Combinatorics 1:
The art of counting
Preface

This is the first of a three-part set of lecture notes on Advanced Combinatorics, for the module MT5821 of that title at the University of St Andrews.

Roughly speaking, Combinatorics is the study of arranging objects according to certain rules. The natural questions are: is the arrangement possible? If so, in how many different ways can it be made? What extra properties such as symmetry do these arrangements have? This part of the notes is mainly concerned with the second question: for basic objects such as subsets of a set, there is no doubt about their existence, and we want to know how many there are with various properties. We will touch briefly on the first question in a section on combinatorics of subsets, treating Ramsey’s theorem and Steiner systems.

The most powerful tool in enumerative combinatorics is the use of formal power series, and we spend some time on these objects and their properties.

The syllabus for the module describes the three options as follows:

1. *Enumerative combinatorics*: basic counting, formal power series and their calculus, recurrence relations, *q*-analogues, group action and cycle index, species, asymptotic results.

2. *Graphs, codes and designs*: strongly regular graphs, *t*-designs, optimality for block designs, codes and weight enumerators, matroids and Tutte polynomial, MacWilliams relations.

3. *Projective and polar spaces*: geometry of vector spaces, combinatorics of projective planes, sesquilinear and quadratic forms and their classification, diagram geometry, classical groups.

These notes refer to the first section, delivered for the first time in the second semester of 2013–2014.

Many images in the notes are taken from that great St Andrews resource, the MacTutor History of Mathematics website.

Counting combinatorial objects can mean various different things:

- Best of all is an exact formula, such as the formula $2^n$ for the number of subsets of a set of size $n$. This formula is also easy to evaluate for given $n$, and tells us how fast the number in question grows as a function of $n$. 

• Some formulae are more complicated, such as the following formula for the number of partitions of an $n$-set with $k$ parts:

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n.$$  

• Even when there is a simple formula, it may be difficult to estimate its magnitude. The number of permutations of an $n$-set is $n!$, and Stirling’s formula gives us the asymptotic behaviour:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$  

• Sometimes even an asymptotic formula is too much to ask, and we have to be content with some kind of approximation.

• Maybe we can’t give an exact or asymptotic formula, but we have a recurrence relation. The number $B(n)$ of partitions of an $n$-set satisfies

$$B(n) = \sum_{i=1}^{n} \binom{n-1}{i-1} B(n-i).$$  

• Sometimes we want to generate all the objects counted by a combinatorial formula. In this case we need either a method of stepping from any object to the next (as the odometer or mileage gauge in a car does, with strings of digits), or a method of computing the $i$th object in some ordering directly from $i$.

The topics covered in the course can be inferred from the contents page.

Peter J. Cameron
April 2014
Contents

1 Counting subsets 6
1.1 Binomial coefficients ............................................. 6
1.2 The Binomial theorem ............................................. 7
1.3 Binomial coefficient identities ................................... 9
1.4 Generating functions .............................................. 11
1.5 Unimodality and estimates ....................................... 14
1.6 Matrices ........................................................... 15

2 Combinatorics of subsets 18
2.1 Ramsey’s theorem .................................................. 18
2.2 Steiner systems ..................................................... 25
2.3 Appendix: The infinite Ramsey theorem ......................... 27

3 Formal power series 32
3.1 Commutative rings with identity .................................. 32
3.2 Formal power series ................................................. 32
3.3 Operations on formal power series ............................... 34
3.4 Elementary properties .............................................. 37
3.5 Connection with analysis .......................................... 41
3.6 Further examples .................................................. 43
3.7 Appendix: selection with repetition .............................. 47

4 Linear recurrences with constant coefficients 50
4.1 The problem and its solution ..................................... 50
4.2 Regular languages .................................................. 52
4.3 C-finite sequences ................................................. 55

5 Catalan objects 60
5.1 Binary trees ........................................................ 60
5.2 Rooted plane trees ................................................. 61
5.3 Dissections of polygons .......................................... 61
5.4 Dyck paths ........................................................ 61
5.5 Ballot numbers ..................................................... 62
5.6 Young diagrams and tableaux ..................................... 63
5.7 Permutation patterns .............................................. 63
5.8 Wedderburn–Etherington numbers ............................... 64
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Gaussian coefficients</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>6.1 The definition</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>6.2 Vector spaces over finite fields</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>6.3 Relations between Gaussian coefficients</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>6.4 Lattice paths</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>Möbius inversion</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>7.1 The Principle of Inclusion and Exclusion</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>7.2 Applications</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>7.3 The Möbius function of a poset</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>7.4 Some examples</td>
<td>82</td>
</tr>
<tr>
<td>8</td>
<td>Number partitions</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>8.1 Definition and generating function</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>8.2 Euler’s Pentagonal Numbers Theorem</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>8.3 Jacobi’s Triple Product Identity</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>Set partitions and permutations</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>9.1 Partitions</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>9.2 Permutations</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>9.3 The Möbius function of the partition lattice</td>
<td>104</td>
</tr>
<tr>
<td>10</td>
<td>Stirling’s formula</td>
<td>107</td>
</tr>
<tr>
<td>11</td>
<td>Solutions to the exercises</td>
<td>111</td>
</tr>
<tr>
<td>12</td>
<td>Revision problems</td>
<td>133</td>
</tr>
</tbody>
</table>
1 Counting subsets

In this section, we count the subsets of an \( n \)-element set. The counting numbers are the binomial coefficients, familiar objects but there are some new things to say about them. These will lead on to topics later in the notes.

We begin with a simple observation.

**Proposition 1.1** The number of subsets of an \( n \)-element set is \( 2^n \).

**Proof** A subset is specified by saying which elements of the set it contains. For each element, we have the choice of including it or leaving it out. The result of \( n \) binary choices is that there are \( 2^n \) possible subsets.

Alternatively, we can match up the subsets of \( \{1, \ldots, n\} \) with the \( n \)-tuples of zeros and ones (of which there are \( 2^n \)): the subset \( A \) is matched to the \( n \)-tuple \( e \), where

\[
e_i = \begin{cases} 1 & \text{if } i \in A, \\ 0 & \text{if } i \notin A. \end{cases}
\]

We will sometimes abbreviate “\( n \)-element set” to “\( n \)-set”.

### 1.1 Binomial coefficients

**Definition** The binomial coefficient \( \binom{n}{k} \) is the number of \( k \)-subsets of an \( n \)-set.

**Proposition 1.2**

\[
\binom{n}{k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k(k-1)(k-2) \cdots 1},
\]

the product of \( k \) descending integers starting at \( n \) divided by a similar product starting at \( k \).

**Proof** We choose a \( k \)-subset of the \( n \)-set by picking its elements one at a time: there are \( n \) choices for the first, \( n-1 \) for the second, \ldots, and \( n-k+1 \) for the \( k \)th.

However, we have over-counted, since the same elements in a different order make up the same subset. The number of orders in which \( k \) elements can be chosen is \( k(k-1) \cdots 1 \), and we have to divide by this to get the answer.

Immediately from the definition and the result of the preceding section we have

**Proposition 1.3** \( \sum_{k=0}^{n} \binom{n}{k} = 2^n \).
**Definition**  The *factorial* function is defined by

\[ n! = n(n-1)(n-2) \cdots 1, \]

the product of the integers from 1 to \( n \). Using this, we can write the binomial coefficient more succinctly as

\[ \binom{n}{k} = \frac{n!}{k!(n-k)!}, \]

since the \((n-k)!\) in the denominator cancels the factors in \( n! \) from \( n-k \) down to \( 1 \), leaving only those in the numerator in the proposition.

However, this nice compact formula is not always the best in practice. Consider evaluating \( \binom{200}{2} \). The formula in the Proposition gives

\[ \binom{200}{2} = \frac{200 \cdot 199}{2 \cdot 1} = 19900, \]

while the one using factorials gives

\[ \binom{200}{2} = \frac{200!}{2!198!} = \frac{78865786736479050355236321393218506229513597768717326329474253324435944996340334292030428401198462390417721213891963883025}{2 \cdot 19815524305648002601818171204326257846611456725808373449616646563} \]

Then what do you do?

You probably think of the expression \( \binom{n}{k} \) as making sense when \( n \) and \( k \) are positive integers and \( k < n \). However, the formula in the Proposition works fine in other cases too:

- \( \binom{n}{0} = 1 \) since an \( n \)-set has a unique subset with no elements (the empty set). Similarly \( \binom{n}{n} = 1 \).
- If \( k > n \), then the formula gives 0, which is correct since there are no \( k \)-subsets in this case.

### 1.2 The Binomial theorem

The reason for the name “binomial coefficients” is that these numbers occur as coefficients in the *binomial theorem*:
Theorem 1.4 (Binomial Theorem) For any non-negative integer $n$,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k.$$

How is such a theorem proved? We have two ways of thinking about the coefficients (as counting something, or as a formula), so there are potentially two ways to prove the theorem. It is instructive to compare them.

First proof The formula suggests a proof by induction. It is clearly true when $n = 0$. So let us assume it for a value $n$ and prove it for $n + 1$. We have

$$(x + y)^{n+1} = (x + y)(x + y)^n = (x + y) \left( \sum_{k=0}^{n} x^{n-k} y^k \right).$$

On the right, expanding the brackets will give us a sum of terms, each of which will have the sum of the exponents of $x$ and $y$ equal to $n + 1$. Consider the term $x^{n+1-k} y^k$ with $1 \leq k \leq n$. This will be the sum of two contributions,

$$x \cdot \binom{n}{k} x^{n-k} y^k \text{ and } y \cdot \binom{n}{k-1} x^{n-k+1} y^{k-1}.$$ 

So we need:

Lemma 1.5 \( \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \).

This can be proved using the formula for the binomial coefficient.

The terms for $k = 0$ and $k = n + 1$ each only come from one place in the expression, and the coefficient of each of them is 1, as required. So the induction step is complete.

Second proof Consider

$$(x + y)^n = (x + y) \cdots (x + y) \quad (n \text{ factors}).$$

Expanding all the brackets, we get a sum of terms, each a product of a power of $x$ and a power of $y$. The term $x^{n-k} y^k$ comes by selecting $y$ from $k$ of the $n$ brackets
and $x$ from the remaining ones. There are \( \binom{n}{k} \) ways to choose the $k$ brackets from which to select $y$. The result follows.

In this case the counting proof is much simpler! Very often we will have two proofs of a result, one by counting and one analytic; judge for yourself which you prefer.

### 1.3 Binomial coefficient identities

There is a huge industry involving proving identities connecting binomial coefficients. Here are a few. I have given hints for counting proofs in some cases. You should write out detailed proofs of some of these, and compare them to proofs using the formulae.

**Proposition 1.6** \( \binom{n}{k} = \binom{n}{n-k} \).

This says that you can match up the $k$-subsets of an $n$-set with the $(n-k)$-subsets; match each set with its complement.

**Proposition 1.7** \( \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \).

We saw this earlier, as a lemma in the proof of the Binomial Theorem. Suppose you have to pick a team of $k$ from a class of $n+1$ pupils. You can do this in \( \binom{n+1}{k} \) ways. Alternatively, focus attention on one of the children, say A. There are \( \binom{n}{k-1} \) teams including A, since the remainder of the team must be picked from the remaining $n$ pupils. Similarly there are \( \binom{n}{k} \) teams not including A. The result follows.

This is the basic property which is used to construct *Pascal’s Triangle*, a triangular array in which the $k$th entry in the $n$th row is \( \binom{n}{k} \). (Both row numbers, and entry numbers in each row, start from 0.) The rule is: each entry of the table is the
sum of the two entries in the row above, to the left and the right of the position we are looking at.

\[
\begin{array}{cccccc}
1 & & & & & \\
1 & 2 & 1 & & & \\
1 & 3 & 3 & 1 & & \\
1 & 4 & 6 & 4 & 1 & \\
1 & 5 & 10 & 10 & 5 & 1
\end{array}
\]

Pascal’s triangle was not invented by Pascal; it was known in China some centuries earlier, and possibly brought to Europe by missionaries. Figure 1 shows Chu Shi-Chieh’s version, from his book *Ssu Yuan Yu Chien*, dated 1303.

Figure 1: Chu Shi-Chieh’s Triangle

**Proposition 1.8** \( k \binom{n}{k} = n \binom{n-1}{k-1} \).

Consider choosing a team of \( k \) from a class of \( n \) pupils, and then choosing a captain for the team. We could alternatively choose the captain first, and then pick the remaining \( k - 1 \) members of the team.
Proposition 1.9 $\sum \binom{m}{i} \binom{n}{k-i} = \binom{m+n}{k}$, where the sum is over all values of \( i \) for which the binomial coefficients make sense (that is, \( 0 \leq i \leq m \) and \( 0 \leq k-i \leq n \)).

Pick a team of \( k \) players from a class of \( m \) girls and \( n \) boys.
This result is known as the Vandermonde convolution.

