaches used for mixed choice in Section 8; the ideas are very similar.

7. Extremal Graph Theory, Walks and Connectedness

Extremal graph theory is a large and active area of current research in graph theory, and this section is an introduction to the kind of results and methods which are common in this area, including in particular the proof of a result we will need when we discuss trees. In the previous section we saw results describing copies of a given graph within a fixed larger subgraph (e.g. copies of P_2 in K_n). However, often we want to consider not just one single larger graph, but instead a wide class of larger graphs. That is, we want to find conditions which ensure that *any* graph which satisfies these conditions must contain the subgraph which we are looking for. For example, the next lemma shows that *any* graph with minimum degree at least two contains a cycle (i.e. contains a copy of C_k for some k).

Lemma 7.1. *Let G be a graph with $\delta(G) \geq 2$. Then G contains a cycle.*

Proof. Consider a longest path P in G, and let $x_1, x_2, \ldots, x_{r-1}, x_r$ be the vertices of P, ordered as in P. Then x_{r-1} is a neighbour of x_r, and since $\deg(x_r) \geq 2$ there must be another neighbour v of x_r in G. Note that v must lie in P, since otherwise the vertices $x_1, x_2, \ldots, x_{r-1}, x_r, v$ (in that order) would form a longer path in G. So $v = x_i$ for some $i \leq r-2$, but then the vertices $x_i, x_{i+1}, \ldots, x_{r-1}, x_r, x_i$ (in that order) form a cycle in G. □

Observe that this lemma is best-possible in the sense that if we replaced the condition $\delta(G) \geq 2$ by the condition $\delta(G) \geq 1$ the result would not be true. For example, for each n the path P_n would provide a counterexample. Also note the idea of studying a longest path in G; considering extremal structures (e.g. longest path, shortest cycle, largest clique, etc.) within a graph so that you can then contradict their extremality (as in this proof) is a common approach in this area which we will see again.

The next lemma improves on Lemma 7.1 by replacing the minimum degree condition with a weaker condition on the number of edges. Specifically it shows that any graph with at least as many edges as vertices contains a cycle.

Theorem 7.2. *For every $n \in \mathbb{N}$, every graph of order n and size at least n contains a cycle.*

Proof. Suppose for a contradiction that there exist natural numbers n for which there exist graphs of order n and size at least n which do not contain a cycle. Let G be such a graph, and moreover choose G such that the order n of G is the smallest[1] for which such graphs exist. Note that we must have $n \geq 3$, since any graph with at most two vertices has fewer edges than vertices. Moreover we cannot have $\delta(G) \geq 2$, since then G would contain a cycle by Lemma 7.1. It follows that there must be some vertex $x \in V(G)$ with $\deg(x) = 0$ or $\deg(x) = 1$. Let G' be the graph formed by deleting x from G, along with the edge incident to x in the case where $\deg(x) = 1$. Then G' has $n-1$ vertices (since we deleted one vertex) and at least $n-1$ edges (since we deleted at most one edge). Furthermore G' does not contain a cycle, since this would also be a cycle in G. The existence of G' therefore contradicts the minimality of G, and we conclude from this contradiction that the theorem is true. □

[1] To know that there is a smallest order for which counterexamples exist, we are appealing to the *well-orderedness of the natural numbers*, which is the fact that every non-empty set of natural numbers contains a smallest element. This can be proved by induction; moreover the principle of mathematical induction can equally be deduced from the well-orderedness of the natural numbers, showing that the two statements are equivalent (and it is a defining property of the natural numbers that these hold). Through this equivalence any proof by induction can be translated to a proof by minimal counterexample, and vice versa. You might like to try this here by giving a proof of Theorem 7.2 by induction on n, in which for the inductive step you use the idea used in this proof of deleting a vertex of degree zero or one.

Again paths show that this theorem is best-possible, in the sense that the path P_{n-1} is a counterexample if we replace "size at least n" by "size at least $n-1$". This proof illustrates another important proof idea in this area, namely that you consider a minimal/maximal counterexample to the result you want to prove (in this case, a counterexample of smallest order), and then edit this hypothetical counterexample so as to contradict the fact that it was minimal/maximal, whereupon this contradiction proves that no counterexample exists.

We can also consider specific cycles (or other graphs), and we will illustrate this by giving analogous results for the triangle K_3. For this we use the following simple lemma.

Lemma 7.3. *Let xy be an edge in a graph G of order n. If $\deg(x)+\deg(y)>n$, then G contains a copy of K_3.*

Proof. The neighbourhoods $N(x)$ and $N(y)$ have sizes $\deg(x)$ and $\deg(y)$ respectively. Moreover their union has size $|N(x)\cup N(y)|\leq n$ since there are ony n vertices of G in total. By inclusion-exclusion it follows that

$$|N(x)\cap N(y)|=|N(x)|+|N(y)|-|N(x)\cup N(y)|>n-n=0,$$

so we may choose a vertex $z\in N(x)\cap N(y)$, so that z is a neighbour of both x and y. The vertices x,y and z then form a copy of K_3 in G. □

The following corollary follows immediately, since the minimum degree condition implies that every edge satisfies the condition of Lemma 7.3.

Corollary 7.4. *Let G be a graph of order $n\geq 3$ with $\delta(G)>n/2$. Then G contains a copy of K_3.*

The degree condition of Corollary 7.4 is best-possible in the sense that it is not true if the condition $\delta(G)>n/2$ is relaxed to the condition $\delta(G)\geq n/2$. Indeed, for each even n a counterexample is given by the graph with vertex set $A\cup B$ and edge set $\{ab : a\in A, b\in B\}$ for disjoint sets A and B with $|A|=|B|=n/2$ (we will see this graph in a subsequent section as the complete bipartite graph $K_{\frac{n}{2},\frac{n}{2}}$). Again we can also give a version based on the size of G; this celebrated theorem was proved by Mantel in 1907, and is widely-regarded as the founding result of extremal graph theory.

