# Combinatorics and Probability

Colva, Duncan, Ian, Peter, Ruth

October 15, 2014

# Contents

# About the Course

## Teaching:

- Lectures will be held at 11am. In odd numbered weeks there will be lectures on Monday, Wednesday and Friday. In even numbered weeks there will be lectures on Wednesday and Friday. Dr Roney-Dougal will give the first 12 lectures, and Dr King will give the final 12 lectures.
- Every other week there will be a *compulsory* small group tutorial, starting in week 2. Please sign up for a tutorial slot via MMS.
- Every other week there will be a *compulsory* examples class, starting in week 3. Please sign up for an examples class slot via MMS.
- In weeks 3 and 4 there will be a lecture in the maths computer classroom and in weeks 5 and 6 there will be a supervised computer classroom session for you to work on your Maple computer project. Please sign up for a computer classroom slot via MMS.

## Assessment:

- The final examination will count for 70%.
- The tutorial sheets for Weeks 4, 6, 8 and 10 will contain one or more assessed questions for you to hand in. Each sheet will count for 3% of your final mark, an extra 3% will be given to everyone who hands in genuine attempts for all four sheets.
- The Maple computer project will count for 15% of the final mark.

## Recommended books:

- Norman L Biggs, Discrete Mathematics, Oxford, 1985 or 2002.
- James A Anderson, Discrete Mathematics with Combinatorics, Prentice Hall, 2001.
- John A. Rice, Mathematical Statistics and Data Analysis, Belmont, CA: Brooks/Cole CENGAGE, 2007.
- Richard D. De Veaux, Paul F. Velleman & David E. Bock, Stats: Data and Models, Pearson/Addison Wesley, 2005.

All are available in the mathematics library.

**MMS:** See MMS for lecture notes, copies of handouts and additional information. You will be required to use MMS to upload your microlab projects. To log on to MMS, use your university username and password at

`https://www.st-and.ac.uk/mms/maths.html`

# Chapter 1

# Sets, Counting, Elementary Probability

In this introductory chapter we will establish some basic results on how to count sets, functions, and permutations, and introduce the axioms of probability theory. These results will then be used throughout the rest of the course.

**Example 0.1.** [Cards] Suppose that I have two cards: one is red on both sides; the other is red on one side and green on the other. Now I choose one card at random and place it down on the table. The visible (face up) side of the card is red. What is the probability that the other (face down) side of the card is also red?

We will answer this question later using the techniques developed in this chapter.

## 1. Unions of sets

For additional information on this section, see Biggs §6 and §10.

**Notation 1.1.** Let $A$ and $B$ be sets. We write $|A|$ for the *size* or *order* of $A$, the number of elements in $A$. We write $\emptyset$ for the empty set. It has no elements and size 0. We write $A \cup B$ for the *union* of $A$ and $B$: the set of all elements that lie in $A$ or $B$ (or both). We write $A \cap B$ for the *intersection* of $A$ and $B$: the set of elements in both $A$ and $B$. If $A$ is a subset of some fixed set, $\Omega$ then $A^c$ is the set of elements $\Omega$ that do not lie in $A$. We write $A \setminus B$ for the set of elements that lie in $A$ but not in $B$.

**Definition 1.2.** An *experiment* (or *trial*) is any process with a random outcome. The *sample space*, denoted by $\Omega$, is the set of all possible outcome of an experiment. Each $\omega \in \Omega$ is *sample point*: a possible outcome. A subset $A$ of $\Omega$ is an *event*. The emptyset $\emptyset$ is a *null event*: an impossible event.

**Example 1.3.** Consider an experiment where we roll one die and note the number shown. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$, so $|\Omega| = 6$.

Let $A$ be the event "the number shown is even". Then $A = \{2, 4, 6\}$ and $|A| = 3$.

**Definition 1.4.** Two sets $A$ and $B$ are *disjoint* (or *mutually exclusive*) if they have no elements (or events) in common, that is if $A \cap B = \emptyset$. The sets $A_1, A_2, \ldots, A_n$ are *pairwise disjoint* if for $1 \leq i \leq n$, $1 \leq j \leq n$ the sets $A_i$ and $A_j$ are disjoint whenever $i \neq j$.

**Theorem 1.5.** *For any two disjoint finite sets $A$ and $B$*

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

**Proof.** Since $A$ is finite, it has $m$ elements for some natural number $m$, so we may write $A = \{a_1, \dots, a_m\}$. Similarly $B = \{b_1, \dots, b_n\}$ for some $n$. Therefore since $A \cap B = \emptyset$ we have $A \cup B = \{a_1, \dots, a_m, b_1, \dots, b_n\}$, where no entries are repeated, and so $|A \cup B| = |A| + |B|$. ■

**Notation 1.6.** For positive integers $n$ and $m$, we write $n \mid m$ to indicate that $n$ divides $m$.

**Example 1.7.** What is wrong with the following argument?

Since one half of the numbers in the range $1, 2, \dots, 60$ are multiples of 2, including 2 itself, 29 of them cannot be primes. Since one third of them are multiples of 3, including 3 itself, 19 of them cannot be primes. Since $60 - 29 - 19 = 12$ there are at most 12 primes in $1, \dots, 60$.

**Answer:** The sets $\{n : 1 \leq n \leq 60,\ 2 \mid n\}$ and $\{n : 1 \leq n \leq 60,\ 3 \mid n\}$ are not disjoint.

Some primes: $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \dots$ Notice that there are more than 12 of them!

We now generalise Theorem 1.5 to more than two sets.

**Theorem 1.8.** *If $A_1, A_2, \dots, A_n$ are pairwise disjoint sets then*

$$|A_1 \cup \cdots \cup A_n| = |A_1| + \cdots + |A_n|.$$

**Proof.** We prove this by induction.

Base case: $n = 2$ is Theorem 1.5.

Inductive hypothesis: Suppose that the theorem holds for any collection of up to $n$ disjoint sets.

Inductive step: Consider $n + 1$ sets. By Theorem 1.5 it follows that

$$|A_1 \cup \cdots \cup A_{n+1}| = |(A_1 \cup \cdots \cup A_n) \cup A_{n+1}| = |A_1 \cup \cdots \cup A_n| + |A_{n+1}|.$$

Then by the inductive hypothesis

$$|A_1 \cup \cdots \cup A_{n+1}| = |A_1| + \cdots + |A_n| + |A_{n+1}|$$

as required, so the result holds for all $n$. ■

## 2. Axioms of Probability

For more information on this section, see Rice Sections 1.3 and 1.4 or Anderson Section 8.5.

**Definition 2.1.** Let $\Omega$ be a sample space. A *probability* $\mathbb{P}$ is any function which assigns a real number $\mathbb{P}(A)$ to each event $A$ such that:

**Axiom 1** $0 \leq \mathbb{P}(A) \leq 1$ for any event $A \subseteq \Omega$;

**Axiom 2** $\mathbb{P}(\Omega) = 1$ – this is the *honesty condition*;

**Axiom 3** If $A_1, \dots, A_n$ are a finite collection of mutually exclusive events, then,

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Sometimes a stronger version of Axiom 3 is needed:

**Axiom 3a** If $A_1, A_2, \ldots$ is an infinite sequence of disjoint events then,

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

In frequentist terms, the probability of an event $A$ happening is the long-term relative frequency with which the event occurs. A probability can also be interpreted in a non-frequentist, more subjective, framework. In any case, it always adheres to the conditions and properties described in this chapter.

**Lemma 2.2.** *Let $\Omega = \{\omega_1, \ldots, \omega_n\}$, let $A \subset \Omega$ be an event, and assume that each outcome is equally likely. Then*

$$\mathbb{P}(A) = \frac{\textit{number of sample points in } A}{\textit{number of sample points in } \Omega} = \frac{|A|}{n}.$$

**Proof.** Since each of the $n$ possible outcomes is equally likely, it follows from Axiom 2 that $\mathbb{P}(\omega_i) = 1/n$ for $1 \leq i \leq n$. Now by Axiom 3 we deduce

$$\mathbb{P}(A) = \sum_{\omega_i in A} \mathbb{P}(\{\omega_i\}) = |A| \cdot \frac{1}{n}.$$

■

**Example 2.3.** In Example 1.3 the probability of rolling an even number is 1/2.

**Remark 2.4.** If the events are *not* equally likely, then Lemma 2.2 will not hold, in general. For example, suppose that when tossing a coin there is a chance of $1/1,000,000$ chance of it staying permanently on its side. There are now three possible outcomes: heads, tails or side. Then $\mathbb{P}(\text{heads or tails})$ is much greater than 2/3!