1.4 Generating functions

Our formulation of the Binomial Theorem is slightly wasteful: since the exponents of \( x \) and \( y \) sum to \( n \), one can be deduced from the other. It will be more useful to state the theorem in this form:

Theorem 1.10 (Binomial Theorem) For a non-negative integer \( n \),

\[
(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.
\]

The expression on the right is what is known as the generating function for the binomial coefficients. In general, if we have a sequence of numbers \( a_0, a_1, a_2, \ldots \), the generating function for the sequence is \( \sum_{n\geq0} a_n x^n \), so that the coefficient of the \( n \)th power of the indeterminate is the \( n \)th number in the sequence. The sum is over all relevant values. In the case of the binomial theorem, it doesn’t matter if we don’t stop in time, since the binomial coefficients vanish for \( k > n \); so even if we think of the sum as over all natural numbers \( k \), there are only finitely many non-zero terms.

This form is already useful. Here is an example.

A special case of the Vandermonde convolution (taking \( m = n = k \) and using Proposition 1.6) is

\[
\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.
\]

If we put alternating signs in the sum, something different happens:

Proposition 1.11 $\sum_{k=0}^{n} (-1)^k \binom{n}{k}^2 = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ (-1)^{n/2} \binom{n}{n/2} & \text{if } n \text{ is even.} \end{cases}$
Proof  We start with the identity
\[
(1 - x)^n(1 + x)^n = (1 - x^2)^n.
\]
Now we calculate the coefficient of \(x^n\) on both sides. On the left, taking the coefficient of \(x^k\) in the first factor and \(x^{n-k}\) in the second, we obtain
\[
\sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{n}{n-k} = \sum_{k=0}^{n} (-1)^k \binom{n}{k}^2.
\]
On the right, we have only even powers of \(x\), so the coefficient is zero if \(n\) is odd. If \(n\) is even, the required coefficient is obtained from the binomial expansion of \((1 - x^2)^n\).

The binomial coefficient \(\binom{n}{k}\) has two parameters. So one could ask about the generating function for the parameter \(n\), namely
\[
\sum_{n \geq 0} \binom{n}{k} y^n
\]
for fixed \(k\), or even the bivariate generating function, where we have one indeterminate for each parameter. Can we calculate these?

We have
\[
\sum_{n \geq 0} \sum_{k \geq 0} \binom{n}{k} x^k y^n = \sum_{n \geq 0} (1 + x)^n y^n = \frac{1}{1 - (1 + x)y}.
\]
Putting \(x = 1\) says that we don’t care about the value of \(k\), so we get the generating function for the total number of subsets of an \(n\)-set; what we find is
\[
\frac{1}{1 - 2y} = \sum_{n \geq 0} 2^n y^n,
\]
in agreement with Proposition 1.1.

Continuing the calculation above,
\[
\sum_{n \geq 0} \sum_{k \geq 0} \binom{n}{k} x^k y^n = \frac{1}{1 - y} \cdot \frac{1}{1 - (y/(1 - y)x)} = \sum_{k \geq 0} y^k \frac{1}{(1 - y)^{k+1}} x^k.
\]
Now we can equate the coefficients of $x^k$ on the two sides to deduce

$$\sum_{n \geq 0} \binom{n}{k} y^n = \frac{y^k}{(1-y)^{k+1}}.$$  

This is the required generating function.

Note that the power series for the right-hand side begins with the term $y^k$; this is because the binomial coefficients vanish for $n < k$.

We can deduce from this a form of the Binomial Theorem for negative exponent. Divide by $y^k$ and put $m = k+1$, $x = -y$, and $l = n - m + 1 = n - k$:

$$(1+x)^{-m} = \sum \binom{n}{n-m+1} (-x)^{n-m+1}$$

$$= \sum \frac{n(n-1)\cdots m}{(n-m+1)\cdots 1} (-1)^{n-m+1} x^{n-m+1}$$

$$= \sum \frac{(-m)(-m-1)\cdots (-m-l+1)}{l!} x^l$$

$$= \sum \binom{-m}{l} x^l.$$  

Here we have defined the binomial coefficient $\binom{-m}{l}$ by the same rule as for the usual binomial coefficients. Indeed, we can go further:

**Definition**  Let $k$ be a non-negative integer and $a$ any real number. Then let

$$\binom{a}{k} = \frac{a(a-1)\cdots (a-k+1)}{k!}.$$  

**Theorem 1.12 (General binomial theorem)**  $(1+x)^a = \sum_{k \geq 0} \binom{a}{k} x^k$.

This is, at present, a theorem of analysis, since we need analysis to define the function $(1+x)^a$ for real numbers $a$ and to calculate its Taylor series. However, we will come back to this result later.
1.5 Unimodality and estimates

The binomial coefficients form a *unimodal* sequence:

**Proposition 1.13** For fixed $n$, the binomial coefficients \( \binom{n}{k} \) increase with $k$ for $k < n/2$, and decrease as $k$ increases for $k > n/2$. If $n$ is even, then the central binomial coefficient \( \binom{n}{n/2} \) is the largest; if $n$ is odd, then the two on either side of the centre (that is, \( \binom{n}{(n-1)/2} \) and \( \binom{n}{(n+1)/2} \)) are equal and are larger than all the others.

**Proof** From the formula, we see that

\[
\binom{n}{k+1} = \frac{n-k}{k+1} \binom{n}{k}.
\]

So \( \binom{n}{k+1} \) is greater than, equal to, or less than \( \binom{n}{k} \) according as \( n-k \) is greater than, equal to, or less than \( k+1 \); that is, according as \( k \) is less than, equal to, or greater than \( (n-1)/2 \). The result follows from this.

We will meet other sequences of combinatorial numbers which also have the unimodal property. In cases where we don’t have a formula for the numbers in question, we cannot use such a simple argument. Instead, we will develop a test for unimodality, based on properties of the generating function.

As we have seen, the $n+1$ binomial coefficients sum to $2^n$. So we have:

**Proposition 1.14** If $n$ is even, then

\[
\frac{2^n}{n+1} \leq \binom{n}{n/2} \leq 2^n.
\]

A similar result holds for odd $n$. In fact, the precise asymptotics of the central binomial coefficient can be calculated using *Stirling’s formula*, which we will meet later. Its value is “close to” \( c \cdot 2^n / \sqrt{n} \) for a suitable constant $c$. 

14
1.6 Matrices

By left-aligning the entries of Pascal’s Triangle, we can convert it into an infinite lower-triangular matrix $B$, with $(i, j)$ entry $\binom{i}{j}$. (Note that this correctly gives $B_{ij} = 0$ for $i < j$.)

Multiplication of lower-triangular matrices is possible; no infinite sums are involved. For if $A$ and $B$ are lower-triangular, then the $(i, j)$ entry of $AB$ is $\sum A_{ik}B_{kj}$, and the terms in the sum are non-zero only if $k \leq j \leq i$. This also shows that the product $AB$ is lower-triangular. This means that, in practice, if we multiply the top left-hand corners of lower-triangular matrices, we get the top left-hand corner of the product.

For example,

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 2 & 1 & 0 \\
1 & 3 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 \\
1 & -2 & 1 & 0 \\
-1 & 3 & -3 & 1
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}.
\]

This is a special case of something much more general:

**Theorem 1.15** Let $B$ be the matrix of binomial coefficients defined above, and let $B^*$ be the matrix with $(i, j)$ entry $(-1)^{i-j}\binom{i}{j}$. Then $B^*$ is the inverse of $B$: that is, $BB^* = B^*B = I$.

**Proof** The set of all real polynomials is a real vector space, with basis $\{1, x, x^2, x^3, \ldots\}$. Another basis is $\{1, x+1, (x+1)^2, (x+1)^3, \ldots\}$. Now the Binomial Theorem shows that $B$ is the transition matrix expressing the vectors of the second basis in terms of the first. So its inverse is the transition matrix in the reverse direction. If $y = x + 1$, then $x = y - 1$, and so the coefficients of the inverse are found from the expansions of powers of $y - 1$ in terms of $y$, giving the result.

In fact this can be extended. For any real number $c$, let $B(c)$ be the matrix with $(i, j)$ entry $e^{c(i-j)}\binom{i}{j}$. Then $B(1) = B$ and $B(-1) = B^*$. More generally, $B(c)$ is the transition matrix from the basis $\{1, x, x^2, x^3, \ldots\}$ to the basis $\{1, x + c, (x + c)^2, (x + c)^3, \ldots\}$. So arguing as above we find that $B(c_1)B(c_2) = B(c_1 + c_2)$. 

15
In other words, the map \( c \mapsto B(c) \) is a homomorphism from the additive group of real numbers to the multiplicative group of upper triangular matrices with diagonal entries 1.

**Exercises**

1.1.

(a) In Vancouver in 1984, I saw a Dutch pancake house advertised “a thousand and one combinations” of toppings. What do you deduce?

(b) More recently McDonald’s offered a meal deal with a choice from eight components of your meal, and advertised “40,312 combinations”. What do you deduce?

1.2. Prove the Vandermonde convolution using generating functions.

1.3. Show that the number of subsets of \( \{1, \ldots, n\} \) of even cardinality is \( 2^{n-1} \). Calculate the number of subsets of cardinality divisible by 4. Your answer should depend on the congruence of \( n \) mod 8. (*Hint:* calculate \( (1+i)^n \)).

1.4. Write down Pascal’s triangle mod 2; that is, record only whether each entry is odd or even. You should find that the triangle has a fractal structure; can you explain why?

1.5. *Lucas’ Theorem* gives an expression for the congruence of binomial coefficients modulo a prime. Let \( p \) be prime, and let \( a = a_r p^r + a_{r-1} p^{r-1} + \cdots + a_1 p + a_0 \) and \( b = b_r p^r + b_{r-1} p^{r-1} + \cdots + b_1 p + b_0 \) be the expressions for \( a \) and \( b \) in base \( p \), so that \( 0 \leq a_i, b_i \leq p - 1 \) for all \( i \). Prove that

\[
\binom{a}{b} \equiv \binom{a_r}{b_r} \binom{a_{r-1}}{b_{r-1}} \cdots \binom{a_1}{b_1} \binom{a_0}{b_0} \pmod{p}.
\]

*Hint:* Show first that, if \( a = ps + c \) and \( b = pt + d \) with \( 0 \leq c, d \leq p - 1 \), then

\[
\binom{a}{b} \equiv \binom{s}{t} \binom{c}{d} \pmod{p}.
\]

*Remark:* This exercise may help you with the preceding one.
1.6. The multinomial coefficient \( \binom{n}{k_1, \ldots, k_r} \), where \( k_1, \ldots, k_r \) are non-negative integers with sum \( n \), is defined to be
\[
\frac{n!}{k_1! \cdots k_r!}.
\]
(So, in this notation, \( \binom{n}{k} = \binom{n}{k, n-k} \).)

Prove the Multinomial Theorem:
\[
(x_1 + x_2 + \cdots + x_r)^n = \sum_{k_1+k_2+\cdots+k_r=n} \binom{n}{k_1, k_2, \ldots, k_r} x_1^{k_1} x_2^{k_2} \cdots x_r^{k_r}.
\]

1.7. Construct a bijection between the set of all \( k \)-element subsets of \( \{1, 2, \ldots, n\} \) containing no two consecutive elements, and the set of all \( k \)-element subsets of \( \{1, 2, \ldots, n-k+1\} \). Hence show that the number of such subsets is \( \binom{n-k+1}{k} \).

In the UK National Lottery, six numbers are chosen randomly from the set \( \{1, \ldots, 49\} \). What is the probability that the selection contains no two consecutive numbers?

1.8. Prove that the generating function for the central binomial coefficients is
\[
\sum_{n \geq 0} \binom{2n}{n} x^n = (1 - 4x)^{-1/2},
\]
and deduce that
\[
\sum_{k=1}^{n} \binom{2k}{k} \binom{2(n-k)}{n-k} = 4^n.
\]
[Note: Finding a counting proof of this identity is quite challenging!]
2 Combinatorics of subsets

Many combinatorial objects can be expressed in terms of the subsets of a set. For example, a graph consists of a set $V$ of vertices and a subset $E$ of the set of all 2-subsets of $V$; the elements of $E$ are called edges. Typically we look for graphs with certain nice properties, or investigate properties of arbitrary graphs.

Graph theory would take us too far afield, but in this chapter I will discuss two special topics: Ramsey’s theorem, and Steiner systems.

2.1 Ramsey’s theorem

The party problem

The party problem states:

Six people are at a party. Show that either there are three of them, any two of whom know each other, or there are three people, no two of whom know each other.

Using the mathematician’s prerogative to use words with any defined meaning, we will abbreviate “any two know each other” to “mutual friends”, and “no two know each other” to “mutual strangers”. We assume also that friendship is an irreflexive symmetric relation: that is,

- nobody is his/her own friend; and
- if A is B’s friend, then B is A’s friend.

Thus friendship is represented by a subset of the set of 2-element subsets, and “strangership” is the complementary subset. We can represent this in a diagram by drawing a red edge between two friends, and a blue edge between two strangers.

First, observe that five people do not suffice for the assertion of the party problem. For they might form the configuration shown in Figure 2.