Theorem 7.5 (Mantel's theorem)**.** *Every graph with n vertices and more than $n^2/4$ edges contains a copy of K_3.*

Proof. Suppose that G is a graph on n vertices with no copy of K_3; we will show that G has at most $n^2/4$ edges. Observe first that

$$\sum_{v\in V}\deg(v)^2=\sum_{xy\in E}(\deg(x)+\deg(y)),$$

since for any vertex v of G, the sum on the right hand side counts the degree $\deg(v)$ once for each edge incident to v (in other words, $\deg(v)$ is counted $\deg(v)$ times). Also, since G does not contain K_3, for every edge xy of G we must have $\deg(x)+\deg(y)\leq n$ by Lemma 7.3, and so

$$\sum_{xy\in E}(\deg(x)+\deg(y))\leq |E|n.$$

On the other hand, by the Cauchy-Schwarz inequality[2] we have

$$\frac{1}{n}\left(\sum_{v\in V}\deg(v)\right)^2\leq\sum_{v\in V}\deg(v)^2.$$

Together with the Handshaking lemma, which states that $\sum_{v\in V}\deg(v)=2|E|$, combining the three displayed inequalities gives

$$\frac{(2|E|)^2}{n}\leq |E|n,$$

and so we have $|E|\leq n^2/4$ as required. □

Again the graph $K_{\frac{n}{2},\frac{n}{2}}$ described above demonstrates that this theorem is best-possible, in the sense that the "more than" in the statement cannot be relaxed to "at least".

[2] Or, if you prefer, by convexity of x^2, which implies that $(\frac{1}{n}\sum_{v\in V}\deg(v))^2\leq\frac{1}{n}\sum_{v\in v}\deg(v)^2$ (in words, the average of the squares of some collection of real numbers is at least the square of the average of the collection).

Walks and Connectedness

Loosely speaking, a *walk* in a graph is a route which can be traced through the graph by moving from vertex to vertex along edges, and a walk is closed if the first and last vertices of the walk are the same, that is, if you finish at the same vertex at which you started.

Definition. A walk W in a graph is an ordered sequence $(v_0, v_1, \dots, v_k)$ of vertices such that $v_i v_{i+1}$ is an edge for any $0 \leq i \leq k-1$. The length of W is the number of edges traversed, that is, k. A walk W is closed if $v_0 = v_k$. We say that W is a walk from x to y if $v_0 = x$ and $v_1 = y$.

So the vertices and edges traversed by a walk W form a path if and only if W has no repeated vertices, and if W is closed then they form a cycle if and only if the only repeated vertex in W is the first and last vertex. One important use of walks in graphs, which unfortunately we do not have time to investigate, is the study of *random walks*, where at each step you select at random which edge incident to the current vertex will be traversed next. Such walks have a very wide range of applications (to give one example, they can be used to model Brownian motion in Physics). Walks are also the basis of Eulerian tours[3], such as those used for the famous "Bridges of Königsberg" problem.

Our use of walks will mainly be to describe what it means for a graph to be *connected*, a property which, roughly speaking, means that you can 'get from any vertex to any other vertex' in the graph. For a non-connected graph we can also define *connected components*, the connected parts of the graph. The next definition defines these terms formally, but I think it is much easier understood visually, as in Figure 1 and Figure 2.

Definition. A graph G is connected if for all vertices u and v of G there is a walk in G from u to v. A connected component of G is a maximal connected subgraph of G, so two distinct vertices are in the same connected component if and only if there is a walk between them in G.

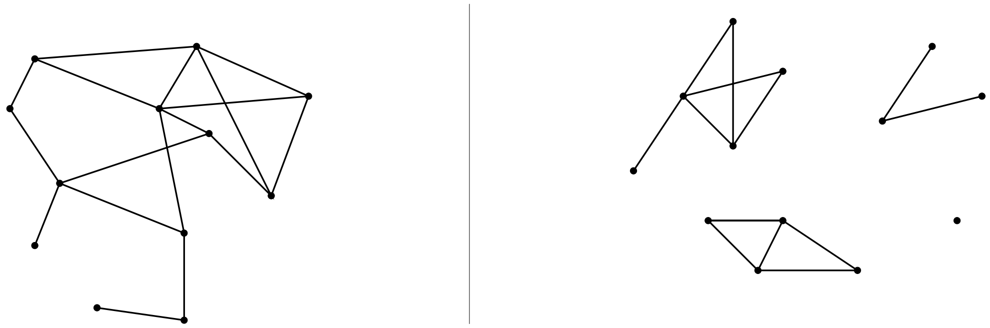

Figure 1: The graph on the left is *connected* – there is a walk from any vertex to any other vertex. However, the graph on the right is not connected. Observe how this graph has four connected 'parts' (circled in Figure 2). These are the *connected components* of the graph.

Any graph consists of one or more connected components (see Figures 1 and 2 for examples), and a graph is connected if and only if it has exactly one connected component.