**Theorem 2.5.** *Let $A$ and $B$ be events.*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.
3. *If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

**Proof.** 1. Since $A \cap A^c = \emptyset$, Axiom 3 says that $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$. But Axiom 2 then gives $\mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$. So $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$.

2. Each element of $A$ either lies in $B$, and so lies in $A \cap B$, or does not lie in $B$, and so lies in $A \cap B^c$. So $A = (A \cap B) \cup (A \cap B^c)$, and this is a disjoint union. The result now follows from Axiom 3.

3. Since $B$ contains $A$, we can write $B = A \cup (B \setminus A)$. Then by Axiom 3, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. The result now follows from Axiom 1 (all probabilities are non-negative). ■

**Notation 2.6.** We often write $\mathbb{P}(A \cap B) = \mathbb{P}(A, B) = \mathbb{P}(A \text{ and } B)$.

## 3. Inclusion–Exclusion

See Anderson §8.2 and §12.3 for more information on this section.

In this section we measure the size of unions of sets that are not necessarily disjoint. When we add the size of a set $S$ to the size of a set $T$, any elements that lie in both $S$ and $T$ will be counted twice. Using this, we can start thinking in more detail about the probability of pairs of events.

**Theorem 3.1.** *Let $S$ and $T$ be sets. Then the number of elements in $S \cup T$ is equal to $|S| + |T| - |S \cap T|$, that is*

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

**Proof.** The set

$$S \cup T = (S \setminus T) \cup (T \setminus S) \cup (S \cap T),$$

and the three sets in this expression are pairwise disjoint. Therefore, by Theorem 1.8

$$|S \cup T| = |S \setminus T| + |T \setminus S| + |S \cap T|.$$

Now, $|S| = |S \setminus T| + |S \cap T|$, and similarly for $T$, so rearranging we get

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

as required. ■

**Theorem 3.2.** *For any 2 events $A$ and $B$,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

**Proof.** This is an exercise on Tutorial Sheet 1. Base your proof on the proof of Theorem 3.1. ■

**Notation 3.3.** For a fraction $n/m$, we write $\lfloor n/m \rfloor$ to denote the largest whole number that is less than or equal to $n/m$; i.e. the result of rounding down $n/m$ to the nearest whole number.

**Example 3.4.** Find the probability that a positive integer less than 1001 is divisible by 3 or 5.

Let $A$ be the set of positive integers that are less than 1001 and divisible by 3. Then the size of $A$ is $\lfloor 1000/3 \rfloor = 333$, so $\mathbb{P}(\text{divisible by } 3) = 333/1000$. Let $B$ be the set of positive integers that are less than 1001 and divisible by 5. Then the size of $B$ is $\lfloor 1000/5 \rfloor = 200$, so $\mathbb{P}(\text{divisible by } 5) = 1/5$. The elements in $A \cap B$ are divisible by both 5 and 3, so they are divisible by 15. Therefore $|A \cap B| = \lfloor 1000/15 \rfloor = 66$, and $\mathbb{P}(\text{divisible by } 15) = 66/1000 = 33/500$. Thus $\mathbb{P}(\text{divisible by 3 or 5}) = 333/1000 + 1/5 - 33/500 = 467/1000$

We now generalise Theorem 3.1 to three sets.

**Theorem 3.5 (The inclusion-exclusion principle)** *Let $A_1, A_2, A_3$ be finite sets. The number of elements in $A_1 \cup A_2 \cup A_3$ is*

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$

*That is, add the sizes of the individual sets, then subtract the sizes of their intersections in pairs, then add the size of their three-way intersection.*

**Proof.** Let $B = A_1 \cup A_2 \cup A_3$. We want to show that each element of $B$ is counted exactly once on the right hand side of the above equation. Assume that an element $x$ is contained in exactly $m$ of the sets $A_1, A_2, A_3$. In $|A_1| + |A_2| + |A_3|$ the element $x$ is counted $m$ times, once for each of the $m$ sets.

If $m = 1$ then the element $x$ is counted in none of the remaining terms, so is counted once in total. If $m = 2$ then $x$ is counted in exactly one of the sets $A_i \cap A_j$, $i \neq j$, and not at all in $A_1 \cap A_2 \cap A_3$, so $x$ is counted $2 - 1 = 1$ times in total. If $m = 3$ then $x$ is in all three of the sets $A_i \cap A_j$ with $i \neq j$, and is also in $A_1 \cap A_2 \cap A_3$, so is counted $3 - 3 + 1 = 1$ times in total. ■

**Example 3.6.** In a survey of 100 students, it was found that 50 take basket-weaving, 53 take history of volleyball, 42 take yoghurt-making, 15 take basket-weaving and yoghurt-making, 20 take yoghurt-making and history of volleyball, 25 take history of volleyball and basket-weaving, and 5 take all three.

(a) How many students take at least one of these subjects?

By the inclusion-exclusion principle, we must add together the three numbers for basket-weaving, history of volleyball and yoghurt-making, then subtract the numbers for people taking two subjects and finally add the number taking all three. This gives:

$$50 + 53 + 42 - 15 - 20 - 25 + 5 = 90.$$

(b) How many students take none of them?

There are 100 students in total, and 90 take at least one subject, so $100 - 90 = 10$ students take none.

(c) How many students take history of volleyball only?

We start with the number taking history of volleyball and subtract the number who take history of volleyball and yoghurt-making and the number who take history of volleyball and basket-weaving. In doing so we have twice taken away the number who do all three, so we add it back in again:

$$53 - 20 - 25 + 5 = 13.$$

## 4. Conditional Probability and Bayes' Theorem

For more information on this section, see Rice, Sections 1.5 and 1.6.

**Definition 4.1.** Let $A$ and $B$ be any two events such that $\mathbb{P}(B) > 0$. Then, the *conditional probability of $A$, given that the event $B$ has occurred* is written as $\mathbb{P}(A|B)$ and is defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

If $\mathbb{P}(B) = 0$, then $\mathbb{P}(A|B)$ is not defined.

**Example 4.2.** Recall Example 3.6.

(a) Given that a student takes basket-weaving ($B$), what is the probability that they also take history of volleyball ($H$)?

We need to calculate $\mathbb{P}(H|B) = \mathbb{P}(H \cap B)/\mathbb{P}(B)$. There are 25 students studying both subjects, and 50 studying basket-weaving, so $\mathbb{P}(H|B) = 25/50 = 1/2$.

(b) What is the probability that the same student does *not* take history of volleyball? $1 - 1/2 = 1/2$.

**Theorem 4.3 (The multiplication rule for probabilities)** *1. Let $A$ and $B$ be events. Then*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B).$$

*2. Let $A_1, A_2, \ldots, A_n$ be events, then*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

**Proof.** 1. This is just a rearrangment of Definition 4.1.
2. This is a direct calculation. Let's write out the right hand side in more detail, using Definition 4.1 for each part.

$$\begin{aligned}
&\mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1\cap A_2)\times\cdots\times\mathbb{P}(A_n|A_1\cap\cdots\cap A_{n-1})\\
&\qquad= \mathbb{P}(A_1)\cdot\tfrac{\mathbb{P}(A_2\cap A_1)}{\mathbb{P}(A_1)}\cdot\tfrac{\mathbb{P}(A_3\cap(A_1\cap A_2))}{\mathbb{P}(A_1\cap A_2)}\cdot\dots\cdot\tfrac{\mathbb{P}(A_n\cap(A_1\cap\cdots\cap A_{n-1}))}{\mathbb{P}(A_1\cap\cdots\cap A_{n-1})}\\
&\qquad= \mathbb{P}(A_1\cap\cdots\cap A_n),
\end{aligned}$$

since $X\cap Y = Y\cap X$ for all sets $X$ and $Y$. ■

**Example 4.4.** Consider the set of all possible genders of children for families with two children, so that

$$\Omega = \{(g,g),(g,b),(b,g),(b,b)\}.$$

We shall assume that each outcome is equally likely. Given that a family has a boy, what is the probability that both children are boys?