We have to prove that six people do suffice. Suppose that the six people are A, B, C, D, E, F. Select one person, say A. Then, of the other five people, either three or more are friends of A, or else at most two are friends of A (in which case three or more are strangers to A). We have to deal with both cases, but if we find an argument for the first case, then by swapping red and blue, it will work for the second case.

So suppose that A has at least three friends. It doesn’t matter who they are, so suppose that B, C and D are all friends of A. If any pair of B, C, D are friends (say
B and C), then we have three mutual friends A, B, C. On the other hand, if no pair of them are friends, we have found our three mutual strangers B, C, D. So we are done in either case.

The party problem is the basis of a simple two-player game. Mark six dots on a piece of paper; one player takes a red pen and the other a blue pen. Players take turns in selecting two dots (not already joined) and joining them with an edge of their own colour. The first player to complete a triangle in their colour loses. The proof above guarantees that the game never ends in a draw; but it is not at all clear what a good strategy for the game should be!

The party problem is the first non-trivial case of a much more general theorem. We have coloured the 2-element subsets with 2 colours, and required there to be a monochromatic 3-element subset (one all of whose 2-subsets have the same colour). It turns out that the three numbers 2, 2, 3 here can all be generalised. I will state and prove one generalisation, and then state the most general case with a sketch of the proof.

The party problem generalised
Suppose that, rather than simply requiring three mutual friends or three mutual strangers, we wanted larger monochromatic sets. Our first generalisation of the party problem shows that this is always possible.

**Theorem 2.1** Suppose that $k$ and $l$ are given positive integers. Then there exists a positive integer $n$ such that, if the $2$-subsets of an $n$-set are coloured red or blue in any manner, there is either a $k$-set all of whose $2$-subsets are red, or an $l$-set all of whose $2$-subsets are blue.

**Proof** We will show that $n = \binom{k + l - 2}{k - 1}$ will do. Note that this is equal to $inom{k + l - 2}{l - 1}$, by Proposition 1.6. Our proof is by induction on the pair $(k, l)$; that is, for the inductive step, we assume the result for $(k - 1, l)$ and $(k, l - 1)$, and prove it for $(k, l)$. The party problem gives us the result for $(k, l) = (3, 3)$, but you might like to check for yourself that the result is true when one of $k$ and $l$ is equal to 2.

So suppose that $n$ has this value, and that the $2$-subsets of $\{1, \ldots, n\}$ are coloured red and blue. We claim that

- **either** there are at least $\binom{k + l - 3}{k - 2}$ points $i$ for which $\{1, i\}$ is red;

- **or** there are at least $\binom{k + l - 3}{k - 1}$ points $j$ for which $\{1, j\}$ is blue.

To see this, suppose that neither of these conditions held. Then there would be at most $\binom{k + l - 3}{k - 2} - 1$ points $i$ such that $\{1, i\}$ is red, and at most $\binom{k + l - 3}{k - 1} - 1$ points $j$ such that $\{1, j\}$ is blue. Together with the point 1, this gives at most

$$1 + \binom{k + l - 3}{k - 2} - 1 + \binom{k + l - 3}{k - 1} - 1 = \binom{k + l - 2}{k - 1} - 1$$

points altogether (using Proposition 1.7), contrary to our assumption.

So we have two possible cases.

**Case 1:** At least $\binom{k + l - 3}{k - 2}$ points are joined to 1 by red edges. Let $S$ be the set of these points. By the induction hypothesis for $(k - 1, l)$, we see that
either there is a \((k - 1)\)-set with all its 2-subsets red, or there is a \(l\)-set with all its 2-subsets blue. In the second case we are done; in the first, we can add 1 to the set to get the required \(k\)-set.

**Case 2:** At least \(\binom{k + l - 3}{k - 1}\) points are joined to 1 by blue edges. Now the argument is very similar, and I leave it to you: just interchange red and blue.

**Definition** The Ramsey number \(R_2(k, l)\) is defined to be the smallest number \(n\) for which the conclusions of the theorem hold.

We saw in the last section that \(R_2(3, 3) = 6\); six people suffice, but five do not, for the existence of either three mutual friends or three mutual strangers. Our theorem gives an upper bound:

\[
R_2(k, l) \leq \binom{k + l - 2}{k - 1}.
\]

However, the bound is not exact in general. For example, it shows that \(R_2(3, 4) \leq 10\) and \(R_2(4, 4) \leq 20\); the actual values are 9 and 18 respectively. In order to prove lower bounds, we need constructions (like the one we gave for five people earlier).

Finding exact values of Ramsey numbers is one of the most difficult open problems in combinatorics. \(R_2(k, k)\) is not known for any \(k\) greater than 4.

**Ramsey’s Theorem**

We can generalise the result in two further directions. First, we can specify the number of colours; we are not restricted to two colours. Second, we can colour subsets of any size, we are not restricted to colouring edges (though drawing pictures is much harder in general).

**Theorem 2.2 (Ramsey’s Theorem)** Let \(k, r, l_1, l_2, \ldots, l_r\) be given positive integers with \(k \leq l_i\) for \(i = 1, \ldots, r\). Then there exists a number \(n\) with the following property: if the \(k\)-subsets of \(\{1, \ldots, n\}\) are coloured with \(r\) colours \(c_1, \ldots, c_r\), then for some value of \(i \in \{1, \ldots, r\}\), there exists a set of \(l_i\) points, all of whose \(k\)-subsets have colour \(c_i\).

We will denote the smallest value of \(n\) for which this is true by \(R_k(l_1, l_2, \ldots, l_r)\).
I will postpone the general proof; but first, here are a couple more special cases which illustrate the new ideas required.

**Proposition 2.3** If the $2$-subsets of a set of (at least) $17$ points are coloured with three colours, then there exists a $3$-subset with all its $2$-subsets of the same colour.

**Proof** Choose a point $a$ of the set. There are at least $16$ further points, so there must be at least $6$ of them joined to $a$ by edges of the same colour. (If this failed, there could be at most $5+5+5=15$ points different from $a$.) Suppose, without loss of generality, that there are six points joined to $a$ by red edges, say $b, c, d, e, f, g$. If any two of these points were joined by red edges, there would be a red triangle (including $a$). If not, they are joined by blue or green edges only, and so by the party problem there will exist either a blue triangle or a green triangle.

**Proposition 2.4** Suppose that the $3$-subsets of a set of $n$ points are coloured red and blue, where $n > R_2(4,4) + 1$. Then there is either a $4$-set with all its $3$-sets red, or a $4$-set with all its $3$-sets blue.

**Proof** Choose a point $a$. Colour the $2$-subsets of the remaining points with two colours, say scarlet and turquoise, by the rule that $\{b, c\}$ is scarlet if $\{a, b, c\}$ is red, and $\{b, c\}$ is turquoise if $\{a, b, c\}$ is blue. Since there are at least $R_2(4,4)$ points different from $a$, we can find a set $A$ of size $4$ all of whose $2$-subsets are of the same colour; without loss of generality, scarlet. This means that $\{a, b, c\}$ is red for all $b, c \in A$.

If there exist three points $b, c, d \in A$ such that $\{b, c, d\}$ is red, then the $4$-set $\{a, b, c, d\}$ has all its $3$-subsets are red. If not, then all the $3$-subsets of $A$ are blue; since $|A| = 4$, we are done in this case too.

**Remark** We showed that $R_2(4,4) \leq 20$, so $R_3(4,4) \leq 21$. In fact the true value of $R_2(4,4)$ is $18$, and $R_3(4,4) = 13$.

There is also an infinite version of Ramsey’s theorem:

**Theorem 2.5 (Ramsey’s Theorem, infinite version)** Let $k$ and $r$ be given positive integers. Suppose that the $k$-subsets of an infinite set $X$ are coloured with $r$ colours $c_1, \ldots, c_r$ in any manner. Then, for some $i$, there exists an infinite subset $Y$ of $X$, all of whose $k$-subsets have colour $c_i$. 

22
This theorem is interesting to logicians for various reasons. (Frank Ramsey was a logician and mathematical economist.) One of these is that it is possible to deduce the finite form of the theorem from the infinite by a result from logic known as the Compactness Theorem; but it is not possible to deduce the infinite form from the finite. This leads to a statement which is true but unprovable from the axioms of arithmetic (the Paris–Harrington Theorem, which is a slight variant of the finite form of Ramsey’s Theorem.

I will prove the infinite Ramsey theorem in the appendix to this chapter.

The doocot principle

There is one special case of Ramsey’s Theorem where everything is known: the case \( k = 1 \). Invented by Dirichlet for applications in number theory in 1834, it is referred to as Dirichlet’s drawer principle (Schubfachprinzip in German) or pigeonhole principle. In St Andrews it seems natural to call it the doocot principle.

![A doo](image1.jpg) ![A doocot](image2.jpg)

**Proposition 2.6 (Doocot principle, simple form)** If \( n + 1 \) doos are in \( n \) holes in a doocot, there must be one hole which contains more than one doo.

**Proposition 2.7 (Doocot principle, general form)** Let positive integers \( r \) and \( l_1, l_2, \ldots, l_r \) be given. Suppose that \( n \geq l_1 + \cdots + l_r - r + 1 \). If \( n \) doos are in \( r \) holes \( H_1, \ldots, H_r \) in a doocot, then for some \( i \in \{1, \ldots, r\} \), hole \( H_i \) contains at least \( l_i \) doos. In other words,

\[
R_1(l_1, \ldots, l_r) = l_1 + \cdots + l_r - r + 1.
\]

This holds because if the conclusion fails, so that \( H_i \) contains at most \( l_i - 1 \) doos for \( i = 1, \ldots, r \), then the total number of doos is at most

\[
(l_1 - 1) + \cdots + (l_r - 1) = l_1 + \cdots + l_r - r,
\]
contraru to assumption. Also, it is the exact answer, since if we had one fewer
do, namely \( l_1 + \cdots + l_r - r \), then we could put \( l_i - 1 \) doos into \( H_i \).

In the solution to the party problem, we used this principle in the form \( R_1(3,3) = 5 \), when we said that of the five people B, C, D, E, F, either at least three are friends with A or at least three are strangers to A. Similarly, in the generalised party problem, we used

\[
R_1\left(\binom{k + l - 3}{k - 2}, \binom{k + l - 3}{k - 1}\right) = \binom{k + l - 3}{k - 2} + \binom{k + l - 3}{k - 1} - 1 = \binom{k + l - 2}{k - 1} - 1.
\]

Similarly, in the proof of \( R_2(3,3,3) \leq 17 \), we used \( R_1(6,6,6) = 16 \).

**The proof of Ramsey’s Theorem**  The proof is an induction. We show that

\[
R_k(l_1, \ldots, l_r) \leq 1 + R_{k-1}(A_1, \ldots, A_r),
\]

where

\[
A_i = R_k(l_1, \ldots, l_{i-1}, l_i - 1, l_{i+1}, \ldots, l_r).
\]

For suppose that we have a set \( X \) of \( n \) points, where \( n = 1 + R_{k-1}(A_1, \ldots, A_r) \), and colour the \( k \)-subsets of \( X \) with \( r \) colours \( c_1, \ldots, c_r \) in any manner. Choose a point \( x \in X \), and colour the \( (k-1) \)-subsets of \( X \setminus \{x\} \) with \( r \) new colours \( c_1^*, \ldots, c_r^* \) by the rule that a \( (k-1) \)-set \( B \) has colour \( c_i^* \) if and only if \( \{x\} \cup B \) has colour \( c_i \) in the original colouring. By assumption, there exists some value of \( i \), and a subset \( Y \) of size \( A_i \), such that all the \( (k-1) \)-subsets of \( Y \) have colour \( c_i^* \); that is, for any such subset \( B \), \( \{x\} \cup B \) has colour \( c_i \).

By definition of \( A_i \) we see that either

- for some \( j \neq i \), there is a subset \( Z \) of \( Y \) with \( |Z| = l_j \) and all \( k \)-subsets of \( Z \) have colour \( c_j \) (in which case we have succeeded); or
- there is a subset \( W \) of \( Y \) with \( |W| = l_i - 1 \) such that all the \( k \)-subsets of \( W \) have colour \( c_i \). In this case, by construction, all the \( k \)-subsets of \( \{x\} \cup W \) have colour \( c_i \), and \( |\{x\} \cup W| = l_i \), so again we have succeeded.

The induction is complete.

The induction on \( l_1, \ldots, l_r \) begins when some \( l_i \) is equal to \( k \). But we have

\[
R_k(l_1, \ldots, l_{r-1}, k) = R_k(l_1, \ldots, l_{r-1})
\]

[WHY?], so we can deal with these cases by an induction on \( r \). With some careful thought it can be seen that this induction scheme really is valid.
2.2 Steiner systems

It happens very frequently that mathematical discoveries are named after the wrong person. This is a good example: Kirkman beat Steiner by twelve years, but the name Kirkman systems has been applied something different, so the name Steiner systems has stuck.

**Definition** Let \( t, k, n \) be positive integers with \( t < k < n \). A *Steiner system* \( S(t,k,n) \) is a set \( S \) of \( k \)-subsets of an \( n \)-set \( X \), with the property that any \( t \)-subset of \( X \) is contained in a unique \( k \)-set in \( S \). The elements of \( S \) are called *blocks*.