Proposition 7.6. *Let u and v be vertices of a graph G. Then G contains a path from u to v if and only if G contains a walk from u to v.*

Proof. The vertices of any path from u to v in G, in order as they appear on the path, form a walk from u to v in G. So we only need to show that if G contains a walk from u to v then G contains a path from u to v. Let $W = (v_0, v_1, \dots, v_\ell)$ be a shortest walk from u to v in G (so $u = v_0$ and $v = v_\ell$). If $v_i = v_j$ for some $0 \leq i < j \leq \ell$, then

$$(v_0, v_1, \dots, v_{i-1}, v_i, v_{j+1}, v_{j+2}, \dots, v_{\ell-1}, v_\ell)$$

[3] These are covered in the module 2AC, which should be available to you as an option next year.

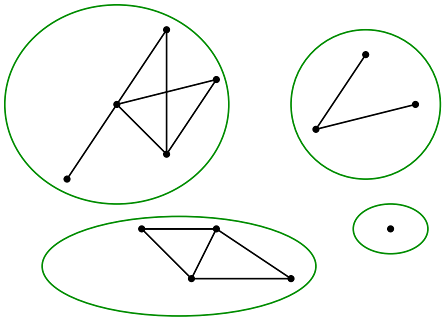

Figure 2: The connected components of the graph on the right hand side of Figure 1 are circled.

is a shorter walk from u to v in G, contradicting our choice of W. So the vertices of W must all be distinct, meaning that these vertices together with the edges traversed by W form a path from u to v in G. □

Proposition 7.6 allows us to give an equivalent definition in terms of paths: a graph G is connected if and only if for all vertices u and v of G there is a path from u to v in G. We use this form of the definition to show that a connected graph on n vertices must have at least $n-1$ edges.

Theorem 7.7. *Every graph of order n with size at most $n-2$ is not connected.*

Proof. We argue by induction. The theorem statement holds vacuously for $n = 1$, and it is straightforward to check the statement for $n = 2$. Now suppose that the theorem statement holds for some $k \in \mathbb{N}$, and let G be an arbitrary graph on $k+1$ vertices with at most $k-1$ edges. By the Handshaking Lemma we then have $\sum_{v \in V} \deg(v) \leq 2(k-1) < 2(k+1)$, so G must contain a vertex v with $\deg(v) < 2$.[4] Suppose for a contradiction that G is connected. Then we cannot have $\deg(v) = 0$, so we must have $\deg(v) = 1$. Let G' be the subgraph of G formed by deleting v and the edge incident to v. Then G' has k vertices and at most $k-2$ edges, so our induction hypothesis implies that G' is not connected. It follows that there exist vertices x and y of G' for which there is no path in G' from x to y. Since G is connected there is a path P in G from x to y. However, v cannot be an endvertex of P since $v \neq x$ and $v \neq y$ (because $x, y \in V(G')$ and $v \notin V(G')$), and so v cannot be a vertex of P since $\deg_G(v) = 1$ but every vertex of a path other than the endvertices has degree at least two. It follows that P is a path in G' from x to y, contradicting the fact that no such path exists by choice of x and y. We conclude that G is not connected, and since G was arbitrary it follows that every graph on $k+1$ vertices with at most $k-1$ edges is not connected, that is, that the theorem statement holds for $k+1$, completing the inductive step. □

[4] Because the average degree of G is less then 2, so there must be a vertex with degree less than 2.

8. Trees and Bipartite Graphs

For brevity, we say that a graph is acyclic if it does not contain a cycle.

Definition. A tree is a connected acyclic graph. A leaf of a tree is a vertex v with $\deg(v) = 1$.

Trees are an important class of graphs which have many applications: phylogenetic trees, search trees, decision trees and so forth. They are called trees due to the way they 'branch out' – see Figures 1 and 2 for examples. The next few results specify a number of important properties of trees, beginning with the observation that Theorems 7.2 and 7.7 together allow us to specify exactly how many edges a tree has.

Corollary 8.1. *Every tree of order n has precisely $n - 1$ edges.*

Proof. Let T be a tree on n vertices. Then by definition T is connected and acyclic. Since T is acyclic, Theorem 7.2 implies that T has at most $n - 1$ edges; since T is connected Theorem 7.7 implies that T has at least $n - 1$ edges. □

Lemma 8.2. *Every tree of order $n \geq 2$ has a leaf.*[1]

Proof. Let T be a tree on $n \geq 2$ vertices. Then T has $n - 1$ edges by Corollary 8.1. By the Handshaking Lemma this implies that $\sum_{v \in V} \deg(v) = 2(n-1)$, so the average degree of vertices in T is $\frac{2(n-1)}{n} < 2$. So T contains a vertex v with $\deg(v) \leq 1$. Since T is connected and has at least two vertices, T has no isolated vertices, so $\deg(v) \neq 0$, and therefore $\deg(v) = 1$. So v is a leaf of T. □

Our next theorem states that every connected graph contains a tree as a spanning subgraph, a result with many important applications.[2]

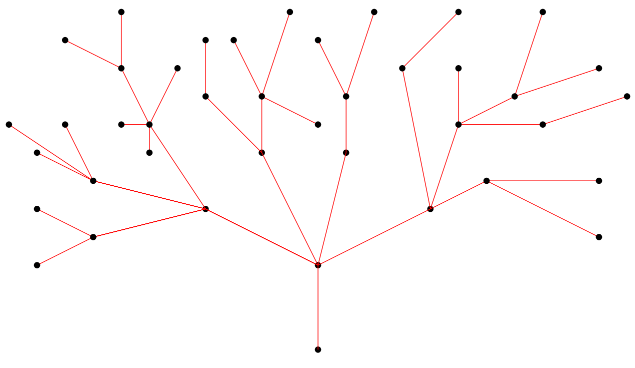

Figure 1: A tree.