Let $A$ be the event "both children are boys" and $B$ the event "family has at least one boy". Then $A = \{(b,b)\}$, $B = \{(b,b),(g,b),(b,g)\}$ and $A\cap B = \{(b,b)\}$. Since each outcome is equally likely, $\mathbb{P}(B) = \frac{3}{4}$ and $\mathbb{P}(A\cap B) = \frac{1}{4}$. Then,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1}{3}.$$

**Definition 4.5.** Two events $A$ and $B$ are *independent* if

$$\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

**Lemma 4.6.** *The events $A$ and $B$ are independent if and only if*

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

In other words, two events are independent if and only if knowing that event $B$ has occurred does not give any information relating to the occurrence of $A$.

**Proof.** Suppose that $A$ and $B$ are independent. Then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

Conversely, suppose that $\mathbb{P}(A|B) = \mathbb{P}(A)$. Then

$$\mathbb{P}(A\cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B),$$

so $A$ and $B$ are independent. ■

**Example 4.7.** Recall Example 3.6 again. Are the events "Student $S$ is taking yoghurt-making ($Y$)" and "Student $S$ is taking basket-weaving ($B$)" independent?

Here $\mathbb{P}(B) = 50/100 = 1/2$, whilst $\mathbb{P}(Y) = 42/100 = 21/50$, and $P(B\cap Y) = 15/100 = 3/20$. Since $3/20 \neq (1/2)\cdot(21/50)$, the events are *not* independent.

**Remark 4.8.** Suppose that $A$ and $B$ are disjoint events. Then this does NOT imply that $A$ and $B$ are independent.

**Example 4.9.** Suppose that we toss a coin. Let $A$ be the event of a head ($H$) and $B$ the event of a tail ($T$). Clearly, $A$ and $B$ are disjoint and

$$\mathbb{P}(A) = \frac{1}{2} = \mathbb{P}(B).$$

However,

$$\mathbb{P}(A\cap B) = 0,$$

since this is the probability of a $H$ AND a $T$, which is impossible.

**Definition 4.10.** Let $A_1, \ldots, A_n$ be subsets of a set $\Omega$. If $\Omega = \cup_{i=1}^n A_i$ and the sets $A_i$ are pairwise disjoint then $A_1, \ldots, A_n$ are a *partition* of $\Omega$.

**Example 4.11.** 1. Let $\Omega = \{1, 2, 3, 4\}$. Then $\{\{1, 2\}, \{3\}, \{4\}\}$ form a partition of $\Omega$, but $\{\{1, 2\}, \{4\}\}$ and $\{\{1, 2\}, \{2, 3, 4\}\}$ do not.
2. The suits form a partition of a deck of cards (without jokers).

**Theorem 4.12 (The Law of Total Probability)** *Let $\Omega = \cup_{i=1}^n A_i$ be a partition of $\Omega$ into non-empty parts, and let $B$ be an event. Then*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

**Proof.** Since $\cup_{i=1}^n A_i$ is a partition, $B \cap A_i$ and $B \cap A_j$ are mutually exclusive events whenever $i \neq j$. Hence, by Axiom 3, we can write $\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i)$. Now apply the Multiplication Rule (Theorem 4.3) to each term. ■

**Theorem 4.13 (Bayes' Theorem)** *Let $A_1, A_2, \ldots, A_n$ be a partition of $\Omega$. Let $B$ be an event such that $\mathbb{P}(B) > 0$. Then*

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

**Proof.** By definition, $\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \cap A_j)}{\mathbb{P}(B)}$. Now use the Multiplication Rule (Theorem 4.3) to rewrite $\mathbb{P}(B \cap A_j)$ as $\mathbb{P}(B|A_j)\mathbb{P}(A_j)$, and use the Law of Total Probability (Theorem 4.12) to rewrite $\mathbb{P}(B)$ as $\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)$. ■

This has become a very important formula in recent years within statistical research, on which a whole range of statistical methodology rests (although the theorem actually dates back about 250 years). See honours module MT4531/MT5831 (Advanced) Bayesian Inference.

## 5. The pigeonhole principle

Next we have a key theorem about distributing objects:

**Theorem 5.1 (Pigeonhole principle)** *Let $m, n, r$ be natural numbers with $m > nr$. If $m$ objects are partitioned into $n$ sets, then at least one set contains at least $r + 1$ objects.*

**Proof.** If there are no more than $r$ objects per set then the maximum number of objects is at most $nr$. However, $m > nr$. ■

As an easy application, we can see that in any set of 13 people at least 2 must have birthdays in the same month.

We finish this section with a series of applications of the pigeonhole principle.

**Example 5.2.** Show that in any set of six people there are either three people who all know each other or three mutual strangers.
**Answer:** Let the people be $A, B, C, D, E$ and $F$. Consider person $A$. Divide the remaining 5 people into two sets: one contains the people that know $A$, and the other contains people that $A$ does not know. Then by the pigeonhole principle, since $5 > 2 \times 2$ at least one set must contain at least three people.

Suppose first that at least three people know $A$. Without loss of generality we may suppose that $B$, $C$ and $D$ all know $A$. Then either none of $B$, $C$ and $D$ know each other (in which case $B$, $C$ and $D$ are mutual strangers) or at least two of them, say $B$ and $C$, know each other (in which case $A$, $B$ and $C$ all know each other).

Suppose instead that at least three people do not know $A$. Without loss of generality we may suppose that $B$, $C$ and $D$ do not know $A$. Then either $B$, $C$ and $D$ all know each other, in which case we have found a set of three people who all know each other, or at least two of them, say $B$ and $C$, do not know each other, in which case $A$, $B$ and $C$ are mutual strangers.

**Example 5.3.** Consider a five a side football competition, with the rather odd rule that the members of each team must have birthdays in the same month. How many people need to turn up to be sure of making at least one team?
**Answer:** We calculate that $(12 \times 4) + 1 = 49$, so if at least 49 people show up for the competition then at least one team of 5 can be made.

**Example 5.4.** Show that in any finite set $X$ of people there are two members who have the same number of friends in $X$.
**Answer:** Let $|X| = m$, and consider $m$ sets $B_0, B_1, \ldots, B_{m-1}$, such that the $i$th set contains the people with $i$ friends.

If $B_0$ contains someone then at least one person knows nobody. Therefore noone knows everybody, and hence $B_{m-1}$ is empty. If $B_{m-1}$ contains someone then that person knows everybody, so $B_0$ is empty.

Therefore at least one of $B_0$ and $B_{m-1}$ is empty, and so at most $m-1$ sets contain people. Hence by the pigeonohole principle there is a set containing (at least) two people.

## 6. Counting sets of pairs

See Biggs §10.2 for more information on this section.

**Definition 6.1.** Let $X$ and $Y$ be sets, then by $X \times Y$ we denote the *direct product* of $X$ and $Y$. This is the set of *ordered pairs*

$$X \times Y = \{(x, y)\ : x \in X, y \in Y\}.$$

Note that the order of the elements in the pair is important: $(x, y) \neq (y, x)$.

**Notation 6.2.** We write $A \subseteq B$ to mean that $A$ is a *subset* of $B$: all elements of $A$ are elements of $B$. We write $x \in A$ to mean that $x$ is an element of $A$.

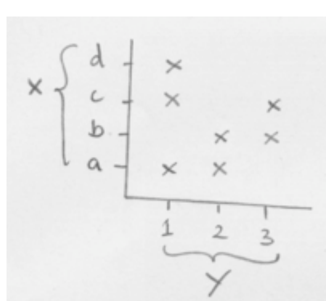
Let $S \subseteq X \times Y$, then for $x \in X$ and $y \in Y$ we denote by

$$r_x(S) = |\{y : (x, y) \in S\}|,\ c_y(S) = |\{x : (x, y) \in S\}|.$$

If we represent $X \times Y$ as a grid with rows labelled by elements of $X$ and columns labelled by elements of $Y$ then $r_x(S)$ is the number of elements of $S$ in the row labelled $x$ and $c_y(S)$ is the number of elements of $X$ in the column labelled $y$.

**Example 6.3.** Let $X = \{a, b, c, d\}$, $Y = \{1, 2, 3\}$. Then one choice for $S$ would be

$$\{(a, 1), (b, 2), (c, 3), (d, 1), (a, 2), (b, 3), (c, 1)\}.$$



In this case $r_b(S) = 2$, $c_1(S) = 3$.

**Remark 6.4.** Notice that our labelling here is similar to the way that we index entries of a matrix: the *first* entry describes the row and the *second* describes the column. It is the opposite of the way we normally label $x$ and $y$ coordinates when drawing $y = f(x)$.