The figure shows a \( S(2,3,7) \).

There are two convenient descriptions of this Steiner system. In the first, we number the points 0, 1, 2, \ldots, 6, by elements of the integers mod 7; we take one block to be \( B_0 = \{0,1,3\} \), and the other blocks are obtained by successively adding 1 \pmod{7}. Thus, \( B_i = \{i, i+1, i+3\} \), where the entries are interpreted \pmod{7}.

The second representation takes \( X = \{1,2,\ldots,7\} \), and

\[
S = \{\{1,2,3\}, \{1,4,5\}, \{1,6,7\}, \{2,4,6\}, \{2,5,7\}, \{3,4,7\}, \{3,5,6\}\}.
\]

To see how this works, write the numbers from 1 to 7 to the base 2; thus 1 = 001, 2 = 010, etc. Then each block consists of three triples whose sum mod 2 is zero; for example, \( \{3,5,6\} = \{011,101,110\} \), and the sum of these three binary vectors is 0.
If you have ever played Nim, you will recognise these as the non-trivial winning positions with three heaps of at most seven matches each. (The game is played with any number of heaps of matches, with any number of matches in each heap; a move for either player consists of taking any number of matches from a single heap. The last player to take a match wins.)

The big question is: for which values of \( t, k, n \) do Steiner systems exist? We are a very long way from solving this question! Below, you will learn much of what is known.

Let \((X, S)\) be a Steiner system \( S(t, k, n) \). Take \( i < t \), and let \( I \) be a set of \( i \) points of \( X \). Now take \( X' = X \setminus I \), and let \( S' \) be the collection of all \( (k - i) \)-subsets of \( X' \) formed as follows: take just those blocks \( B \in S \) which contain \( I \), and remove \( I \) from each of these blocks.

**Proposition 2.8** With the above construction, \((X', B')\) is a Steiner system \( S(t - i, k - i, n - i) \).

For clearly each block contains \( k - i \) points. Given a set of \( t - i \) points of \( X' \), together with \( I \) we have \( t \) points of \( X \), and so a unique block of \( S \) contains this set, giving a unique block of \( S' \) containing the given \( t - i \) points, as required.

**Proposition 2.9**

(a) The number of blocks of a Steiner system \( S(t, k, n) \) is

\[
\binom{n}{t} / \binom{k}{t}.
\]

(b) A necessary condition for \( S(t, k, n) \) to exist is that \( \binom{k - i}{t - i} \) divides \( \binom{n - i}{t - i} \) for \( i = 0, 1, \ldots, t - 1 \).

For (a), each of the \( \binom{n}{t} \) \( t \)-sets of points is contained in a unique block; but each block contains \( \binom{k}{t} \) \( t \)-sets.

For (b), we simply observe that the number of blocks in the Steiner system, and the number in each of the systems constructed in the preceding Proposition, must be a whole number.

The necessary conditions given by this result are not sufficient. However, in January 2014, Peter Keevash posted a paper on the arXiv proving the existence
of a function $F(t,k)$ such that, if $n \geq F(k,t)$ and $n$ satisfies the conditions of the above proposition, then a Steiner system $S(t,k,n)$ exists.

The smallest non-trivial example is $t = 2$, $k = 3$. In this case we refer to a Steiner triple system. Our necessary condition shows that $2$ divides $n - 1$ and $3$ divides $n(n - 1)/2$. The first condition shows that $n$ is odd, while the second rules out $n \equiv 5 \pmod{6}$. So

**Proposition 2.10** If a Steiner triple system on $n$ points exists, then $n \equiv 1 \text{ or } 3 \pmod{6}$.

In this case, the necessary condition is also sufficient, as was shown by Kirkman in 1847:

**Theorem 2.11** A Steiner triple system on $n$ points exists if and only if $n$ is congruent to $1$ or $3$ $(\text{mod} \ 6)$.

Unfortunately, I can’t give the proof here; the required constructions are quite ingenious.

For $t = 2$ and $k > 3$, an asymptotic result due to Richard Wilson shows that the necessary conditions are sufficient for large enough $n$. For $t = 3$, infinitely many examples are known, but (apart from the case $k = 4$) the gap between necessary and sufficient conditions is very wide. For $t > 3$, only a small finite number of examples is known. The most famous of these are the $S(5,6,12)$ and $S(5,8,24)$ systems, whose automorphism groups are the Mathieu groups $M_{12}$ and $M_{24}$ (two of the “sporadic” finite simple groups).

### 2.3 Appendix: The infinite Ramsey theorem

As the earlier comments suggest, we can’t prove the infinite form of Ramsey’s Theorem by a simple modification of the finite proof; something different is required.

I will take a fairly relaxed attitude to infinite sets; I will just assume that any infinite set $A$ contains a countable sequence $a_0, a_1, \ldots$ of distinct elements. (This requires a weak form of the Axiom of Choice; if you prefer, you can just replace “infinite” by “countably infinite” everywhere.)

**Theorem 2.12 (The infinite Doocot Principle)** If infinitely many doos live in a doocot with only finitely many holes, then some hole contains infinitely many doos.
This is obvious: the union of a finite collection of finite sets is finite.

**Theorem 2.13 (The infinite Ramsey Theorem)**  Let $k$ and $r$ be positive integers. Suppose that the $k$-element subsets of an infinite set $A$ are coloured with $r$ colours. Then there is an infinite subset $B$ of $A$, all of whose $k$-element subsets have the same colour.

**Proof**  The proof is by induction on $k$. The case $k = 1$ is just the infinite doocot principle. Assume that the result holds for $k - 1$, and suppose we have a colouring of the $k$-subsets of $A$ with $r$ colours.

We start by constructing three infinite sequences:

- an infinite sequence $a_0, a_1, a_2, \ldots$ of elements of $A$;
- an infinite descending sequence $B_0 \supseteq B_1 \supseteq B_2 \cdots$ of subsets of $A$, with $a_i \in B_j$ if and only if $i > j$;
- an infinite sequence $c_0, c_1, c_2, \ldots$ of colours

with the property that any $k$-subset of $\{a_i\} \cup Y_i$ containing $a_i$ has colour $c_i$.

We start with all three sequences empty.

Suppose we have constructed the sequences up to $a_{i-1}, Y_{i-1}, c_{i-1}$. Choose $a_i$ to be any point in $Y_{i-1}$. (If this set is countable, we can choose the lowest-numbered point.) Now colour the $(k - 1)$-subsets of $Y_{i-1} \setminus \{a_i\}$ with new colours in one-to-one correspondence with the old colours, where a set $L$ gets colour $c^*$ if $\{a_i\} \cup L$ had colour $c$ in the original colouring. By induction, there is a colour $c_i^*$ and an infinite subset $Y_i$ of $Y_{i-1} \setminus \{a_i\}$ such that every $(k - 1)$-subset of $Y_i$ has colour $c_i$. This completes the inductive step.

We end up with sequences of points, subsets, and colours with the above properties. Note that the points $a_i$ are all distinct.

Consider the sequence of colours $c_1, c_2, \ldots$. There are infinitely many terms, but they are all chosen from the finite set of colours; so some colour, say $c$, occurs infinitely often. Let $B = \{a_i : c_i = c\}$, an infinite subset of $A$. We claim that this is the required subset: all its $k$-subsets have colour $c$.

Take any $k$-subset of $B$, say $K = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\}$, where (we may assume) $i_1 < i_2 < \cdots < i_k$. By construction, $\{a_{i_1}, \ldots, a_{i_k}\}$ is a $(k - 1)$-subset of $Y_{i_1}$, and so has colour $c^*_{i_1} = c^*$; thus, $K$ has colour $c$, as required.
Exercises

2.1. Show that $R_2(2, \ell) = \ell$ for all $\ell$. More generally, $R_k(k, \ldots, k, \ell) = \ell$.

2.2. Prove the Handshaking Lemma:

**Lemma 2.14** In any group of people at a meeting, the number of people who have shaken hands an odd number of times is even.

Now use this lemma to show that $R_2(3, 4) \leq 9$, as follows. Examine carefully the proof in the notes to show that, if we have coloured the 2-subsets of $\{1, \ldots, 9\}$ so as to avoid both 3-sets with all edges red and 4-sets with all edges blue, then necessarily each point lies on three red and five blue edges. Now interpreting the red edges as handshakes, show that the Handshaking lemma is contradicted.

2.3. Show that $R_2(3, 4) \geq 9$ by finding a configuration of eight points with the edges coloured red and blue in such a way as to avoid sets of 3 with all edges red and sets of 4 with all edges blue. (*Hint:* Start with a red octagon.)

2.4. Dirichlet’s principle is used to prove the following result.

**Theorem 2.15** Let $\alpha$ be an irrational number. Then there exist infinitely many rational numbers $p/q$ for which

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2}.$$

Prove this theorem, by showing that for every natural number $n$ we can find a rational number $p/q$ with $q < n + 1$ and $\left| \alpha - \frac{p}{q} \right| < 1/nq$. For this, let $c_n = n\alpha - \lfloor n\alpha \rfloor$ for any $n$. Now the $n + 1$ numbers $c_1, c_2, \ldots, c_{n+1}$ lie in the unit interval. Divide the interval into $n$ equal parts. By the doocot principle, two of the numbers $c_i$ and $c_j$ must lie in the same part, and so differ by at most $1/n$.

2.5. A finite set of points in the plane is said to be in general position if no three of the points are collinear. A convex $n$-gon is a set of $n$ points in general position, none of which is in the interior of the polygon formed by the others.

(a) Prove that, given five points in general position in the plane, some four of the points form a convex 4-gon.

(b) Use Ramsey’s Theorem to prove that there is a function $F$ with the property that, given $F(n)$ points in the plane, some $n$ of them form a convex $n$-gon. (*Hint:* $F(n) = R_4(5, n)$. You need to show that given $n$ points, if any four of them form a convex 4-gon, then all the points form a convex $n$-gon.)
Remark  This result is called the *Happy End Theorem*. The picture below shows George and Esther Szekeres. Esther proved (a), and then George proved (b) (some years before the photograph was taken).

2.6. Suppose that $t+2$ is a prime number. Show that the divisibility conditions of Proposition 2.9(b) are satisfied for $k = t+1$ and $n = 2t+2$. [This does not mean that such Steiner systems exist. For $t = 9$, Mendelsohn and Hung showed by computer that $S(4,5,15)$ does not exist; hence by Proposition 2.9(a), $S(9,10,20)$ does not exist either.]

2.7. This exercise gives a product construction for Steiner triple systems. Let $(Y,T)$ and $(Z,U)$ be Steiner triple systems. Now let $X$ be the Cartesian product $Y \times Z$, and let $S$ consist of all the 3-sets $\{(y_1,z_1),(y_2,z_2),(y_3,z_3)\}$ for which

- either $y_1 = y_2 = y_3$, or $\{y_1,y_2,y_3\}$ is a block of $T$;
- either $z_1 = z_2 = z_3$, or $\{z_1,z_2,z_3\}$ is a block of $U$.

Show that $(X,S)$ is a Steiner triple system. [Note that we cannot have the first option in both clauses of the definition, since if $y_1 = y_2 = y_3$ and $z_1 = z_2 = z_3$ then the putative 3-set would have only one element.]

2.8. This exercise gives a “doubling” construction for Steiner triple systems. Let $(Y,T)$ be a Steiner triple system on $n$ points. Take $X = Y \times \{0,1\} \cup \{\infty\}$, and let $T$ consist of the 3-sets of one of the following types:
• \{∞, (y, 0), (y, 1)\}, for all \(y \in Y\);

• \{(y_1, e_1), (y_2, e_2), (y_3, e_3)\}, where \(\{y_1, y_2, y_3\}\) is a block of \((Y, T)\), \(e_1, e_2, e_3 \in \{0, 1\}\), and \(e_1 + e_2 + e_3\) is even.

Prove that \((X, S)\) is a Steiner triple system on \(2n + 1\) points.
3 Formal power series

Generating functions are the most powerful tool available to combinatorial enumerators. This week we are going to look at some of the things they can do.

3.1 Commutative rings with identity

In studying formal power series, we need to specify what kind of coefficients we should allow. We will see that we need to be able to add, subtract and multiply coefficients; we need to have zero and one among our coefficients. Usually the integers, or the rational numbers, will work fine. But there are advantages to a more general approach. A favourite object of some group theorists, the so-called Nottingham group, is defined by power series over a finite field.

A commutative ring with identity is an algebraic structure in which addition, subtraction, and multiplication are possible, and there are elements called 0 and 1, with the following familiar properties:

- addition and multiplication are commutative and associative;
- the distributive law holds, so we can expand brackets;
- adding 0, or multiplying by 1, don’t change anything;
- subtraction is the inverse of addition;
- 0 ≠ 1.

Examples include the integers Z (this is in many ways the prototype); any field (for example, the rationals Q, real numbers R, complex numbers C, or integers modulo a prime p, F_p).

Let R be a commutative ring with identity. An element u ∈ R is a unit if there exists v ∈ R such that uv = 1. The units form an abelian group under the operation of multiplication. Note that 0 is not a unit (why?).