[1] In fact, any tree of order $n \geq 2$ has at least two leaves; as an exercise try to adapt the proof to show this.

[2] Spanning trees are important because they are minimal connected subgraphs, that is, the smallest number of edges you need to be able to get from any vertex to any other vertex. For example, if your graph is a road network, then provided you can keep open the roads (edges) of a spanning tree, then traffic will still be able to travel from any city (vertex) to any other city, even if all of the other roads are closed.

A natural related question is whether a graph will still be connected even after the closure of an arbitrary set of at most $k - 1$ cities (vertices) or roads (edges); this leads to the notions of *k-connectedness* and *k-edge-connectedness*, which are explored in detail in the Year 3 Graph Theory course.

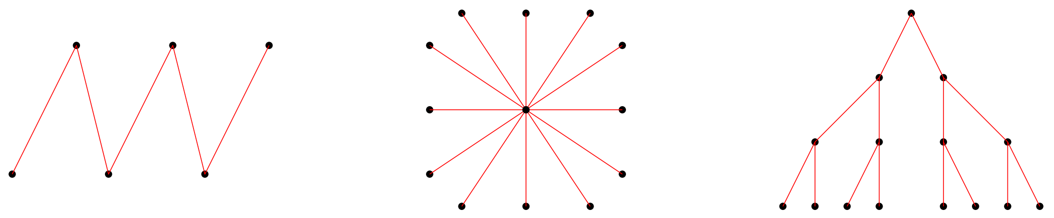

Figure 2: Examples of three specific types of trees, namely a *path*, a *star* and a *binary tree*.

Theorem 8.3. *Every connected graph contains a spanning tree (that is, has a spanning subgraph which is a tree).*

Proof. We begin with the following observation: if H is a connected graph which contains a cycle C, and $e = uv$ is an edge of C, then the graph H' formed by deleting e from H is also connected. To see this, let $x_1, x_2, \dots, x_\ell$ be the vertices of C, in order, with $x_1 = u$ and $x_\ell = v$. Choose any vertices $y, z \in V(H)$. Since H is connected there is a walk W in H from y to z. Form a new walk W' by replacing each instance of (x_1, x_ℓ) in W with $(x_1, x_2, \dots, x_{\ell-1}, x_\ell)$, and replacing each instance of (x_ℓ, x_1) in W with $(x_\ell, x_{\ell-1}, \dots, x_2, x_1)$. In other words, each time W traverses the edge uv, W' instead moves between u and v by going the 'other way' around C. Then by construction W' doesn't traverse the edge e, and so is a walk from y to z in H'. Since y and z were arbitrary it follows that H' is connected, as claimed.

We now prove the theorem. Let G be a connected graph, and let T be a connected spanning subgraph of G of smallest size (that is, with the fewest edges among all connected subgraphs of G). If T contains a cycle then by the observation above we may choose and delete an edge of that cycle to obtain a subgraph $T' \subseteq T$ which is also a connected spanning subgraph of G, and which has fewer edges than T, contradicting the choice of T. We conclude that T must be acyclic; since T is connected it follows that T is a tree, so T is a spanning tree in G, as required. □

A tree T on n vertices is connected and acyclic by definition, and has $n-1$ edges by Corollary 8.1. To conclude this section, the next two results show that any graph which satisfies at least two of these three properties must be a tree (and therefore satisfies the third property).

Corollary 8.4. *Every connected graph of order n with precisely $n-1$ edges is a tree.*

Proof. Let G be a connected graph on n vertices with precisely $n-1$ edges. By Theorem 8.3 G contains a spanning tree T. Then T is a tree on n vertices, so has $n-1$ edges by Corollary 8.1. This means that T is a subgraph of G with the same number of vertices and edges as G, that is, T must be equal to G. So G is a tree. □

Lemma 8.5. *Every acyclic graph of order n with precisely $n-1$ edges is a tree.*

Proof. Let G be an acyclic graph on n vertices with precisely $n-1$ edges. Let $C_1, \dots, C_m$ be the connected components of G, and for each $1 \leq i \leq m$ let n_i be the order of C_i. So $\sum_{i=1}^m n_i = n$. Since G is acyclic, each connected component C_i must be acyclic, so by Theorem 7.2 it follows that C_i has at most $n_i - 1$ edges. So, in total, G has at most $\sum_{i=1}^m (n_i - 1) = \sum_{i=1}^m n_i - m = n - m$ edges. Since G has $n-1$ edges we must have $m = 1$, that is, that G has only one connected component. So G is connected, and so G is a tree. □

To summarize, consider the following possible properties of a graph T.

(i) T is connected.
(ii) T is acyclic.
(iii) T has $n-1$ edges, where n is the order of T.

If T is a tree, then satisfies (i) and (ii) by definition, and also satisfies (iii) by Corollary 8.1. Conversely, Corollary 8.4, Lemma 8.5, and the definition of a tree imply that if T satisfies any two of these properties then T is a tree (and therefore satisfies the third property also).

Bipartite Graphs

Bipartite graphs are an important class of graphs in which every edge crosses a partition of the vertex set into two parts (i.e. is incident to a vertex on either side). We begin with a formal definition.

Definition. A graph G is bipartite if its vertex set V can be written as $V = V_1 \cup V_2$ where V_1 and V_2 are disjoint and every edge of G is incident to one vertex of V_1 and one vertex of V_2 (so no edges have both endvertices in V_1 or both endvertices in V_2). We refer to the sets V_1 and V_2 as vertex classes of G.