We collect some facts about sets of ordered pairs:

**Theorem 6.5.**

1. $|S| = \sum_{x\in X} r_x(S) = \sum_{y\in Y} c_y(S)$.
2. *If there exists a constant $r$ such that $r_x(S) = r$ for all $x \in X$ then*
$$|S| = r|X|.$$
3. *If there exists a constant $c$ such that $c_y(S) = c$ for all $y \in Y$ then*
$$|S| = c|Y|.$$
4. *The multiplication principle for sets: $|X \times Y| = |X||Y|$.*
5. $|X_1 \times X_2 \times \cdots \times X_n| = |X_1||X_2|\cdots|X_n|$.

**Proof.**

1. The set of elements of $S$ with first coordinate $x$ has size $r_x(S)$. Since the set of elements with first coordinate $a$ is disjoint from the set of elements with first coordinate $b$ whenever $a \neq b$, the fact that $|S|$ is equal to the sum of the sizes follows from Theorem 1.8. The result for columns is proved in the same way.
2. If $r_X(S) = r$ for all $x \in X$ then there are $|X|$ terms in the sum from part 1, and each term contributes $r$. So the result follows.
3. The argument for $y$ is similar to part 2.
4. In the special case when $S = X \times Y$ the row total is $|Y|$ for each $x \in X$, so the result follows by part 2.
5. Prove this by induction.

■

**Example 6.6.** In a calculus class, 32 of the students are boys. Each boy knows five girls and each girl knows eight boys. If I choose a student at random, what is the probability that the student is a girl?
**Answer:** Here $\Omega$ is the set of students.

Let $g$ be the number of girls. Let

$$S = \{(x, y) : x \text{ is a boy}, y \text{ is a girl, and they know each other}\}.$$

Then $r_x(S) = 5$ for all boys $x$, and $c_y(S) = 8$ for all girls $y$. Therefore

$$|S| = 32 \cdot r_x(S) = 32 \cdot 5 = g \cdot c_y(S) = g \cdot 8$$

so $g = 20$. The total number of students is therefore $|\Omega| = 20 + 32 = 52$. So $\mathbb{P}(\text{student is a girl}) = 20/52$.

**Example 6.7.** A licence plate consists of two letters followed by two numbers followed by three letters. How many different licence plates are there?
**Answer:** Assuming that all letters of the alphabet are used, there are $26^2 \cdot 10^2 \cdot 26^3 = 26^5 \cdot 10^2$.

How many different licence plates are there per year?
**Answer:** There are two numbers used per year, so $26^5 \cdot 2$.

**Example 6.8.** In an experiment, we randomly choose an integer between 0 and 1000 (inclusive). What is the probability that is has exactly one digit equal to 6?
**Answer:** Here $\Omega = \{0, 1, \ldots, 1000\}$

First count the outcomes satisfying the condition. One digit numbers: 1.
Two digit numbers: $6X$ or $Y6$ where $X \in \{0, 1, 2, 3, 4, 5, 7, 8, 9\}$ and $Y \in \{1, 2, 3, 4, 5, 7, 8, 9\}$ gives $9 + 8 = 17$.
Three digit numbers: $6XX$ or $Y6X$ or $YX6$ gives $(9 \cdot 9) + (8 \cdot 9) + (8 \cdot 9) = 225$.
Hence the total is $1 + 17 + 225 = 243$.
So $\mathbb{P}(\text{exactly one digit } 6) = 243/1001$.

The following is an immediate consequence of Theorem 6.5.

**Corollary 6.9.** *Suppose that we perform $n$ experiments where experiment $i$ has $|\Omega_i| = r_i$ outcomes for $1 \leq i \leq n$. Then in total there are $|\Omega_1 \times \cdots \times \Omega_n| = r_1 \cdots r_n$ possible outcomes for the $n$ experiments together.*

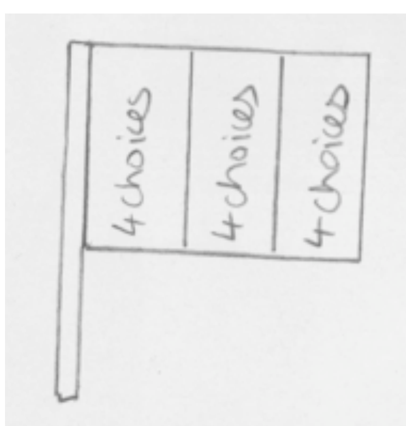
## 7. Functions, subsets and permutations

For more information on this section, see Biggs §§10.4–10.6.

**Theorem 7.1.** *Let $X$ and $Y$ be finite nonempty sets, with $|X| = m$ and $|Y| = n$. If $F$ is the set of all functions from $X$ to $Y$ then $|F| = n^m$.*

**Proof.** Let $X = \{x_1, \ldots, x_m\}$. A function $f : X \to Y$ is uniquely determined by the ordered $m$-tuple $(f(x_1), f(x_2), \ldots, f(x_m))$ of elements of $Y$. This $m$-tuple can be any element of $Y \times \cdots \times Y$ (where there are $m$ factors), so the number of functions is $|Y \times \cdots \times Y| = |Y^m| = n^m$ by Theorem 6.5.5. ■

**Question** How many national flags can be contructed from three equal vertical strips using the colours red, white, blue and green (assume repeats are allowed)?




**Answer:** $4^3 = 64$.

**Example 7.2.** Keys are made by cutting incisions of various depths in a number of positions on a blank key. If there are eight possible depths, how many positions are required to make 1000000 different keys?
**Answer:** Let $p$ be the number of positions. We need $8^p \geq 1000000$, so $p \log 8 \geq \log 1000000$ and so $p \geq 6.64$. Thus at least 7 positions are needed.

**Theorem 7.3.** *If $X = \{x_1, \ldots, x_n\}$ is a finite set with $n$ elements, then $X$ has precisely $2^n$ subsets (including $X$ and $\emptyset$).*

**Proof.** There is a correspondence between the subsets of $X$ and (ordered) $n$-tuples of 0s and 1. We match up $S \subseteq X$ with $W(S) = a_1 a_2 \ldots a_n$, where $a_i = 1$ if $x_i \in S$ and $a_i = 0$ if $x_i \notin S$. The map $S \mapsto W(S)$ is a bijection between the set of subsets of $X$ and the set of ordered $n$-tuples of 0s and 1s. Therefore, the number of subsets is equal to the number of $n$-tuples, which is $2^n$ by Theorem 6.5.5. ■

As an example of the correspondence in the previous theorem, let $X = \{1, 2, \ldots, 7\}$ and $S = \{1, 3, 4, 6\}$, then $W(S) = 1011010$.

**Example 7.4.** We list all subsets of $\{1, 2, 3\}$ and their corresponding 3-tuples:

| Subset | Word |
|---|---|
| $\emptyset$ | 000 |
| $\{1\}$ | 100 |
| $\{2\}$ | 010 |
| $\{3\}$ | 001 |
| $\{1, 2\}$ | 110 |
| $\{1, 3\}$ | 101 |
| $\{2, 3\}$ | 011 |
| $\{1, 2, 3\}$ | 111 |

**Theorem 7.5.** *Let $m, n \in \mathbb{N}$, with $n \geq m$. The number of ordered selections,* without repetition, *of $m$ objects from a set $X$ of size $n$ is*

$$n(n-1)\cdots(n-m+1).$$

**Proof.** The first selection can be any one of the $n$ objects in $X$. The second must be one of the remaining $n-1$ objects, since repetition is not allowed. Similarly, when we come to select object number $m$ a total of $m-1$ objects have already been selected, and so there are $n-(m-1) = n-m+1$ objects to choose from. ■

**Corollary 7.6.** *There are $n! = n \cdot (n-1) \cdots 2 \cdot 1$ permutations of elements of an $n$-element set.*

**Proof.** Set $n = m$ in Theorem 7.5. ■

**Example 7.7.** In how many ways can one select a batting order of 11 from a pool of 14 cricketers?
**Answer:** $14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \sim 1.45 \times 10^{10}$.

**Example 7.8.** Suppose that there are $m$ girls and $n$ boys in a class. What is the number of ways of arranging them in a line so that all of the girls stand together?
**Answer:** The $m$ girls can be arranged in a line in $m!$ ways. Then consider them as a single block, and arrange the $n+1$ objects: $m!(n+1)!$

# Chapter 2

# Recursion and generating functions

## 1. Binomial coefficients

See Biggs §11.1 for more information on this section.