3.2 Formal power series

A formal power series is, informally, an expression of the form

\[ a(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots, \]

like a polynomial but continuing for ever. However, if you start asking what x is, or what the infinite sum means, you will quickly run into trouble. So we proceed
in a more careful mathematical manner. One spin-off is that we get a definition of a polynomial at no extra cost.

**Definition** A formal power series is an infinite sequence \((a_0, a_1, a_2, \ldots)\) of elements taken from a commutative ring with identity \(R\). A polynomial is a formal power series \((a_0, a_1, a_2)\) for which there is some natural number \(n\) such that \(a_i = 0\) for \(i > n\); the smallest such \(n\) is the degree of the polynomial.

We always think of a formal power series as represented in the form

\[ a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots = \sum_{n \geq 0} a_n x^n. \]

At the moment, this is just a convenient way of writing it, but we will see that it is not unconnected with the notion of power series in analysis.

We use language based on this; we often call \(a_n\) the “coefficient of \(x^n\)” and refer to \(a_0\) as the “constant term” of the power series.

There are two typical occurrences of formal power series in combinatorial enumeration. Suppose we have numbers \(b_0, b_1, \ldots,\) for which \(b_n\) solves some counting problem on a set of size \(n\) (for example, the number of partitions of the set). Then we define

- the ordinary generating function or o.g.f. of the numbers is the formal power series
  \[ (b_0, b_1, b_2, b_3, \ldots) = \sum_{n \geq 0} b_n x^n; \]

- the exponential generating function or e.g.f. of the numbers is the formal power series
  \[ (b_0 / 0!, b_1 / 1!, b_2 / 2!, b_3 / 3!, \ldots) = \sum_{n \geq 0} \frac{b_n x^n}{n!}. \]

The reason for the name “exponential generating function” is the fact from analysis that the Taylor series of the exponential function is

\[ e^x = \sum_{n \geq 0} \frac{x^n}{n!}, \]

the exponential generating function of the very simple series \(1, 1, 1, \ldots\) of numbers. Incidentally, we will write the exponential function as \(\exp(x)\) rather than \(e^x\).
3.3 Operations on formal power series

There are many operations on formal power series; this is what gives them their flexibility and wide applicability. Here we meet some of these.

In every case, although our informal way of writing power series appears to involve an infinite sum, the definitions only ever ask us to add or multiply a finite number of elements.

**Addition** We define the sum of two formal power series term by term:

\[(a_0, a_1, \ldots) + (b_0, b_1, \ldots) = (a_0 + b_0, a_1 + b_1, \ldots),\]

or said otherwise,

\[\sum_{n \geq 0} a_n x^n + \sum_{n \geq 0} b_n x^n = \sum_{n \geq 0} (a_n + b_n) x^n.\]

From now on we will just use the second form; but we can always go back to the first form if required.

**Multiplication** We define the product of two formal power series by the convolution formula

\[\left(\sum_{n \geq 0} a_n x^n\right) \left(\sum_{n \geq 0} b_n x^n\right) = \sum_{n \geq 0} c_n x^n,\]

where

\[c_n = \sum_{k=0}^{n} a_k b_{n-k}.\]

Note that this is just what the informal representation of power series would lead you to expect, though it appears somewhat unmotivated if power series are written as infinite sequences.

**Proposition 3.1** Let \(R\) be a commutative ring with identity. Then the set \(R[[x]]\) of all formal power series over \(R\) is a commutative ring with identity, and the subset \(R[x]\) of all polynomials over \(R\) is also a commutative ring with identity.

We will not stop to prove this, which just involves somewhat tedious verification of axioms. But there is much more that formal power series can do!
Example: Summing a geometric series. Verify that

\[(1 - ax) \sum_{n \geq 0} a^n x^n = 1,\]

so it makes sense to write

\[\sum_{n \geq 0} a^n x^n = \frac{1}{1 - ax} = (1 - ax)^{-1}.\]

Infinite sums Let \(a^{(0)}, a^{(1)}, a^{(2)}, \ldots\) be formal power series over \(R\). Suppose that the following condition holds:

- for any \(n \geq 0\), there exists \(m \geq 0\) such that the first \(n\) coefficients of \(a^{(i)}\) are all zero if \(i > m\).

Then the infinite sum

\[a^{(0)} + a^{(1)} + a^{(2)} + \cdots = \sum_{i \geq 0} a^{(i)}\]

makes sense, since if we look in position \(n\) (the coefficient of \(x^n\)), there are only finitely many non-zero coefficients to add. (We adopt the convention that adding zero infinitely often has no effect.)

Various nice properties of infinite sums, such as the infinite distributive law, now hold.

Note that our interpretation of the formal power series \((a_0, a_1, \ldots)\) as the infinite sum \(\sum_{n \geq 0} a_n x^n\) is justified by this definition, since \(a_n x^n\) is itself a very simple formal power series (with \(a_n\) in position \(n\) and zero elsewhere), and these series satisfy our condition for an infinite sum to exist.

Infinite products Just as adding 0 infinitely often is supposed to leave things unchanged, similarly multiplying by 1 infinitely often does not change things.

As a result, if \(a^{(0)}, a^{(1)}, a^{(2)}, \ldots\) satisfy our condition for infinite sums to exist, then the infinite product

\[(1 + a^{(0)})(1 + a^{(1)}) \cdots = \prod_{n \geq 0} (1 + a^{(n)})\]

is defined. For we get a term in \(x^n\) in the product by selecting terms where the exponents of \(x\) sum to \(n\), and choosing 1 from all the remaining factors; and by assumption, only finitely many terms can be obtained in this way.
This is less familiar than infinite sums, so let us have a small example. Consider

\[ \prod_{m \geq 1} (1 + x^m). \]

If we write this as a power series \( \sum_{n \geq 0} a_n x^n \), what is the coefficient \( a_n \)?

We obtain a term in \( x^n \) by writing \( n = m_1 + m_2 + \ldots + m_r \), where \( m_1, m_2, \ldots, m_r \) are distinct positive integers, and choosing \( x^{m_i} \) from the bracket \( 1 + x^{m_i} \) for \( i = 1, \ldots, r \) and 1 from all the other brackets. So

The coefficient \( a_n \) is the number of expressions for \( n \) as a sum of distinct positive integers.

Thus,

\[ \prod_{m \geq 1} (1 + x^m) = 1 + x + x^2 + 2x^3 + 2x^4 + 3x^5 + 4x^6 + \cdots, \]

where the coefficient 4 of \( x^6 \) comes from the expressions

\[ 6 = 5 + 1 = 4 + 2 = 3 + 2 + 1. \]

**Differentiation** We can differentiate formal power series; no calculus involved, except that we steal from calculus the idea that the derivative of \( x^n \) is \( nx^{n-1} \). So

\[
\frac{d}{dx} \sum_{n \geq 0} a_n x^n = \sum_{n \geq 1} na_n x^{n-1}.
\]

We usually write the differential operator as \( D \) or \( D_x \) instead of \( d/dx \).

Now the usual rules for differentiating a sum or product of two functions hold. Indeed, *Leibniz’ rule* for arbitrary derivatives of a product hold:

\[
D^n (fg) = \sum_{k=0}^{n} \binom{n}{k} D^k f \cdot D^{n-k} g,
\]

as is easily proved by induction.

**Substitution** When can we substitute a power series \( g \) into a power series \( f \)? If \( f = \sum_{n \geq 0} a_n x^n \), we would like to have

\[ f(g) = \sum_{n \geq 0} a_n g^n. \]
So we need the powers $g^n$ to satisfy the condition for infinite sums to exist, that is, non-zero terms are pushed further to the right in higher powers. This will hold if the “constant term” of $g$ is equal to 0, since then the first $n$ terms of $g^n$ will all be zero. So the rule is:

Substitution $f(g)$ is defined if the constant term of $g$ is equal to zero.

Now everything you would expect to hold does hold, including the Chain Rule:

$$D(f(g)) = (Df)(g) \cdot Dg.$$ 

More generally, Faà di Bruno’s formula describes the $n$th derivative of $f(g)$, for any two formal power series $f$ and $g$. I won’t describe it here. (To my knowledge, Faà di Bruno is the only mathematician who was also a Catholic saint; among his students were C. Segre and G. Peano.)

### 3.4 Elementary properties

#### Multiplicative inverses

**Proposition 3.2** A formal power series $\sum_{n \geq 0} a_n x^n$ is a unit in the ring $R[[x]]$ if and only if its constant term $a_0$ is a unit in the ring $R$.

**Proof** Suppose that

$$\left( \sum_{n \geq 0} a_n x^n \right) \left( \sum_{n \geq 0} b_n x^n \right) = 1.$$

Then, considering the constant term, we see that $a_0 b_0 = 1$, so that $a_0$ is a unit.

Conversely, suppose that $a_0$ is a unit, and we want to find a formal power series $\sum_{n \geq 0} b_n x^n$ satisfying the above displayed equation. We see that $a_0 b_0 = 1$, so $b_0$ is the inverse of $a_0$, which is unique (Why?)

For $n > 0$, we want

$$\sum_{k=0}^{n} a_k b_{n-k} = 0.$$

Moving all terms except $a_0 b_n$ to the right and multiplying by $b_0$ (the inverse of $a_0$), we see that

$$b_n = -b_0 \left( \sum_{k=1}^{n} a_k b_{n-k} \right).$$

Thus $b_n$ is determined by the known coefficients $a_i$ and the coefficients $b_0, b_1, \ldots, b_{n-1}$. These can be found recursively, obtaining the inverse of the given power series.
Example: Fibonacci numbers  The Fibonacci numbers $F_0, F_1, F_2, \ldots$ satisfy the recurrence relation

$$F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2.$$ 

There is some argument about what the first two terms are: three common conventions are

- $F_0 = 0, F_1 = 1$;
- $F_0 = F_1 = 1$;
- $F_0 = 1, F_1 = 2$.

Clearly, once the first two terms are chosen, the rest of the sequence is determined; and the sequences resulting from these three conventions differ only by a shift.

All three conventions have something to recommend them. The second and third solve counting problems: the second gives the number of compositions of $n$ as a sum of ones and twos, while the third gives the number of binary sequences with no two consecutive ones. [Exercise: Prove these!] The first convention has no such simple counting interpretation, but has the nice properties that $\gcd(F_m, F_n) = F_{\gcd(m,n)}$, and that $F_n$ is prime only if $n$ is prime; also that $F_{12} = 12^2$ (the only time this happens apart from $F_1 = 1^2$).

Fibonacci’s famous rabbits seem to follow the second convention, but Fibonacci actually used the third convention, as this page from his book Liber Abaci shows: the calculation on the right of the picture in Figure 3 (from Wikipedia) gives $F_{12} = 377$.

Here, I will adopt the middle convention.

Let $\phi(x) = \sum_{n \geq 0} F_n x^n$, and calculate the coefficient of $x^n$ in $(1 - x - x^2) \phi(x) = (1 - x - x^2)(1 + x + 2x^2 + \cdots)$.

The constant term is clearly 1, and the coefficient of $x$ is 0. For $n \geq 2$, the coefficient of $x^n$ is

$$F_n - F_{n-1} - F_{n-2} = 0.$$ 

So $(1 - x - x^2) \phi(x) = 1$, and $\phi(x) = (1 - x - x^2)^{-1}$.

From this we can find a formula for the Fibonacci numbers. Write

$$1 - x - x^2 = (1 - \alpha x)(1 - \beta x),$$
Figure 3: Fibonacci’s numbers
where $\alpha$ and $\beta$ are roots of the quadratic $y^2 - y - 1 = 0$. Then write $(1 - x - x^2)^{-1}$ as partial fractions:

$$
\frac{1}{1 - x - x^2} = \frac{A}{1 - \alpha x} + \frac{B}{1 - \beta x} = A \sum_{n \geq 0} \alpha^n x^n + B \sum_{n \geq 0} \beta^n x^n,
$$

by solving two linear equations for $A$ and $B$. Then we can read off the formula:

$$F_n = A \alpha^n + B \beta^n.$$

We see that, for any sequence satisfying the Fibonacci recurrence for $n \geq 2$, whatever the first two terms are, the generating function will have the form $(a + bx)(1 - x - x^2)^{-1}$. So we will obtain a formula of the same shape, with the same $\alpha$ and $\beta$, but with a different $A$ and $B$.

**Example: Connected permutations** The purpose of this example is to convince you that even power series with zero radius of convergence can count something interesting.

A permutation $\pi$ of the set $\{1, \ldots, n\}$ is said to be *connected* if there does not exist a number $k$ with $0 < k < n$ such that $\pi$ permutes the numbers $1, \ldots, k$ among themselves (and so also the numbers $k+1, \ldots, n$ among themselves). Let $c_n$ be the number of connected permutations. Can we calculate $c_n$?