Note that there may be more than one possible way to choose vertex classes of a bipartite graph. Bipartite graphs arise naturally in applications where a graph represents connections between different types of objects. For example, a scheduling problem might consider a graph whose vertices are students and classes, where there is an edge between a student and a class if that student is taking that class; in this context an edge between two students, or between two classes, would not make sense, and the vertex classes would be the set of vertices corresponding to students and the set of vertices corresponding to classes respectively.

Definition. The complete m by n bipartite graph, denoted $K_{m,n}$ has vertex set $V(K_{m,n}) = V_1 \cup V_2$, where V_1 and V_2 are disjoint sets of sizes m and n respectively, and edge set $E(K_{m,n}) = \{xy : x \in V_1, y \in V_2\}$. So V_1 and V_2 are the vertex classes of $K_{m,n}$, and there is an edge between vertices x and y if and only if x and y are not in the same vertex class.

See Figure 3 for some examples. In a bipartite graph any walk must move from one vertex class to the other with each step; this is the essence of the proof of the next lemma.

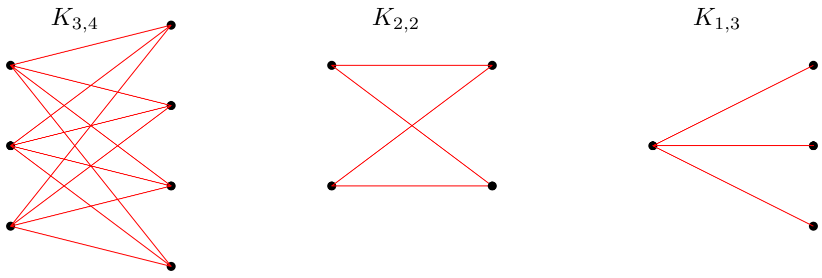

Figure 3: Some complete bipartite graphs.

Proposition 8.6. *Every closed walk in a bipartite graph has even length.*

Proof. Let G be a bipartite graph, and let $W = (v_0, v_1, v_2, \dots, v_\ell)$ be a walk of length ℓ in G. Let V_1 and V_2 be vertex classes of G, chosen so that $v_0 \in V_2$. Since $v_{i-1}v_i$ is an edge of G for each $1 \leq i \leq \ell$, the vertices v_{i-1} and v_i must lie in different vertex classes, and so we must have $v_i \in V_2$ if i is even and $v_i \in V_1$ if i is odd. So v_1 and v_ℓ lie in different vertex classes if and only if ℓ is even. However, if W is closed then $v_0 = v_\ell$, so $v_1 v_\ell$ is an edge of G, which implies that v_1 and v_ℓ lie in different vertex classes, and therefore that ℓ is even. So every closed walk in G has even length. □

The next lemma shows that in any graph, a shortest closed walk of odd length (if such a walk exists) is an odd cycle.

Proposition 8.7. *If a graph contains a closed walk of odd length, then it contains a cycle of odd length.*

Proof. Let $W = (v_0, v_1, \dots, v_\ell)$ be a closed walk of odd length (so $v_0 = v_\ell$); moreover, choose W to be shortest among all closed walks of odd length. Suppose that W has a repeated vertex other than $v_0 = v_\ell$, say $v_i = v_j$ for some $0 \leq i < j < \ell$. Then

$$W^1 = (v_0, v_1, \dots, v_{i-1}, v_i, v_{j+1}, v_{j+2}, \dots, v_{\ell-1}, v_\ell)$$

is a closed walk of length $i + \ell - j$, and

$$W^2 = (v_i, v_{i+1}, \dots, v_{j-1}, v_j)$$

is a closed walk of length $j - i$. Since $(i + \ell - j) + (j - i) = \ell$ is odd, either W^1 or W^2 must have odd length, contradicting the fact that W was shortest among all closed walks of odd length. We conclude that W has no repeated vertices other than $v_0 = v_\ell$, and so W is a cycle of odd length. □

Note that the analogous statement for even length walks and cycles is not true. That is, a graph with no cycles of even length may contain a closed walk of even length[3].

We conclude this section with the central theorem regarding bipartite graphs, that we can give a characterisation of such graphs in terms of the existence of cycles. Specifically, a graph is bipartite if and only if it contains no odd cycles. This exemplifies the kind of beautiful connections which often arise in graph theory; looking at the definition of a bipartite graph it seems quite unexpected that this kind of characterisation should exist.

Theorem 8.8. *Let G be a graph. Then G is bipartite if and only G contains no cycles of odd length.*

Proof. First suppose that G is bipartite. Then any cycle in G forms a closed walk in G, and so has even length by Proposition 8.6. So G has no cycles of odd length.

It remains to prove that if G has no cycles of odd length then G is bipartite. We first prove this statement for connected graphs. So let G be a connected graph with no cycles of odd length. Choose any vertex $x \in V(G)$, and define

$$V_1 = \{y \in V(G) : \text{there is a walk of odd length from } x \text{ to } y\}, \text{and}$$
$$V_2 = \{y \in V(G) : \text{there is a walk of even length from } x \text{ to } y\}.$$

Note that since G is connected every vertex is then either in V_1 or V_2. Now suppose that V_1 and V_2 have a common vertex, say y. Then G contains a walk W_1 from x to y of odd length (since $y \in V_1$) and a walk W_2 from x to y of even length (since $y \in V_2$). Traversing first W_1, and then the reverse of W_2, gives a closed walk in G of odd length, from which Proposition 8.7 implies that G contains a cycle of odd length, contradicting our assumption. Similarly, if there is an edge yy' with $y, y' \in V_1$ then combining a walk of odd length from x to y, a walk of odd length from x to y' and the edge yy' gives a closed walk of odd length in G, whilst if there is an edge yy' with $y, y' \in V_2$ we obtain a closed walk of odd length in G by a similar argument. As before, such a closed walk contradicts our assumption on G. We conclude that the sets V_1 and V_2 are disjoint, that $V_1 \cup V_2 = V(G)$, and that every edge of G has one endvertex in V_1 and one endvertex in V_2, that is, that G is bipartite with vertex classes V_1 and V_2.