In this section we consider the number of ways of choosing subsets of a fixed size from a set of size $n$, and prove some easy consequences.

**Definition 1.1.** The number of $r$-element subsets of an $n$-element set is denoted by $\binom{n}{r}$, pronounced "$n$ choose $r$". The numbers $\binom{n}{r}$ are called the *binomial numbers* or *binomial coefficients*.

We will prove later that

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

**Example 1.2.** Consider the set $\{1,2,3,4\}$. The subsets of size two are: $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$ and $\{3,4\}$. So $\binom{4}{2} = 6 = 4!/(2!2!)$.

**Lemma 1.3.**

$$\binom{n}{0} = 1,\ \binom{n}{1} = n,\ \binom{n}{n} = 1,\ \binom{n}{r} = \binom{n}{n-r},\ \binom{n}{r} = 0 \text{ for } r > n.$$

**Proof.** Exercise! Use Definition 1.1. ■

**Theorem 1.4.** *Let $r, n \in \mathbb{N}$, with $1 \leq r \leq n-1$. Then*

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

**Proof.** Let $|X| = n$ and choose $x \in X$. We define

$$U = \{A \ : \ A \subseteq X,\ x \in A,\ |A| = r\}$$

so that $|U| = \binom{n-1}{r-1}$. Also, let

$$V = \{A \ : \ A \subseteq X,\ x \notin A, |A| = r\},$$

so that $|V| = \binom{n-1}{r}$.

Then $U \cap V = \emptyset$, and

$$\binom{n}{r} = |U| + |V| = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

■

The binomial numbers occur in a famous mathematical object called *Pascal's triangle*. Here are the first 7 rows of the triangle, which is infinite.

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}
$$

Each entry is made by summing the entries above it — we assume that all entries outside the triangle are 0. The $(n+1)$st row has entries $\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n}$. Note that by convention $0! = 1$.

**Theorem 1.5.** *For $n, r \in \mathbb{N}$ with $1 \leq r \leq n$:*

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

**Proof.** We use induction on $n$.

Base case, $n = 1$. We have $\binom{1}{1} = 1 = \frac{1!}{1!0!}$.

Assume true for $n = k$, that is assume that $\binom{k}{r} = \frac{k!}{r!(k-r)!}$ for $1 \leq r \leq n$.

Now test $n = k+1$. If $r \neq k+1$ then we calculate:

$$
\begin{aligned}
\binom{k+1}{r} &= \binom{k}{r-1} + \binom{k}{r} \text{ by Thm 1.4} \\
&= \frac{k!}{(r-1)!(k-r+1)!} + \frac{k!}{r!(k-r)!} \\
&= \frac{(k!r) + (k!(k-r+1))}{r!(k-r+1)!} \\
&= \frac{k!(r+k-r+1)}{r!(k-r+1)!} \\
&= \frac{(k+1)!}{r!(k+1-r)!}
\end{aligned}
$$

If $r = k+1$ we perform the same calculation but with $\binom{k}{k+1} = 0$ instead of the formulae in the second line.

So the result follows for all $n$. ■

**Example 1.6.** $\binom{15}{5} = (15 \cdot 14 \cdot 13 \cdot 12 \cdot 11)/(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) = 3003$.

**Example 1.7.** In how many ways can 10 people be divided into two basketball teams of five?

**Answer:** $\binom{10}{5} = (10 \cdot 9 \cdot 8 \cdot 7 \cdot 6)/(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) = 252$. This is the number of ways of picking a set of size 5. However, the two teams would be the same if we had picked the *remaining* set of size 5, rather than the one we counted here. So in fact we divide by 2 to get 126 pairs of teams.

**Theorem 1.8.** *For all $n \geq 0$*

$$\sum_{i=0}^{n} \binom{n}{i} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n.$$

**Proof.** We note that $\binom{n}{r}$ is the number of $r$-element subsets of a set of size $n$. Then by Theorem 7.3 the total number of subsets is $2^n$. ■

One reason why the binomial numbers are so famous is the following:

**Theorem 1.9 (The Binomial Theorem)** $(x+y)^n = \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}x^0 y^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$.

**Proof.** Consider writing out $(x+y)^n$ as a product of $n$ brackets, and multipliying. The coefficient of $x^{n-i}y^i$ is the number of ways of making $x^{n-i}y^i$ when the brackets are multiplied out. To make $x^{n-i}y^i$ we choose $y$ from exactly $i$ of the brackets, and $x$ from the remaining brackets. So the number of ways we can do this is the number of ways of choosing $i$ elements from a set of size $n$, namely $\binom{n}{i}$. ■

We can use Theorem 1.9 to get an alternative proof of Theorem 1.8: set $x = y = 1$ in the Binomial Theorem to get $2^n$ on the left hand side, and $\sum_{i=0}^{n}\binom{n}{i}$ on the right hand side.

The following theorem will be extremely important in later sections.

**Theorem 1.10.** *If $n$ is any positive integer, then*

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

**Proof.** Set $x = 1$ and $y = -x$ in Theorem 1.9. ■

Finally, we show how to use a counting argument to prove a result about binomial coefficients.

**Theorem 1.11.** *Let $n$, $m$ and $k$ be positive integers, with $k \leq n+m$. Then*

$$\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k}.$$

**Proof.** Take a set of $n+m$ numbers, of which $n$ are odd and $m$ are even. We count the number of subsets of size $k$.

On one hand, there are $\binom{n+m}{k}$ such subsets, by the definition of the binomial coefficients.

On the other, we can choose a subset of size $k$ by first choosing a number $i$ between 0 and $k$, and then choosing $i$ odd numbers and $k-i$ even numbers. The number of ways of doing this is $\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}$.

These two ways of counting must give the same answer, so the result follows. ■

## 2. Choosing with repetition

See Andersen, Section 8.7, for more information about this section.

In this section we'll consider the possibility of choosing $r$ objects from a set of size $n$, where we are allowed to choose the same object more than once, but the order in which we choose them does not matter.

**Definition 2.1.** A *multiset* is a collection of objects where each object may be repeated more than once, but the order of the objects does not matter.

**Example 2.2.** The multisets of size two with elements chosen from $\{1,2,3,4\}$ are

$$1,1 \quad 1,2 \quad 1,3 \quad 1,4 \quad 2,2 \quad 2,3 \quad 2,4 \quad 3,3 \quad 3,4 \quad 4,4.$$

There are 10 of them.

To count the number of multisets of a given size, we first prove an unexpected-looking theorem about solving certain types of equations.

**Theorem 2.3.** *The number of solutions of $x_1+\cdots+x_n = r$ in non-negative integers $x_i$ is*

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

**Proof.** We can represent a solution $x_1, x_2, \ldots, x_n$ by a sequence of 0s and 1s:

$$\underbrace{0,0,\ldots,0}_{x_1\ 0s},1,\underbrace{0,0,\ldots,0}_{x_2\ 0s},1,\ldots,\underbrace{0,0,\ldots,0}_{x_n\ 0s}.$$

Corresponding to $x_1 + \cdots + x_n = r$, there will be $n-1$ 1s and $r$ 0s, and so each sequence has length $n+r-1$.

Conversely, any sequence containing $n-1$ 1s and $r$ 0s describes exactly one solution to the equation $x_1+\cdots+x_n = r$, where all of the $x_i$ are non-negative. So the number of such sequences is exactly the number of solutions of the equation.

The $r$ 0s can be in any of the $n+r-1$ places, so the number of such sequences is $\binom{n+r-1}{r}$. ■

**Example 2.4.** Consider the equation $x_1 + x_2 + x_3 + x_4 = 5$. As an example of the sequences defined in the previous proof, the solution $x_1 = 2$, $x_2 = 0$, $x_3 = 2$, $x_4 = 1$ corresponds to the sequence 00110010. The total number of solutions in non-negative integers is

$$\binom{4+5-1}{5} = \binom{8}{5} = \binom{8}{3} = 56.$$

It's now easy to prove the following theorem.