Given any permutation $\pi$, let $k$ be the smallest positive number for which $\pi$ permutes among themselves the numbers $1, \ldots, k$. Thus $\pi$ is connected if and only if $k = n$. Now $\pi$ is obtained by stitching together a connected permutation of $\{1, \ldots, k\}$ and an arbitrary permutation of $\{k+1, \ldots, n\}$. Thus,

$$n! = \sum_{k=1}^{n} c_k (n-k)!.$$

Put $A(x) = \sum_{n \geq 0} n! x^n$ (a power series which converges only for $x = 0$) and $C(x) = 1 - \sum_{n \geq 1} c_n x^n$. What is the coefficient of $x^n$ in $C(x)A(x)$? It is clearly 1 for $n = 0$. If $n > 0$, this coefficient is

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

So $C(x)A(x) = 1$, and $C(x) = A(x)^{-1}$. 40
Substitution inverses

Recall that, to substitute $g$ into $f$, we require that $g$ has constant term 0. Inverses for substitution should then satisfy $f(g(x)) = x = g(f(x))$. For this we require that both $f$ and $g$ have zero constant term. Arguing in a similar way to the last result, we need the coefficient of $x$ to be a unit in $R$, and this condition is also sufficient.

Lagrange inversion, which we may or may not get to, gives a formula for the inverse of a power series satisfying these two conditions (constant term zero and coefficient of $x$ a unit).

In particular, the set of formal power series $x + \cdots$ (constant term 0, coefficient of $x$ equal to 1) forms a group under the operation of composition. This group is sometimes referred to as the Nottingham group.

3.5 Connection with analysis

We’ve seen that we can manipulate formal power series without paying any regard to whether or not they converge – we can even look at formal power series over rings where convergence would make no sense. But the good news is that, over the real or complex numbers, if our series are convergent for some non-zero values of $x$, then we can use all the tools of analysis on them.

The most important case of this is the following principle:

Any identity between real or complex power series, involving addition, multiplication (possibly infinite sums and products) and substitution, is an identity in the ring of formal power series.

This is because of the uniqueness of the Taylor series for an analytic function. We will see many examples later.

Another, less common but occasionally useful, is the use of Cauchy’s formula to extract the coefficients of a power series. If $f(z) = \sum_{n \geq 0} a_n z^n$ is analytic in a disc, then

$$a_n = \frac{1}{2\pi i} \oint \frac{f(z)}{z^{n+1}} \, dz,$$

where the contour contains the origin and is contained in the disc of convergence.
The binomial series  We already defined the binomial coefficient \( \binom{a}{k} \) for any real or complex number \( a \) and any non-negative integer \( k \), by the rule

\[
\binom{a}{k} = \frac{a(a-1) \cdots (a-k+1)}{k!}.
\]

Now the Binomial Theorem gives the Taylor series of the analytically defined function \((1 + x)^a\) (which is analytic in the open unit disc):

**Theorem 3.3 (Binomial Theorem for arbitrary exponent)**

\[
(1 + x)^a = \sum_{k \geq 0} \binom{a}{k} x^k.
\]

This is a theorem of analysis, since the function \((a + x)^a\) has no combinatorial definition. Indeed, we can take it as a definition of the function \((1 + x)^a\).

Nevertheless, it is very useful to us. For example, it can be proved analytically that

\[
(1 + x)^a(1 + x)^b = (1 + x)^{a+b}
\]

is valid in the unit disc; so it is an identity for formal power series, by our general principle. This means that the Vandermonde convolution

\[
\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}
\]

holds for any real or complex values of \( a \) and \( b \).

The exponential and logarithmic series  These are two further important series:

\[
\exp(x) = \sum_{n \geq 0} \frac{x^n}{n!},
\]

\[
\log(1 + x) = \sum_{n \geq 1} \frac{(-1)^{n-1} x^n}{n}.
\]

The first is convergent everywhere, the second inside the unit disc. They satisfy all the familiar identities, by our general principle: for example,

\[
\log(1 + x)^a = a \log(1 + x),
\]

\[
\exp(\log(1 + x)) = 1 + x,
\]

\[
\log(1 + (\exp(x) - 1)) = x.
\]
Often it is convenient to use a slightly different form of the logarithmic series:

$$-\log(1-x) = \sum_{n \geq 1} \frac{x^n}{n}.$$

Note that we can only substitute a power series into another if its constant term is zero. So the slightly convoluted form of the third identity is because \(\exp(x)\) has constant term 1, so we must subtract 1 before we substitute it; and in the first identity, we have to observe that \((1+x)^a = 1 + g(x)\), where \(g\) is a formal power series with constant term 0.

The logarithm has the property that it converts products into sums; in fact, it converts infinite products into infinite sums. In the next section, I will use this property, for series with non-zero radius of convergence, and so I can appeal to our analytic principle. But in fact it holds generally, and so is a combinatorial property, which really requires a combinatorial proof!

### 3.6 Further examples

We end this chapter with a couple of examples to show generating functions at work: to show the existence of finite fields, and to find an explicit formula for the number of bracketings of a non-associative product. The first involves taking logs, the second the binomial theorem with non-integer exponent.

**Irreducible polynomials** Before we plunge in to this topic, we need an elementary combinatorial result. We know that the number of ways of choosing \(k\) objects from a set of size \(n\), if we are not allowed to repeat an object in the selection and we don’t care about the order, is \(\binom{n}{k}\). What if we are allowed repetitions?

**Proposition 3.4** The number of selections of \(k\) objects from a set of \(n\), with repetitions allowed and order not important, is \(\binom{n+k-1}{k}\).

The proof is in the appendix.

Galois showed that there exists a finite field of any given prime power order. The way to construct a finite field of order \(q^n\), if you know one of order \(q\), is to find an irreducible polynomial of degree \(n\) over the smaller field, and adjoin a root of it. (You can take \(q\) to be prime here, since then the integers mod \(q\) will do for
a starting field.) So we need to prove that there is an irreducible polynomial of degree \( n \) over a field of order \( q \), for any \( n \) and \( q \).

How do we do this? By counting the irreducible polynomials!

We need only consider monic polynomials. We regard \( q \) as fixed, and the starting field \( F \) of order \( q \) also as fixed. Let \( a_n \) be the number of irreducible polynomials of degree \( n \) over \( F \).

The total number of monic polynomials of degree \( n \) is \( q^n \). We have a unique factorisation theorem to say that each is uniquely a product of irreducibles. Suppose that there are \( m_1 \) factors of degree 1, \( m_2 \) of degree 2, and so on, with \( m_1 + 2m_2 + \cdots = n \). The number of ways of choosing \( m_i \) irreducibles from a set of size \( a_i \) (with replacement, and order unimportant) is \( \binom{a_i + m_i - 1}{m_i} \). So

\[
\sum_{m_1, m_2, \ldots} \prod_{i \geq 1} \binom{a_i + m_i - 1}{m_i} = q^n.
\]

This is a recurrence relation, since there is a term \( a_n \) (corresponding to irreducible polynomials) and all the other terms involve \( a \) with smaller index. But it looks like a nightmare to solve! But let’s plunge in . . .

\[
\frac{1}{1 - qx} = \sum_{n \geq 0} q^n x^n = \sum_{n \geq 0} \sum_{m_1 + 2m_2 + \cdots = n} \prod_{i \geq 1} \binom{a_i + m_i - 1}{m_i} x^{im_i}.
\]

In the last line, instead of summing over \( n \), and then for each \( n \) summing over \( m_1, m_2, \ldots \) satisfying \( m_1 + 2m_2 + \cdots = n \), we simply sum over all \( m_1, m_2, \ldots \) independently; and we split the \( x^n \) over the terms in the sum as shown.

Now I claim that the last sum is equal to

\[
\prod_{i \geq 1} \sum_{m \geq 0} \binom{a_i + m - 1}{m} x^{im}.
\]

For this expression is a product of sums; we expand the brackets and get a sum of terms, one from each bracket (but to make things work, choosing 1 from all but finitely many of the brackets). Note that our infinite product condition is satisfied.
Now a simple calculation shows that
\[
\binom{a + m - 1}{m} = (-1)^m \binom{-a}{m};
\]
so finally we obtain
\[
(1 - qx)^{-1} = \prod_{i \geq 1} (1 - x^i)^{-a_i}.
\]

At this point, we can take logs, and find
\[
-\log(1 - qx) = -\sum_{i \geq 1} a_i \log(1 - x^i),
\]
or
\[
\sum_{n \geq 1} \frac{(qx)^n}{n} = \sum_{i \geq 1} a_i \sum_{k \geq 1} \frac{x^{ik}}{k}.
\]
Equating coefficients of powers of \(x\) gives
\[
\frac{q^n}{n} = \sum_{i | n} a_i \frac{n}{i};
\]
or
\[
q^n = \sum_{i | n} i a_i.
\]

This is a much simpler recurrence! It is easy to see directly that we must have \(a_i > 0\), so that irreducible polynomials exist. Moreover, Möbius inversion, which we meet later, can be used to solve it explicitly for \(a_n\). Finally, it is possible to show that there are just enough irreducible polynomials to build one finite field of order \(q^n\); so the uniqueness follows as well.

Just to see how much simpler the new recurrence is, here is the original recurrence for \(q = 2, n = 5\):
\[
a_5 + a_4 a_1 + a_3 a_2 + a_3 \left( \frac{a_1 + 1}{2} \right) + \left( \frac{a_2 + 1}{2} \right) a_1 + a_2 \left( \frac{a_1 + 2}{3} \right) + \left( \frac{a_1 + 4}{5} \right) = 2^5;
\]
if you know \(a_1 = 2, a_2 = 1, a_3 = 2, a_4 = 3\) and have a bit of perseverance, you can find that \(a_5 = 6\). The new recurrence is simply
\[
5a_5 + a_1 = 32.
\]
Bracketings of a formula

Suppose that $\odot$ is a binary operation which is not necessarily associative, so that the value of a product of $n$ terms might depend on the way the terms are bracketed. How many different bracketings are there?

For example, for $n = 4$, there are five bracketings:

\[ ((a \odot b) \odot c) \odot d, (a \odot (b \odot c)) \odot d, (a \odot b) \odot (c \odot d), a \odot ((b \odot c) \odot d), a \odot (b \odot (c \odot d)). \]

Let $C_n$ be the number of bracketings.

**Proposition 3.5** $C_n = \frac{1}{n} \binom{2n - 2}{n - 1}$.

I do not know any simple direct way to prove this formula. We will detour via generating functions. Let $F(x) = \sum_{n \geq 1} C_n x^n$.

Since any bracketing of $n$ terms (for $n > 1$) is obtained by bracketing the first $k$ and the last $n - k$ and finally combining the results, we have

$$C_n = \sum_{k=1}^{n-1} C_k C_{n-k} \text{ for } n > 1.$$  

We see on the right the convolution product, so that $F(x)^2$ is almost the same as $F(x)$: they agree for all powers of $x$ greater than the first. But $F(x)$ has a term $x$, whereas $F(x)^2$ begins with $x^2$. So

$$F(x) = x + F(x)^2.$$  

Now this is a quadratic equation $F^2 - F + x = 0$, which we can solve:

$$F(x) = \frac{1}{2} \left( 1 \pm \sqrt{1 - 4x} \right).$$  

Which sign to choose? Since the constant term is zero, we must choose the negative sign:

$$F(x) = \frac{1}{2} \left( 1 - \sqrt{1 - 4x} \right).$$  

Now we can obtain $C_n$ by extracting the coefficient of $x^n$, using the binomial theorem with exponent $\frac{1}{2}$:

$$C_n = \frac{1}{2} \binom{1/2}{n} (-4)^n$$

46
The numbers \( C_n \) are the Catalan numbers, about which we shall have a lot more to say.

### 3.7 Appendix: selection with repetition

**Proposition 3.6** The number of selections of \( k \) objects from a set of \( n \), with repetitions allowed and order not important, is \( \binom{n+k-1}{k} \).

**Proof** To specify such a selection, we need to say, for each of the \( n \) objects \( o_1, \ldots, o_n \), the number of times it is chosen. These numbers \( x_1, \ldots, x_n \) are non-negative integers and sum to \( k \). So we have to prove that the number of choices of \( n \) non-negative integers with sum \( k \) is \( \binom{n+k-1}{k} = \binom{n+k-1}{n-1} \).

To do this, take a row of \( n+k-1 \) boxes in a row, and put “barriers” into \( n-1 \) of the boxes. Then take \( x_1 \) to be the number of boxes before the first barrier, \( x_2 \) the number between the first and second barriers, \( \ldots \), and \( x_n \) the number of boxes after the last barrier. Clearly the required conditions hold. Conversely, if \( x_1, \ldots, x_n \) are non-negative with sum \( k \), then put the first barrier after \( x_1 \) empty boxes, the second after another \( x_2 \) empty boxes, and so on.

So the required number is the number of choices of positions for the barriers, that is, \( \binom{n+k-1}{n-1} \).

For example, if \( n = 4 \) and \( k = 5 \), then the picture

\[
\begin{array}{|c|c|c|c|c|}
\hline
& & * & * & *
\hline
\end{array}
\]

corresponds to \( x_1 = 3, x_2 = 0, x_3 = 1, x_4 = 1 \), in other words, to the selection \([o_1,o_1,o_3,o_4] \).
Exercises

3.1. Now is the time to have a serious attempt at the last question on Sheet 1.

3.2. In the equation \( \exp(x + y) = \exp(x) \exp(y) \), calculate the terms of total degree \( n \) in the two variables, and deduce the Binomial Theorem for positive integer exponents.