We have now proved that every connected graph with no cycles of odd length is bipartite. Finally, let G be an arbitrary graph with no cycles of odd length; then each connected component of G is a connected graph with no cycles of odd length. So each connected component of G is bipartite. Let $C_1, \dots, C_m$ be the connected components of G, and let V_1^i and V_2^i be vertex classes of C_i; then G is bipartite with vertex classes $V_1 = \bigcup_{i=1}^m V_1^i$ and $V_2 = \bigcup_{i=1}^m V_2^i$. □

[3] Exercise: give an example.

9. Relations

Definition. A binary relation on a set A is a subset $\sim \subseteq A \times A$. We often just say relation instead of 'binary relation', but be aware that there are other types of relation.

However, we usually think of $\sim$ as being a collection of statements: we say that "a is related to b by $\sim$", written $a \sim b$, to mean that $(a, b) \in \sim$, and we say that "a is not related to b", written $a \not\sim b$, to mean that $(a, b) \notin \sim$. So for every $a, b \in A$ we have either $a \sim b$ or $a \not\sim b$. For example, $=$ and $\leq$ are relations on $\mathbb{R}$ with which you are familiar: for every real numbers x and y, either $x = y$ or $x \neq y$, and either $x \leq y$ or $x \not\leq y$. We often define a relation in this way, that is, by saying that "$\sim$ is the relation on <some set> in which $x \sim y$ if <some property>".

Example. The following are all valid relations.[1]

(i) The relation $R = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3)\}$ on the set $\{1, 2, 3\}$.
(ii) The relation $<$ on $\mathbb{R}$ given by the standard definition of $<$.
(iii) The relation $\leq$ on $\mathbb{R}$ given by the standard definition of $\leq$.
(iv) The relation $\mid$ on $\mathbb{Z}$ given by the standard definition of $\mid$ (i.e. $a \mid b$ means a divides b).
(v) The relation $\sim$ on $\mathbb{N}$ in which $x \sim y$ if $|x - y| \leq 2$.
(vi) The relation $\sim$ on $\mathbb{R}^2$ in which $(x, y) \sim (z, w)$ if both $x - z$ and $y - w$ are integers.
(vii) For a given graph G, the relation $\sim_e$ on $V(G)$ in which $u \sim_e v$ if $uv \in E(G)$.
(viii) For a given graph G, the relation $\sim_w$ on $V(G)$ in which $u \sim_w v$ if there is a walk from u to v in G.
(ix) The relation $\overset{r}{\equiv}$ on $\mathbb{Z}$ in which $x \overset{r}{\equiv} y$ if $x \equiv y \pmod{r}$.

There are three commonly-useful properties that relations can have; these are the following.

Definition. Suppose that $\sim$ is a relation on a set A. We say that

(i) $\sim$ is reflexive if for all $a \in A$ we have $a \sim a$.
(ii) $\sim$ is symmetric if for all $a, b \in A$ with $a \sim b$ we have $b \sim a$.
(iii) $\sim$ is transitive if for all $a, b, c \in A$ with $a \sim b$ and $b \sim c$ we have $a \sim c$.

So, for example, to prove that a given relation $\sim$ on A is transitive, one should suppose that a, b, c are elements of A such that $a \sim b$ and $b \sim c$, and (using the given definition of $\sim$) deduce from this that $a \sim c$. On the other hand, to prove that $\sim$ is not transitive one merely has to exhibit some $a, b, c \in A$ such that $a \sim b$ and $b \sim c$ but $a \not\sim c$. Likewise one should give a proof to show that $\sim$ is reflexive or symmetric, but a counterexample will suffice to prove that $\sim$ is not reflexive or not symmetric.

Example. (i) The relation $R = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3)\}$ on the set $\{1, 2, 3\}$ is reflexive since $1R1$, $2R2$ and $3R3$ (so xRx for every $x \in \{1, 2, 3\}$). Likewise, by quickly checking all possibilities we see that R is symmetric and transitive.

(ii) $<$ on $\mathbb{R}$ is not reflexive since, for example, $3 \not< 3$. It is not symmetric since, for example, $2 < 3$ but $3 \not< 2$. It is transitive since for every $a, b, c \in \mathbb{Z}$ with $a < b$ and $b < c$ we have $a < c$.

(iii) $\leq$ on $\mathbb{R}$ is reflexive since $n \leq n$ for every $n \in \mathbb{Z}$. Similarly as for $<$ it is not symmetric but is transitive.

[1] Note that $<, \leq$ and $\mid$ are standard notation symbols for the given relations, but the other relation symbols are just what I use in this chapter for convenience.