**Theorem 2.5.** *The number of unordered choices of $r$ elements from $n$ elements, with repetitions allowed, is*

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

**Proof.** Any choice consists of $x_1$ choices of the first object, $x_2$ of the second, and so on, subject to the conditions that $x_i \geq 0$ for $1 \leq i \leq n$ and

$$x_1 + x_2 + \cdots + x_n = r.$$

So the required number is the number of solutions of $x_1+\cdots+x_n = r$ in non-negative integers. The result now follows from Theorem 2.3. ■

**Example 2.6. Question:** How many solutions are there of $x + y + z = 17$ in **positive** integers?

**Answer:** Here we require $x \geq 1$, $y \geq 1$ and $z \geq 1$, so let's define $x = u+1$, $y = v+1$ and $z = w+1$. The equation becomes

$$(u+1)+(v+1)+(w+1) = 17 \quad \Rightarrow \quad u+v+w = 14.$$

We need solutions to this for $u, v, w$ non-negative, so we can use Theorem 2.3:

$$\binom{14+3-1}{14} = \binom{16}{2} = 120.$$

**Example 2.7.** A florist stocks 10 types of flowers. How many different bouquets of 12 flowers can be made?

Here we are interested in a multiset of size 12, with 10 types of object, so by Theorem 2.5 there are

$$\binom{10+12-1}{12} = \binom{21}{12} = 293930.$$

## 3. Introduction to recursion

See Biggs, §19.1, §19.2 for more information on recursive functions.

Recursive functions occur in both programming and mathematics. They are used to define functions on natural numbers. The basic idea of a recursive function is to state $f(0)$ explicitly, and then give a formula which expresses $f(n+1)$ in terms of $f(n)$ (and maybe also in terms of $f(n-1), f(n-2), \ldots$).

We will illustrate this idea with a series of examples.

**Example 3.1.** Let $a$ be any natural number. Define

$$f(0) = 1, \quad f(n+1) = af(n).$$

Then $f(1) = a$, $f(2) = a^2$, and one may prove by induction that $f(n) = a^n$.

**Example 3.2.** Define a recursive function $f$ as follows:

$$f(0) = 1, \quad f(n+1) = (n+1)f(n).$$

Then $f(1) = 1$, $f(2) = 2$, and one may prove by induction that $f(n) = n!$

**Example 3.3.** Express the function

$$f(n) = \frac{n(n+1)(2n+1)}{6}$$

as a recursive function.

First, we calculate a few values.

$$\begin{array}{ll} f(1) & = 1 \\ f(2) & = 5 = f(1) + 2^2 \\ f(3) & = 14 = f(2) + 3^2 \end{array}$$

Looking at this, it looks as if the function can be recursively defined as $f(1) = 1$, $f(n+1) = f(n) + (n+1)^2$. We need to prove this.

$$\begin{array}{ll} f(n) + (n+1)^2 & = \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \\ & = \frac{n+1}{6}\left(n(2n+1) + 6(n+1)\right) \\ & = \frac{n+1}{6}(\left(2n^2 + n + 6n + 6\right) \\ & = \frac{(n+1)(n+2)(2n+3)}{6} \\ & = \frac{(n+1)(n+2)(2(n+1)+1)}{6} \end{array}$$

So the result holds for all $n$.

It is also possible to define recursive functions of two or more variables, as the following example shows.

**Example 3.4.** We define a recursive function $f(n,k)$, for $n \geq k \geq 0$ and $n > 0$ as follows:

$$f(n,0) = f(n,n) = 1, \quad f(n,k) = f(n-1,k-1) + f(n-1,k) \text{ otherwise.}$$

Then it's easy to prove by induction on $n$ that $f(n,k) = \binom{n}{k}$. We check the first few values:

$$\begin{array}{l} f(1,0) = f(1,1) = 1 \\ f(2,0) = 1,\ f(2,1) = f(1,0) + f(1,1) = 2,\ f(2,2) = 1 \\ f(3,0) = 1,\ f(3,1) = f(2,0) + f(2,1) = 3,\ f(3,2) = 3, f(3,3) = 1 \\ f(4,0) = 1,\ f(4,1) = 4,\ f(4,2) = 6,\ f(4,3) = 4,\ f(4,4) = 1 \end{array}$$

## 4. Fibonacci and Catalan numbers

For more information on the Catalan numbers, see Anderson §12.2.

**Definition 4.1.** The Fibonacci numbers are a famous sequence of natural numbers which are often defined using a recursive formula:

$$f(1) = 1, \quad f(2) = 1, \quad f(n+2) = f(n+1) + f(n).$$

The first few terms are:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

**Example 4.2.** This is the original problem which led to the introduction of the Fibonnacci numbers:

**Question** A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?

**Answer** Let $R_n$ denote the number of rabbits at the beginning of month $n$. Thus when $n = 1$ we have one pair of (newborn) rabbits, so $R_1 = 1$. At the beginning of the second month we still have just one pair of rabbits, but will now become productive, so $R_2 = 1$ but $R_3 = 2$.

In general, if a pair of rabbits are born in month $n$, then it becomes productive in month $n+1$ and will produce a pair of rabbits starting in month $n+2$. That is $R_n = R_{n-1} + R_{n-2}$, and $R_1 = R_2 = 1$. So for all $n$ the numbers $R_n$ are equal to the Fibonacci numbers $f_n$, and for $n = 12$ we have $R_{12} = f_{12} = 144$.

**Example 4.3.** How many ways can we write a number $n$ as an **ordered** sum of 1s and 2s? (By an *ordered sum* we mean that the order of the terms in the sum is important)?

Let $a_n$ denote the number of ways we can write $n$ as an ordered sum of 1s and 2s. Then 1 can only be written in one way, so $a_1 = 1$. We can write $2 = 1 + 1 = 2$, so $a_2 = 2$.

To make an ordered sum of 1s and 2s equal to $n$, we can either take such a sum for $n-1$ and add a 1 on the end, or we can take such a sum for $n-2$ and add a 2 at the end. Every sum for $n$ has either a 1 or a 2 at the end, so arises exactly once in this way.

So $a_1 = 1$, $a_2 = 2$ and $a_n = a_{n-1} + a_{n-2}$, and hence $a_n = f_{n+1}$.

Later, we'll see how to find a non-recursive definition of the Fibonacci numbers.

**Definition 4.4.** We consider one more recursive sequence, called the *Catalan numbers*. These are defined via:

$$C_0 = 1, C_n = \sum_{i=1}^{n} C_{i-1}C_{n-i} = C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-1}C_0.$$

**Example 4.5.** Let's calculate the first few Catalan numbers:

$$\begin{array}{ll}
C_0 & = 1 \\
C_1 & = C_0C_0 = 1 \\
C_2 & = C_0C_1 + C_1C_0 = 2 \\
C_3 & = C_0C_2 + C_1C_1 + C_2C_0 = 5 \\
C_4 & = C_0C_3 + C_1C_2 + C_2C_1 + C_3C_0 = 14 \\
C_5 & = 14 + 5 + 4 + 5 + 14 = 42 \\
C_6 & = 42 + 14 + 10 + 10 + 14 + 42 = 132 \\
C_7 & = 429 \\
C_8 & = 1430 \\
C_9 & = 4862 \\
C_{10} & = 16796
\end{array}$$

We will get an explicit formula for $C_n$ in a somewhat roundabout way.

**Definition 4.6.** Call a word $w$ of length $2n$ in the alphabet $\{0, 1\}$ *balanced* or *good* if

1. $w$ has $n$ 0s and $n$ 1s.
2. Reading $w$ from left to right there are never more 1s than 0s.

For instance, 00101101 is good, whilst 00101110 is not good. If one takes a good word, and replaces all of the 0s with ( (left bracket) and all of the 1s with ) (right bracket), then one gets an arrangement of brackets that is correctly paired.

**Lemma 4.7.** *The number of balanced words of length* $2n$ *is the* $n$*th Catalan number* $C_n$.

**Proof.** Denote the number of balanced words of length $n$ by $b_n$.

We will prove the result by induction: clearly $b_0 = C_0 = 1$ and $b_1 = C_1 = 1$, which founds the induction.

Assume the result holds for all balanced words of length less than $2n$, and consider a balanced word $a_1a_2 \dots a_{2n}$. Certainly $a_1 = 0$, $a_{2n} = 1$. We let $a_{2i}$ be the first 1 which makes the number of 1s up to that point equal to the number of 0s (note that it must occur in an even position).

Then

$$a_1a_2 \dots a_{2n} = 0a_2 \dots a_{2i-1}1a_{2i+1} \dots a_n,$$

where $a_1 \dots a_{2i}$ is good and so is $a_{2i+1} \dots a_n$. Then $a_2 \dots a_{2i-1}$ is also good, and contains $2(i-1)$ letters.