3.3. We saw that 
\[
F_n = A\alpha^n + B\beta^n.
\]
Calculate \( \alpha \) and \( \beta \), and \( A \) and \( B \) for each of the three conventions in the text.

3.4. Show that any sequence satisfying the Fibonacci recurrence 
\[
F_n = F_{n-1} + F_{n-2}
\]
for \( n \geq 2 \) is given by the formula in the preceding question, with the same values of \( \alpha \) and \( \beta \) but possibly different \( A \) and \( B \).

3.5. Let \( A(x) \) be the generating function for a sequence \( (a_n) \). Show that the generating function for the sequence of partial sums is \( A(x)/(1-x) \).

3.6. Let \( F(x) = \prod_{n \geq 1} (1-x^n)^{-1} \).

(a) If \( F(x) = \sum_{n \geq 0} a_n x^n \), show that \( a_n \) is the number of partitions of \( n \), expressions for \( n \) as a sum of positive integers, where the order is not significant.

(b) Let \( F(x)^{-1} = \sum_{n \geq 0} b_n x^n \). Give a description of the coefficients \( b_n \).

(c) Work out the first few numbers \( b_n \). What do you find?

3.7. Find the inverse (under substitution) of the power series \( x + x^2 \). If your answer is the series \( f(x) \), calculate the first few terms of \( f(x) + f(x)^2 \) to check your result.

3.8. In an election, there are two candidates, A and B. The number of votes cast is \( 2n \), and each candidate receives \( n \) votes. Show that the number of ways in which the votes can be counted so that at every intermediate stage, A has more votes than B, is the Catalan number \( C_n \). [Hint: Let the number be \( f(n) \). A leads by just one vote after the first vote is counted. Suppose that this next happens after \( 2i+1 \) votes have been counted. Show that there are \( f(i) \) ways to count the votes before this point, and \( f(n - i) \) ways after this point.]

Now show that the number of ways in which the votes can be counted so that at every intermediate stage, A has at least as many votes as B, is \( C_{n+1} \). [Hint: Give
A one extra vote at the start and B one extra vote at the end; then the conditions of the preceding paragraph are satisfied.]

3.9. A clown stands on the edge of a swimming pool, holding a bag containing \( n \) red and \( n \) blue balls. He draws the balls out one at a time and discards them. If he draws a blue ball, he takes one step back; if a red ball, one step forward. (All steps have the same size.) Show that the probability that the clown remains dry is \( 1/(n+1) \).

3.10. [*] A simple version of the \textsc{Quicksort} algorithm to sort a list \( L \) into the correct order can be specified recursively as follows:

- Let \( a \) be the first element of the list, and partition the remainder into \( L^- \) (the elements less than \( a \)) and \( L^+ \) (the elements greater than \( a \)).
- Sort \( L^- \) and \( L^+ \).
- Return \((\text{sort}(L^-), a, \text{sort}(L^+))\).

Suppose that \( q_n \) is the average number of comparisons required to sort a list \( L \) of length \( n \) (in random order). Prove that

\[
q_n = n - 1 + \frac{1}{n} \sum_{k=1}^{n} (q_{k-1} + q_{n-k}) = n - 1 + \frac{2}{n} \sum_{k=0}^{n-1} q_k.
\]

[Hint: \( a \) is equally likely to be the smallest, second, \ldots, largest element of \( L \).]

Let \( Q(x) = \sum_{n \geq 0} q_n x^n \). Multiply the recurrence relation by \( nx^n \) and sum:

\[
\sum_{n \geq 0} n q_n x^n = \sum_{n \geq 0} n(n-1)x^n + 2 \sum_{n \geq 0} \left( \sum_{i=0}^{n-1} q_i \right) x^n.
\]

Deduce that

\[
xQ'(x) = \frac{2x^2}{(1-x)^3} + \frac{2x}{1-x} Q(x).
\]

Solve this differential equation to obtain

\[
Q(x) = \frac{-2(x + \log(1-x))}{(1-x)^2}.
\]

Deduce that

\[
q_n = 2(n+1) \sum_{i=1}^{n} \left( \frac{1}{i} \right) - 4n = 2n \log n + O(n).
\]
4 Linear recurrences with constant coefficients

The material in this section is standard stuff, not terribly exciting maybe, but quite a few interesting examples fit into this framework. It is one of the few parts of enumerative combinatorics where we know everything. We saw the example of the Fibonacci numbers in the last chapter; the general picture is very similar.

4.1 The problem and its solution

The general set-up here is sequences which satisfy a recurrence relation

\[ a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k} \]

for all \( n \geq k \), where \( c_1, c_2, \ldots, c_k \) are constants. It is clear that, whatever values we give to the first \( k \) terms \( a_0, a_1, \ldots, a_{k-1} \) of the sequence, then the rest of the sequence is uniquely determined by the recurrence relation. We call the number \( k \) here the degree of the recurrence.

Let \( A(x) \) be the generating function \( \sum_{n \geq 0} a_n x^n \).

**Proposition 4.1** There is a polynomial \( p(x) \), of degree strictly less than \( k \), such that

\[ A(x) = \frac{p(x)}{1 - c_1x - c_2x^2 - \cdots - c_kx^k}. \]

**Proof** We consider \( A(x)(1 - c_1x - c_2x^2 - \cdots - c_kx^k) \). For \( n \geq k \), the coefficient of \( x^n \) is

\[ a_n - c_1a_{n-1} - c_2a_{n-2} - \cdots - c_ka_{n-k} = 0. \]

So the product has terms in \( x^0, x^1, \ldots, x^{k-1} \) only.

Note that the coefficients of the polynomial \( p(x) \) can be found from the values of \( a_0, a_1, \ldots, a_{k-1} \).

In order to extract a formula for \( a_n \) from this expression, we use the technique of partial fractions. I will state the result without proof, and do a couple of examples.

**Proposition 4.2** Suppose that \( q(x) = \prod_{i=1}^{r} (1 - \alpha_ix)^{m_i} \) for distinct complex numbers \( \alpha_i \) and positive integers \( m_i \) summing to the degree \( k \) of \( q \). Then, for any polynomial
$p(x)$ of degree strictly less than $k$, there are constants $b_{ij}$ (for $1 \leq i \leq r$ and $1 \leq j \leq m_i$) such that

$$\frac{p(x)}{q(x)} = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{b_{ij}}{(1 - \alpha_i x)^j}.$$ 

Note that, if $q(x) = 1 - c_1 x - c_2 x^2 - \cdots - c_k x^k$, then the $\alpha_i$ are the roots of $y^k - c_1 y^{k-1} - c_2 y^{k-2} - \cdots - c_k = 0$; by the Fundamental Theorem of Algebra, this equation has $k$ roots, not necessarily distinct, so we group equal roots together to apply the proposition.

**Example**  Suppose that $a_0 = 1$, $a_1 = 3$, $a_2 = 9$, and $a_n = 5a_{n-1} - 8a_{n-2} + 4a_{n-3}$ for $n \geq 3$. We have

$$q(x) = 1 - 5x + 8x^2 - 4x^3 = (1 - x)(1 - 2x)^2,$$

so the generating function for the sequence is $p(x)/((1 - x)(1 - 2x)^2)$ for some polynomial $p$. We have

$$p(x) = (1 + 3x + 9x^2 + \cdots)(1 - 5x + 8x^2 - 4x^3) = 1 - 2x + 2x^2.$$

Now we have

$$\frac{1 - 2x + 2x^2}{(1 - x)(1 - 2x)^2} = \frac{c}{1 - x} + \frac{d}{1 - 2x} + \frac{e}{(1 - 2x)^2},$$

$$= c \sum_{n \geq 0} x^n + d \sum_{n \geq 0} (2x)^n + e \sum_{n \geq 0} \left( -\frac{2}{n} \right) (-2x)^n.$$

Now

$$\left( -\frac{2}{n} \right) = \frac{(-2)(-3)\cdots(-n-1)}{n!} = (-1)^n (n+1),$$

so

$$a_n = c + (d + (n+1)e)2^n.$$

The values $c, d, e$ can now be obtained in either of two ways:

- use the values $a_0, a_1, a_2$ to get three equations in three unknowns; or
- obtain them from the continued fraction expansion.
The latter is often easier since we get some information for free. In this case, clearing denominators gives

\[ 1 - 2x + 2x^2 = c(1 - 2x)^2 + d(1 - x)(1 - 2x) + e(1 - x). \]

Putting \( x = 1 \) we obtain \( 1 - 2 + 2 = c \), so \( c = 1 \). Putting \( x = \frac{1}{2} \) gives \( 1 - 1 + \frac{1}{2} = \frac{1}{2} e \), so \( e = 1 \). Only \( d \) remains to be found, and it turns out that \( d = -1 \). So \( a_n = 1 + n2^n \).

Note that \( \binom{-2}{n} \) is a linear polynomial in \( n \). More generally, \( \binom{-j}{n} \) is a polynomial of degree \( j - 1 \).

To summarise:

**Theorem 4.3** Suppose that the sequence \( (a_n) \) satisfies the recurrence

\[ a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k} \]

for all \( n \geq k \), where \( c_1, c_2, \ldots, c_k \) are constants. Let

\[ 1 - c_1x - c_2x^2 - \cdots - c_kx^k = \prod_{i=1}^{r}(1 - \alpha_i x)^{m_i}, \]

where \( \alpha_i \) are distinct complex numbers and \( m_i \) positive integers. Then

\[ a_n = \sum_{i=1}^{r} h_i(n) \alpha_i^n, \]

where \( h_i \) is a polynomial of degree at most \( m_i - 1 \).

Here

\[ h_i(x) = \sum_{j=1}^{m_i} b_{ij} \binom{-j}{n}, \]

in the earlier notation.

### 4.2 Regular languages

We are given an alphabet \( A \) with \( q \) symbols. A word is just a finite string of symbols taken from the alphabet. So the number of words of length \( n \) is \( q^n \). A language over \( A \) is a (finite or infinite) set of words.

We are going to count the words of length \( n \) in languages of a particular class, called regular languages. These can be defined in many ways; the definition I
A finite-state automaton is a machine which can be in any one of a finite set \( S \) of internal states. It reads a tape on which are written symbols from the alphabet \( A \). It operates according to rules which tell it what to do when it is in a given state and is reading a given symbol on the tape. For a deterministic automaton, there is just one rule which applies: it moves one step to the right, to scan the next symbol on the tape, and the internal state changes to one determined by the previous state and the symbol read. I will just say “automaton” rather than “finite-state deterministic automaton”.

There is one more bit of structure. One of the states is distinguished as the initial state, and another (or maybe the same one) as the accepting state. Now what happens when the machine operates? We write a word \( w \) on the tape, and start the automaton in its initial state scanning the first letter of the word. It operates until it has read the entire word. At this point, it may or may not be in the accepting state. We say that the automaton accepts \( w \) if it ends up in the accepting state.

Sometimes the definition is modified by allowing more than one accepting state. Everything we say below can be modified to work in the more general case.

Now a regular language is the set of words which are accepted by some finite-state deterministic automaton.

An automaton can be represented by a diagram in which we take one node for each state, and put an arc labelled with the alphabet symbol \( a \) from state \( s \) to state \( t \) if, when the automaton is in state \( s \) reading \( a \), it changes into state \( t \). Now the defining property of a deterministic automaton is that, for each alphabet symbol \( a \), there is a unique arc labelled \( a \) leaving any given node.

Here is a simple example. The automaton has three states; the left-hand state in the diagram is the initial state and the right-hand state is the accepting state. The alphabet has two symbols, \( a \) and \( b \).

This automaton accepts the words \( a, aaaa, aba, \) and \( bbba \), but not the words \( ab, aa, bbb \), for example. **Problem:** Can you describe the language accepted by the automaton (that is, give a simple criterion for a word to be accepted)?
**Theorem 4.4** Let $L$ be a regular language accepted by an $s$-state automaton, and let $a_n$ be the number of words of length $n$ in $L$. Then the sequence $(a_0, a_1, a_2, \ldots)$ satisfies a linear recurrence with constant coefficients of degree at most $s$.

**Proof** The key observation is:

$a_n$ is equal to the number of walks from the initial state to the accepting state, following arcs in the diagram.

For any such walk generates a word (the sequence of labels on the arcs), and conversely any word describes a walk.