(iv) The relation $|$ on $\mathbb{Z}$ is reflexive, since for every $a \in \mathbb{Z}$ we have that a divides a. It is not symmetric since, for example, 2 divides 4 but 4 does not divide 2. It is transitive since for every $a, b, c \in \mathbb{Z}$ such that a divides b and b divides c we have a divides c.[2]

(v) The relation $\sim$ on $\mathbb{N}$ in which $x \sim y$ if $|x - y| \leq 2$ is reflexive, since for every $n \in \mathbb{N}$ we have $|n - n| = |0| = 0 \leq 2$, so $n \sim n$. It is symmetric as for every $n, m \in \mathbb{N}$ with $n \sim m$ we have $|n - m| \leq 2$, so $|m - n| \leq 2$, so $m \sim n$. It is not transitive since, for example $3 \sim 4$ and $4 \sim 6$ but $3 \not\sim 6$.

(vi) The relation $\sim$ on $\mathbb{R}^2$ in which $(x, y) \sim (z, w)$ if both $x - z$ and $y - w$ are both integers is reflexive since for every $(x, y) \in \mathbb{R}^2$ we have $x - x = y - y = 0$, so $(x, y) \sim (x, y)$. It is symmetric since for every $(x, y), (z, w) \in \mathbb{R}^2$ with $(x, y) \sim (z, w)$ we have that $x - z$ and $y - w$ are both integers, so $z - x$ and $w - y$ are both integers, so $(z, w) \sim (x, y)$. It is transitive since for every $(x, y), (z, w), (u, v) \in \mathbb{R}^2$ with $(x, y) \sim (z, w)$ and $(z, w) \sim (u, v)$ we have that $x - z, y - w, z - u, w - v$ are all integers, so $x - u = x - z + z - u$ and $y - v = y - w + w - v$ are both integers, so $(x, y) \sim (u, v)$.

(vii) Let G be a graph, and let $\sim_e$ be the relation on $V(G)$ in which $u \sim_e v$ if $uv \in E(G)$. Then G is not reflexive since for every $v \in V(G)$ we have $vv \notin E(G)$ (i.e. there is no edge from v to itself). It is symmetric since for every $u, v \in V(G)$ with $uv \in E(G)$ we have $vu \in E(G)$. Whether or not it is transitive depends on G. For example, if G is a path of length at least 3, then G is not transitive, since if we take u, v and w to be consecutive vertices on the path (appearing in that order) then $u \sim_e v$ and $v \sim_e w$ but $u \not\sim_e w$.

(viii) Let G be a graph, and let $\sim_w$ be the relation on $V(G)$ in which $u \sim_w v$ if there is a walk from u to v in G. Then $\sim_w$ is reflexive since for every $v \in V(G)$ the walk (v) is a walk in G from v to v. It is symmetric since for every $u, v \in V(G)$, if $(u, x_1, x_2, \ldots, x_i, v)$ is a walk in G from u to v then $(v, x_i, x_{i-1}, \ldots, x_1, u)$ is a walk in G from v to u. It is transitive since for every $u, v, w \in V(G)$, if $(u, x_1, x_2 \ldots, x_i, v)$ is a walk in G from u to v, and $(v, y_1, y_2, \ldots, y_j, w)$ is a walk in G from v to w, then $(u, x_1, x_2, \ldots, x_{i-1}, v, y_1, y_2, \ldots, y_j, w)$ is a walk in G from u to w.

(ix) The relation $\overset{r}{\equiv}$ on $\mathbb{Z}$ given by $x \overset{r}{\equiv} y$ if $x \equiv y \pmod r$ is reflexive, symmetric and transitive.[3]

Those relations which have all three properties are particularly useful.

Definition. Suppose that $\sim$ is a relation on a set A. We say that $\sim$ is an equivalence relation if it is reflexive, symmetric and transitive.

Example. $=$ is an equivalence relation on any set, as are the relations in (i), (vi), (viii) and (ix) of the previous example.

The principal reason that equivalence relations are useful is that they divide the set on which they are defined into sets called *equivalence classes* which form a *partition* of A, meaning that every element of A lies in precisely one equivalence class. We now define these terms formally.

Definition. A partition P of a set A is a set of non-empty subsets $X \subseteq A$ such that for every $a \in A$ there is precisely one $X \in P$ with $a \in X$.[4]

Informally, you can think of a partition as splitting up a set into one or more non-overlapping pieces.

Example. There are five possible partitions of $\{1, 2, 3\}$. These are

$$\begin{aligned} P_1 &= \{\{1, 2, 3\}\}, \\ P_2 &= \{\{1, 2\}, \{3\}\}, \\ P_3 &= \{\{1, 3\}, \{2\}\}, \\ P_4 &= \{\{2, 3\}, \{1\}\}, \\ P_5 &= \{\{1\}, \{2\}, \{3\}\}. \end{aligned}$$

[2] The latter statement was proved in the Algebra half of the module.

[3] See Lemma 3.5 from the Algebra half of the module.

[4] Equivalently, a partition P of A is a subset of $\mathscr{P}(A) \setminus \{\emptyset\}$ such that $\bigcup_{X \in P} X = A$ and for every $X, Y \in P$, either $X = Y$ or $X \cap Y = \emptyset$. As an exercise, check that these definitions are equivalent.

For every non-empty set A, $\{A\}$ is a partition of A, and $\{\{a\} : a \in A\}$ is a partition of A.

Definition. Suppose that $\sim$ is an equivalence relation on a set A. For every $a \in A$, the equivalence class of a is the set $[a]_\sim = \{b \in A : a \sim b\}$, that is, the set of all elements of A to which a is related (note that we must have $a \in [a]_\sim$ since $\sim$ is reflexive). The set of equivalence classes of $\sim$ (often called the quotient set of $\sim$), is then

$$A/\sim := \{[a]_\sim : a \in A\}.$$

Example. The equivalence classes of the relation $R = \{(1,1),(1,2),(2,1),(2,2),(3,3)\}$ on $\{1,2,3\}$ are

$$[1]_R = \{1,2\}$$
$$[2]_R = \{1,2\}$$
$$[3]_R = \{3\}$$

So the set of equivalence classes of R is[5] $\{1,2,3\}/R = \Big\{\{1,2\},\{3\}\Big\}$.