Therefore

$$\begin{array}{lll}
b_n & = \sum_{i=1}^n b_{i-1}b_{n-i} & \\
& = \sum_{i=1}^n C_{i-1}C_{n-i} & \text{by the inductive hypothesis} \\
& = C_n & \text{by the definition of the Catalan numbers}
\end{array}$$

■

**Theorem 4.8.** *The number of balanced words of length* $2n$ *is*

$$\frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!},$$

*and hence*

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

**Proof.** Let us count *unbalanced* words consisting of $n$ 0s and $n$ 1s.

Consider such a word $a_1a_2\dots a_n$ and let $a_{2i+1}$ be the first 1 that makes it unbalanced (note that it must occur in an odd position):

$$a_1a_2\dots a_{2i}1a_{2i+2}\dots a_{2n}.$$

Then $a_1\dots a_{2i}$ is balanced, whilst in $a_{2i+2}\dots a_{2n}$ there is one more 0 than 1. In $a_{2i+2}\dots a_n$ change every 0 to 1 and vice versa. Then $a_1\dots a_{2n}$ now has $n-1$ 0s and $n+1$ 1s. Therefore we have constructed a map from the unbalanced words with $n$ 0s and $n$ 1s to the words with $n-1$ 0s and $n+1$ 1s.

Conversely, if we have a word $w_1w_2\dots w_{2n}$ with $(n-1)$ 0s and $(n+1)$ 1s, let $w_{2i+1}$ be the first position where there are more 1s than 0s, so that

$$w_1w_2\dots w_{2n} = w_1\dots w_{2i}1w_{2i+2}\dots w_{2n}.$$

Then $w_1\dots w_{2i}$ is balanced. Therefore, in $w_{2i+2}\dots 2_{2n}$ there is one more 1 than 0. Swap the 0s and 1s in $w_{2i+2}\dots 2_{2n}$, so that $w_1\dots w_{2n}$ now has equal numbers of 0s and 1s. However, the word is still not balanced, as there is one more 1 than 0 by position $2i+1$. Thus we have constructed a map from the words with $(n-1)$ 0s and $(n+1)$ 1s to the unbalanced words with $n$ 0s and $n$ 1s.

Notice that this second map is the *inverse* of the map in the previous paragraph: if we use the first map followed by the second, we get back the word that we started with, and similarly if we start with the second and then apply the first we also get back our original word. Therefore the two maps are *bijections*.

Therefore, the number of unbalanced words is equal to the number of words with $(n-1)$ 0s and $(n+1)$ 1s, which is just $\binom{2n}{n-1}$. The number of words with $n$ of each symbol is $\binom{2n}{n}$. The number of balanced words is the difference between these numbers:

$$\begin{aligned}\binom{2n}{n}-\binom{2n}{n-1} &= \frac{(2n)!}{n!n!}-\frac{(2n)!}{(n-1)!(n+1)!}\\ &= \frac{(2n)!}{n!n!}\left(1-\frac{n}{n+1}\right)\\ &= \frac{1}{n+1}\binom{2n}{n}.\end{aligned}$$

So $b_n = \frac{1}{n+1}\binom{2n}{n}$. ■

We verify this for one small value:

$$C_7 = \frac{1}{8}\binom{14}{7} = \frac{1}{8}\frac{14.13.12.11.10.9.8}{7.6.5.4.3.2.1} = 429.$$

## 5. Derangements

In this question we consider some mathematics inspired by the following problem.

**Example 5.1.** Suppose that $n$ people at a party leave their coats in a bedroom. At the end of the party, the light bulb has broken in the bedrom so they each take a coat at random. What is the probability that **no** person gets the right coat?

**Definition 5.2.** A *derangement* of $1,\dots,n$ is a permutation $\pi$ of $1,\dots,n$ such that $\pi(i)\neq i$ for each $i$. We write $d(n)$ for the number of derangements of $1,\dots,n$.

**Example 5.3.** 1. There are no derangements of 1 (just one point), as the only permutation of 1 point sends 1 to 1.

2. There is one derangement of $1,2$, namely $2,1$.
3. There are two derangements of $1,2,3$, namely $2,3,1$ and $3,1,2$.

4. There are nine derangements of $1, 2, 3, 4$:

$$\begin{gathered} 2413,\ 2143,\ 2341,\ 3142,\ 3421, \\ 3412,\ 4123,\ 4312,\ 4321. \end{gathered}$$

In each of them, 1 is not in the first place, 2 is not in the second, 3 is not in the third and 4 is not in the fourth. So $d(4) = 9$. Notice that in 2143, 3412 and 4321 the number 4 has swapped places with one of the other numbers, but in the remaining 6 derangements this hasn't happened.

We will find a recursive definition of the $d(n)$, and use it to find a closed formula.

**Theorem 5.4.** *For $n \geq 3$,*

$$d(n) = (n-1)(d(n-1) + d(n-2)).$$

**Proof.** Write $e(n)$ for the number of derangements where $n$ swaps positions with some other number, and write $f(n)$ for the number of derangments where this does not happen. Thus $d(n) = e(n) + f(n)$.

Suppose first that a derangement $\pi$ sends $n$ to $i$ and $i$ to $n$, for some $i$. We count the number of such $\pi$. There are $n-1$ choices for $i$, and $1, \ldots, i-1, i+1, \ldots, n-1$ must be deranged amongst themselves. This can be done in $d(n-2)$ ways, so

$$e(n) = (n-1)d(n-2).$$

Suppose next that $\pi$ is a derangement such that $n$ does not swap places with any other number. We count the number of such $\pi$. Then some $r \in 1, \ldots, n-1$ satisfies $\pi(r) = n$, but $\pi(n) \neq r$. Let $D = \{1, \ldots, r-1, r+1, \ldots, n\}$ and $I = \{1, 2, \ldots, n-1\}$. Then $\pi$ maps $D$ to $I$, and each element of $D$ has exactly one element of $I$ which is forbidden as an image ($\pi(n) \neq r$ and $\pi(i) \neq i$, otherwise). Thus there are $d(n-1)$ ways of completing the derangement $\pi$, and so $f(n) = (n-1)d(n-1)$.

Summing, we get

$$d(n) = e(n) + f(n) = (n-1)(d(n-1) + d(n-2)).$$

■

We can now use this to get a closed formula for $d(n)$, which answers our question at the beginning of this section.

**Theorem 5.5.** *For $n \geq 1$,*

$$d(n) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

**Proof.** We can rewrite Theorem 5.4 as

$$d(k) - kd(k-1) = -(d(k-1) - (k-1)d(k-2)).$$

Notice that the right hand side is the negative of the left hand side, but with $k$ replaced by $k-1$. Iterating, we therefore get

$$\begin{aligned} d(k) - kd(k-1) &= -(d(k-1) - (k-1)d(k-2)) \\ &= (-1)^2(d(k-2) - (k-2)d(k-3)) \\ &= (-1)^{k-2}(d(2) - 2d(1)) \\ &= (-1)^k(1-0) \\ &= (-1)^k. \end{aligned}$$

Hence $d(k) - kd(k-1) = (-1)^k$.

Next, we divide the previous equation by $k!$ to get

$$\frac{d(k)}{k!} - \frac{d(k-1)}{(k-1)!} = \frac{(-1)^k}{k!}.$$

Now let's sum the resulting expressions from $k = 2$ to $k = n$:

$$\begin{aligned}(\tfrac{d(2)}{2!} - \tfrac{d(1)}{1}) + (\tfrac{d(3)}{3!} - \tfrac{d(2)}{2!}) + (\tfrac{d(4)}{4!} - \tfrac{d(3)}{3!}) + \cdots + (\tfrac{d(n)}{n!} - \tfrac{d(n-1)}{(n-1)!}) \\ = \tfrac{(-1)^2}{2!} + \tfrac{(-1)^3}{3!} + \cdots + \tfrac{(-1)^n}{n!} \\ \tfrac{d(n)}{n!} - \tfrac{d(1)}{1} = \textstyle\sum_{k=2}^{n} \tfrac{(-1)^k}{k!}.\end{aligned}$$

Now, $d(1)$ is the number of derangments of 1 element, which is 0. Also,

$$\frac{(-1)^0}{0!} + \frac{(-1)^1}{1!} = 1 - 1 = 0,$$

so we can rewrite to get

$$\frac{d(n)}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

■

Finally, we note a surprising and pretty consequence of Theorem 5.5

**Corollary 5.6.** *As $n$ tends to infinity, the probability that a random permutation is a derangment tends to $1/e = 0.368$ (to 3 dp).*

**Proof.** By Chapter 1, Corollary 7.6, the number of permutations of $1, \ldots, n$ is $n!$. So the probability that a random permutation is a derangement is

$$\frac{1}{n!} d(n) = \sum_{m=0}^{n} \frac{(-1)^m}{m!}.$$

Now,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

so

$$1/e = e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots = \lim_{n\to\infty} \frac{d(n)}{n!}.$$

■

## 6. Generating functions

[See Anderson, §§13.1 and 13.2 for lots more information about generating functions].

Suppose we wish to describe a sequence $a_0, a_1, \ldots$. If the sequence is finite, we can just write down all of the values. If the sequence is infinite, we've already seen two ways that we might define it:

- We might be able to write down an explicit formula in terms of $n$ for $a_n$, for all $n$.
- We might be able to give a recursive formula for $a_n$ in terms of $a_0, \ldots, a_{n-1}$.

In this section we'll see a third way of describing (finite or infinite sequences).

**Definition 6.1.** The *generating function* of a sequence $a_0, a_1, a_2, \ldots$ is the function

$$f(x) = \sum_{i=0}^{\infty} a_i x^i.$$

**Example 6.2.** The generating function of the Fibonacci sequence is

$$x + x^2 + 2x^3 + 3x^4 + 5x^5 + \cdots.$$

Usually, when we work with generating functions, we want to express them with a finite formula rather than an infinite sum. The following theorem often gives an extremely convenient way of doing this.

**Theorem 6.3.** *The generating function for the sequence*

$$1, 1, 1, 1, \ldots$$

*is*

$$1 + x + x^2 + x^3 + \cdots = \sum_{r=0}^{\infty} x^r = \frac{1}{1-x}.$$

**Proof.** There are lots of ways to prove this. Possibly the easiest is to notice that if $|x| < 1$ then the left hand side is a geometric series, with first term 1 and common ratio $x$. The sum of the series is therefore $1/(1-x)$. ■

**Example 6.4.** Let $c \in \mathbb{Z}$ be a non-zero integer. By substitutin $cx$ for $x$ in Theorem 6.3, we see that the generating function of the sequence $a_n = c^n$ $(n \geq 0)$ is

$$1 + cx + c^2x^2 + \cdots = \frac{1}{1-cx}.$$

**Theorem 6.5.** *Let $g(x)$ be the generating function for the Fibonacci numbers $f_1, f_2, f_3, \ldots$. Then*

$$g(x) = \frac{x}{1-x-x^2}.$$

**Proof.** We may write

$$\begin{aligned}
g(x) &= f_1x + f_2x^2 + f_3x^3 + \ldots \\
&= x + x^2 + (f_1 + f_2)x^3 + (f_2 + f_3)x^4 + \ldots \\
&= x + x^2 + (f_1x^3 + f_2x^4 + f_3x^5 + \ldots) + (f_2x^3 + f_3x^4 + \ldots \\
&= x + x^2(f_1x + f_2x^3 + \ldots) + x(x + f_2x^2 + f_3x^3 + \ldots) \\
&= x + x^2g(x) + xg(x) \\
(1 - x - x^2)g(x) &= x \\
g(x) &= \tfrac{x}{1-x-x^2},
\end{aligned}$$

as required. ■

**Example 6.6.** Consider the recursively defined sequence $a_0 = 0$, $a_1 = 0$, and $a_n = 3 \cdot 2^{n-1} - a_{n-1}$ for $n \geq 2$. The first few terms are

$$a_0 = a_1 = 0,\ a_2 = 3 \cdot 2^1 - 0 = 6,\ a_3 = 3 \cdot 2^2 - 6 = 6,\ a_4 = 3 \cdot 2^3 - 6 = 18$$

Let's calculate a generating function for this sequence, using the same method as in the proof of Theorem 6.5.

By definition, the generating function is $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$. Comparing with the recurrence relation we get

$$\begin{aligned}
f(x) &= a_0 + a_1 x + (3 \cdot 2 - a_1)x^2 + (3 \cdot 2^2 - a_2)x^3 + \cdots \\
&= 0 + 0x + 3(2x^2 + 2^2x^3 + 2^3x^4 + \cdots) - (a_1x^2 + a_2x^3 + a_3x^4 + \cdots) \\
&= 3x(2x + 2^2x^2 + 2^3x^3 + \cdots) - x(a_1x + a_2x^2 + a_3x^3 + \cdots) \\
&= 6x^2(1 + 2x + 2^2x^2 + 2^3x^3 + \cdots) - x(a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots).
\end{aligned}$$

Now, we saw in Example 6.4 that $1 + 2x + 2^2x^2 + \cdots = \frac{1}{1-2x}$, so we can rewrite the previous line as

$$\begin{aligned}
f(x) &= 6x^2 \cdot \tfrac{1}{1-2x} - xf(x) \\
(1+x)f(x) &= \tfrac{6x^2}{1-2x} \\
f(x) &= \tfrac{6x^2}{(1+x)(1-2x)}.
\end{aligned}$$

Let's now see how to get a closed formula for $a_n$, by working out an explicit sum for $f(x)$. Using partial fractions, we write

$$\begin{aligned}
\tfrac{1}{(1+x)(1-2x)} &= \tfrac{A}{(1+x)} + \tfrac{B}{(1-2x)} \\
1 &= A(1-2x) + B(1+x)
\end{aligned}$$

Collecting constant terms we get $1 = A + B$, then collecting $x$ coefficients we get $0 = -2A + B$, so $B = 2A$, and then $A = 1/3$, $B = 2/3$. Thus

$$\frac{1}{(1+x)(1-2x)} = \frac{1/3}{1+x} + \frac{2/3}{1-2x},$$

so

$$f(x) = 2x^2\left(\frac{1}{1+x} + \frac{2}{1-2x}\right).$$

Using the series expansion for $1/(1-x)$ again, we can write this as

$$f(x) = 2x^2\left((1 - x + x^2 - \cdots) + 4x^2(1 + 2x + 2^2x^2 + \cdots)\right).$$

Now, remember that the coefficient of $x^n$ is $a_n$. We can now just read off the coefficient of $x^n$ to get

$$a_n = 2 \cdot (-1)^{n-2} + 4 \cdot 2^{n-2} = 2 \cdot (-1)^n + 2^n,$$

which is a closed formula for $a_n$, valid for $n \geq 2$.

Let's check a few values. For $n = 2$ we have $2 \cdot (-1)^2 + 2^2 = 2 + 4 = 6$, for $n = 3$ we have $2 \cdot (-1)^3 + 2^3 = -2 + 8 = 6$, and for $n4$ we have $2 \cdot (-1)^4 + 2^4 = 2 + 16 = 18$.

**Theorem 6.7.** *A closed formula for the Fibonacci numbers $f_1, f_2 \ldots$ is:*

$$f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

**Proof.** We use partial fractions to find a closed formula for $f_n$, given the generating function $g(x)$ from Theorem 6.5.

First, one may verify that that $1 - x - x^2 = (1-\alpha x)(1-\beta x)$, where $\alpha = (1+\sqrt{5})/2$ and $\beta = (1-\sqrt{5})/2$. Then we write

$$\begin{aligned}
g(x) &= \tfrac{x}{(1-\alpha x)(1-\beta x)} \\
\tfrac{x}{(1-\alpha x)(1-\beta x)} &= \tfrac{A}{1-\alpha x} + \tfrac{B}{1-\beta x} \\
x &= A(1-\beta x) + B(1-\alpha x)
\end{aligned}$$

From this we deduce that $A = 1/\sqrt{5}$ and $B = -A = -1/\sqrt{5}$. So

$$g(x) = \frac{1}{\sqrt{5}}\left(\frac{1}{1-\alpha x} - \frac{1}{1-\beta x}\right).$$

Now let's use the series expansion for $1/(1-x)$ again, to write this as

$$\begin{aligned} g(x) &= \tfrac{1}{\sqrt{5}}((1+\alpha x+\alpha^2x^2+\ldots) - (1+\beta x+\beta^2x^2+\cdots)) \\ &= \tfrac{1}{\sqrt{5}}((\alpha-\beta)x + (\alpha^2-\beta^2)x^2+\cdots). \end{aligned}$$

The result follows by considering the coefficient of $x^n$. ■