Example. The equivalence classes of the relation $\overset{3}{\equiv}$ on $\mathbb{Z}$ are

$$[-2]_{\overset{3}{\equiv}} = \{\dots,-5,-2,1,4,7,10,\dots\}$$
$$[-1]_{\overset{3}{\equiv}} = \{\dots,-7,-4,-1,2,5,8,\dots\}$$
$$[0]_{\overset{3}{\equiv}} = \{\dots,-6,-3,0,3,6,9,\dots\}$$
$$[1]_{\overset{3}{\equiv}} = \{\dots,-5,-2,1,4,7,10,\dots\}$$
$$[2]_{\overset{3}{\equiv}} = \{\dots,-7,-4,-1,2,5,8,\dots\}$$
$$[3]_{\overset{3}{\equiv}} = \{\dots,-6,-3,0,3,6,9,\dots\}$$
$$[4]_{\overset{3}{\equiv}} = \{\dots,-5,-2,1,4,7,10,\dots\}$$

So the set of equivalence classes of $\overset{3}{\equiv}$ is

$$\mathbb{Z}/\overset{3}{\equiv} = \Big\{\{\dots,-5,-2,1,4,7,10,\dots\},\{\dots,-7,-4,-1,2,5,8,\dots\},\{\dots,-6,-3,0,3,6,9,\dots\}\Big\}.$$

Example. Let G be a graph, and let $\sim_w$ be the relation on $V(G)$ in which $u \sim_w v$ if there is a walk from u to v in G. Then the equivalence classes of G are the vertex sets of the connected components of G, and the set $V(G)/\sim_w$ of equivalence classes of $\sim_w$ is the of these sets.

Theorem 9.1. *Suppose that $\sim$ is an equivalence relation on a set A. Then $A/\sim$ is a partition of A.*

Proof. As we noted in the definition, since $\sim$ is reflexive we have $a \in [a]_\sim$ for every $a \in A$. So $A/\sim$ is a set of non-empty subsets of A, and moreover every $a \in A$ is in at least one element of $A/\sim$ (namely $[a]_\sim$). It remains to show that every $a \in A$ is in at most one element of $A/\sim$. So fix an arbitrary $a \in A$; we want to show that a does not lie in any equivalence class which is *distinct* from $[a]_\sim$, or, to put it another way, that for every $b \in A$, if $a \in [b]_\sim$ then $[a]_\sim = [b]_\sim$.

So suppose that $b \in A$ is such that $a \in [b]_\sim$, which means that $b \sim a$; since $\sim$ is symmetric we also have $a \sim b$. Then:

(i) For every $c \in [a]_\sim$ we have $a \sim c$, which together with $b \sim a$ implies that $b \sim c$ (since $\sim$ is transitive). This means that $c \in [b]_\sim$.

(ii) For every $c \in [b]_\sim$ we have $b \sim c$, which together with $a \sim b$ implies that $a \sim c$. This means that $c \in [a]_\sim$.

[5] Remember that a set cannot have repeated elements!

So $[a]_\sim = [b]_\sim$, completing the proof.[6] □

Theorem 9.1 shows that an equivalence relation on a set A generates a partition of A; the next proposition shows that the reverse is also true. So for any set A, partitions of A and equivalence relations on A are essentially the same thing: each equivalence relation on A corresponds to a partition of A and vice versa.

Proposition 9.2. *Let A be a set and let P be a partition of A. Define a relation $\sim$ on A by*

$$a \sim b \textit{ if there is some } X \in P \textit{ with } a, b \in X.$$

Then $\sim$ is an equivalence relation on A, and $A/{\sim} = P$ (in other words, the equivalence classes of $\sim$ are the elements of P).

Proof. To see that $\sim$ is reflexive, let $a \in A$. Then by the definition of a partition there exists a set $X \in P$ with $a \in A$, so $a \sim a$.

To see that $\sim$ is symmetric, let $a, b \in A$ be such that $a \sim b$. Then there is some $X \in P$ with $a \in X$ and $b \in X$. But the order of these statements is irrelevant, that is, we have $b \in X$ and $a \in X$. So $b \sim a$, and we conclude that $\sim$ is symmetric.

Finally, to see that $\sim$ is transitive, let $a, b, c \in A$ be such that $a \sim b$ and $b \sim c$. The first implies that there is some $X \in P$ with $a \in X$ and $b \in X$, and the second implies that there is some $Y \in P$ with $b \in Y$ and $c \in Y$. But since there is precisely one set in P which contains b (by the definition of a partition), we must have $X = Y$, and so we have $c \in X$ as well as $a \in X$. So $a \sim c$, and we conclude that $\sim$ is transitive.

We conclude that $\sim$ is an equivalence relation on A. Finally, fix any $a \in A$, and let X be the unique member of P such that $a \in X$. Then by definition of $\sim$ we have $a \sim b$ if and only if $b \in X$. In other words, the equivalence class of a is $[a]_\sim = X$. So the set of equivalence classes of $\sim$ is $A/{\sim} = P$. □

[6] Note that we used all three parts of the definition of an equivalence relation in this proof. As an exercise, give examples to show that no two parts of the definition suffice to ensure that the set of 'equivalence classes' of $\sim$ is a partition of A.