Suppose the states are $s_1, s_2, \ldots, s_n$, where $s_p$ is the initial state and $s_q$ is the accepting state. We construct an $n \times n$ matrix $M$ by the rule that $M_{ij}$ is the number of arcs from state $i$ to state $j$. Now the rule for matrix multiplication, and induction, show that the $(i,j)$ entry of $M^n$ is equal to the number of walks of length $n$ from $s_i$ to $s_j$. So

$$a_n = (M^n)_{pq}. $$

The matrix $M$ has a minimal polynomial of degree at most $n$, a polynomial $p(x)$ such that $p(M) = 0$ (and $p$ is chosen to be the monic polynomial of smallest possible degree such that this holds). Suppose that

$$p(x) = x^r + b_{r-1}x^{r-1} + \cdots + b_0. $$

Substituting $x = M$ and multiplying by $M^{n-r}$, we find

$$M^{n-r}p(M) = M^n + b_{r-1}M^{n-1} + \cdots + b_0M^{n-r} = 0. $$

Taking the $(p,q)$ entry of this matrix equation gives

$$a_n + b_{r-1}a_{n-1} + \cdots + b_0a_{n-r} = 0, $$

and taking all terms except $a_n$ to the right-hand side gives us the required recurrence relation.

**Remark** To allow for a set $Q$ of accepting states, we replace the single matrix entry $(M^n)_{pq}$ above by $\sum_{q \in Q} (M^n)_{pq}$. The argument still works.

Let us return to our example. The matrix $M$ has zero diagonal and all off-diagonal elements equal to 1. So $M + I$ is the all-1 matrix, which implies that $(M + I)^2 = 3(M + I)$, so $M^2 - M - 2I = 0$. So the numbers $a_n$ satisfy

$$a_n = a_{n-1} + 2a_{n-2}. $$

54
We have $a_0 = 0$ and $a_1 = 1$, since the automaton does not accept the empty word but accepts the word $a$. So the generating function $f(x) = \sum a_n x^n$ satisfies
\[ f(x)(1 - x - 2x^2) = c + dx, \]
and we get $c = 0$, $d = 1$.

Now calculation shows that
\[ f(x) = \frac{x}{1 - x - 2x^2} = \frac{-1/3}{1 + x} + \frac{1/3}{1 - 2x}, \]
so
\[ a_n = \frac{1}{3} (2^n - (-1)^n). \]
For example, $a_3 = (2^3 + 1)/3 = 3$, and the three words of length 3 in the language are $aab$, $aba$ and $baa$.

### 4.3 C-finite sequences

A sequence is called *C-finite* if it satisfies a finite-length linear recurrence relation with constant coefficients.

What we have seen earlier shows the following:

**Proposition 4.5** A sequence is C-finite if and only if its generating function has the form $p(x)/q(x)$, where $p$ and $q$ are polynomials with $\deg(p) < \deg(q)$ and $q(0) \neq 0$.

**Proposition 4.6** The set of C-finite sequences is closed under addition, scalar multiplication, convolution product, differentiation, and partial sum. In particular, the C-finite sequences form a vector space over the rational numbers.

**Proof** All these operations have simple effects on the generating function: addition, scalar multiplication, convolution product and derivative of sequences correspond to addition, scalar multiplication, ordinary multiplication, and differentiation of the generating function. All we have to do is to show that these operations preserve the class of rational functions specified in the preceding proposition.

For example, if $A_i(x) = p_i(x)/q_i(x)$ for $i = 1, 2$, where $\deg(p_i) < \deg(q_i)$ and $q_i(0) \neq 0$, then
\[ A_1(x)A_2(x) = \frac{q_1(x)p_2(x) + q_2(x)p_1(x)}{q_1(x)q_2(x)}, \]
and \( q_1(0)q_2(0) \neq 0 \). Moreover, the degrees of \( q_1(x)p_2(x) \) and \( q_2(x)p_1(x) \) are both less than that of \( q_1(x)q_2(x) \), and hence the same is true for their sum.

If \( b_n = \sum_{k=0}^{n} a_k \) is the sequence of partial sums of \((a_n)\), then \((b_n)\) is the convolution product of \((a_n)\) and the all-1 sequence, which is obviously C-finite.

But there are other closure properties too.

**Definition** Let \((a_n)\) and \((b_n)\) be sequences. The **Hadamard product** of these sequences is the sequence \((c_n)\) given by \( c_n = a_nb_n \).

**Proposition 4.7** The Hadamard product of two C-finite sequences is C-finite.

In particular, the sequence whose \(n\)th term is the square (or cube, or any fixed power) of the \(n\)th term of a C-finite sequence is C-finite. We had on an earlier problem sheet the exercise of showing that the squares of the Fibonacci numbers satisfy a 4-term recurrence relation; this follows from the general result.

We begin with a lemma. If \((c_n)\) is a sequence, its \(k\)th shift is the sequence with \(n\)th term \(c_{n+k}\) for \(n \geq 0\).

**Lemma 4.8** A sequence is C-finite if and only if it and all its shifts are contained in a finite-dimensional vector subspace of the space of all infinite sequences.

**Proof** Suppose that \(a = (a_n)\) satisfies a recurrence relation, which we re-write as

\[
a_{n+k} = c_1a_{n+k-1} + \cdots + c_k a_n
\]

for all \(n \geq 0\). Let \(a^{(i)}\) denote the \(i\)th shift of \(a\). The recurrence shows that

\[
a^{(k)} = c_1a^{(k-1)} + \cdots + c_k a^{(0)},
\]

so the vector subspace spanned by \(a^{(0)}, \ldots, a^{(k-1)}\) contains \(a^{(k)}\). By induction it contains all shifts \(a^{(l)}\) for \(l \geq k\). So all shifts lie in a finite-dimensional subspace.

Conversely, suppose that all shifts of \(a\) lie in a finite-dimensional subspace. Let \(a^{(l_1)}, \ldots, a^{(l_r)}\) be a maximal linearly independent set of shifts. Let \(m\) be the maximum of \(l_1, \ldots, l_r\). Then \(a^{(m+1)}\) is a linear combination of \(a^{(0)}, \ldots, a^{(m)}\), so the sequence satisfies a linear recurrence.
**Proof of the theorem** Suppose the sequences \((a_n)\) and \((b_n)\) satisfy the recurrence relations

\[
\begin{align*}
a_n &= c_1a_{n-1} + \cdots + c_ka_{n-k} \quad \text{for } n \geq k, \\
b_n &= d_1b_{n-1} + \cdots + d_lb_{n-l} \quad \text{for } n \geq l.
\end{align*}
\]

Consider first the set \(W\) of infinite arrays \((x_{ij})\) (with rows and columns indexed by the natural numbers) such that the rows satisfy the recurrence relation for the sequence \((a_n)\) and the columns satisfy the recurrence relation for the sequence \((b_n)\). If we are given the entries in the top \(l \times k\) rectangle, then all the other entries are determined: the \((a_n)\) recurrence determines the elements in the first \(l\) rows, then the \((b_n)\) recurrence determines the remaining entries. But \(W\) is closed under addition and scalar multiplication, so it is a vector space. Thus it is a vector space of dimension at most \(kl\).

Now the array given by \(x_{ij} = b_ia_j\) satisfies the recurrence relations, and so belongs to the vector space \(V\), as do all its “shifts”. Now consider the sequence \((z_n)\) which is the Hadamard product of the original sequences; that is, \(z_n = a_nb_n\). This sequence and its shifts belong to a finite-dimensional vector space, the projection of the space \(W\) onto the diagonal. So, by the characterisation of C-finite sequences, we see that \((z_n)\) is C-finite, as required. Note that it is contained in a vector space of dimension at most \(kl\), so it satisfies a recurrence of degree at most \(kl\).

**Remark** It follows that the sequence of squares of the terms in a C-finite sequence is C-finite. If \((a_n)\) has degree \(k\), then our argument above gives a bound of \(k^2\) for the degree of \((a_n^2)\). But this can be improved, to \(k(k+1)/2\). For the array in the proof is determined by the elements in the top \(k \times k\) subsquare; but this subsquare is symmetric, so we only have to look at the elements on and above the diagonal, of which there are \(k(k+1)/2\).

For example, the sequence of squares of the Fibonacci numbers satisfies the recurrence

\[
F_n^2 = 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2
\]

of degree 3.

One application of the notion of C-finiteness is that, if we guess a formula for the terms in a C-finite sequence, then we can prove it with a finite amount of calculation:
**Proposition 4.9** Let \((a_n)\) and \((b_n)\) be C-finite sequences, of degrees \(k\) and \(l\) respectively. If \(a_n = b_n\) for \(0 \leq n \leq k + l - 1\), then \(a_n = b_n\) for all \(n \in \mathbb{N}\).

**Proof** The sequence \((a_n)\) and its shifts belong to a \(k\)-dimensional vector space \(U\), and the sequence \((b_n)\) and its shifts belong to an \(l\)-dimensional vector space \(V\). Thus, \((a_n - b_n)\) and its shifts belong to \(U + V\), a vector space of dimension at most \(k + l\), so this sequence satisfies a recurrence of length at most \(k + l + 1\). So, if the first \(k + l\) terms of this sequence are zero, then the sequence is identically zero.

Here is an example, given by Doron Zeilberger in his exposition of C-finiteness. I will not give the details. The problem was originally posed by Don Knuth, and solved by Roberto Tauraso.

**Proposition 4.10** Let \(S_n\) be the region obtained from an \((n + 2) \times (n + 2)\) square by removing an \((n - 2) \times (n - 2)\) square with centre at the same point. Then

\[
a_n = 4(2F_n^2 + (-1)^n)^2.
\]

for all \(n \geq 2\).

**Proof**Zeilberger gives a general argument why the sequence \((a_n)\) is C-finite, and indeed has degree at most 5. (This uses a technique known as the transfer matrix method, which is important in statistical mechanics.) It follows from general considerations that \(4(2F_n^2 + (-1)^n)^2\) satisfies a recurrence relation of degree at most 5. So computation of ten terms of the two sequences (starting at \(n = 2\)) suffices to prove the equality. This is not straightforward, but it is a job which takes a computer only a short time.

**Exercises**

4.1. Solve the recurrence relation \(a_n = a_{n-2}\) for \(n \geq 2\). How does your solution relate to the general solution method for such sequences?

4.2. Prove directly that the squares of the Fibonacci numbers satisfy the recurrence relation

\[
F_n^2 = 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2
\]

of degree 3.

The general method does not help you find this relation directly, but it does guarantee that there will be such a relation. If it has the form

\[
F_n^2 = aF_{n-1}^2 + bF_{n-2}^2 + cF_{n-3}^2,
\]
show how the coefficients $a, b, c$ could be found by computing a few terms of the Fibonacci sequence.

4.3. Solve the recurrence relation

$$a_n = 5a_{n-1} - 6a_{n-2}$$

for $n \geq 2$, with initial conditions $a_0 = 1, a_1 = 0$. Solve it also for the initial conditions $a_0 = 0, a_1 = 1$. Explain how you can now construct the solution of the recurrence for arbitrary initial conditions $a_0 = p, a_1 = q$. 
5 Catalan objects

In chapter 3 of the notes, we defined the Catalan number $C_n$ to be the number of legal bracketings of a product of $n$ terms, and showed that

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

The Catalan numbers look at many things other than bracketings. In this chapter we will look at some of these.

The obvious ways of verifying that a class of objects is enumerated by Catalan numbers are either

(a) to verify the Catalan recurrence and initial condition; or

(b) to find a bijection to a known class of Catalan objects.

There are sometimes other less obvious ways, as we will see in the case of Dyck paths.

Where possible I have given an illustration of the five Catalan objects counted by $C_4$.

5.1 Binary trees

A binary tree has a root of degree 2; the other vertices have degree 1 or 3. So every non-root vertex is either a leaf or has two descendants, which we specify as left and right descendants.

The number of binary trees with $n$ leaves is $C_n$. Figure 4 shows the correspondence with bracketed products: the tree is a “parse tree” for the product.

![Binary trees and bracketed products](image-url)

Figure 4: Binary trees and bracketed products

60
5.2 Rooted plane trees

The number of rooted plane trees with \( n \) edges is \( C_{n+1} \). Figure 5 shows the rooted plane trees with three edges.

![Rooted plane trees](image)

Figure 5: Rooted plane trees

5.3 Dissections of polygons

An \( n \)-gon can be dissected into triangles by drawing \( n - 2 \) non-crossing diagonals. There are \( C_{n-1} \) dissections of an \( n \)-gon. Figure 6 shows dissections of a pentagon.

![Dissections of a polygon](image)

Figure 6: Dissections of a polygon

5.4 Dyck paths

A Dyck path starts at the origin and ends at \((2n, 0)\), moving at each step to the adjacent lattice point in either the north-easterly or south-easterly direction and never going below the X-axis. (An even number of steps is required since each step either increases or decreases the Y-coordinate by 1.)

Figure 7 shows the Dyck paths with \( n = 3 \).

The number of Dyck paths is \( C_{n+1} \), and of these, \( C_n \) never return to the X-axis before the end. I will indicate the proof since it illustrates another technique.

Let \( D_n \) be the number of Dyck paths, and \( E_n \) the number which never return to the axis. Now a Dyck path begins by moving from \((0, 0)\) to \((1, 1)\) and ends by...
moving from \((2n - 1, 1)\) to \((2n, 0)\); if it did not return to the axis in between, then removing these “legs” gives a shorter Dyck path. So

\[ E_n = D_{n-1}. \]

Suppose that a Dyck path first returns to the axis at \((2k, 0)\). Then it is a composite of a non-returning Dyck path of length \(2k\) with an arbitrary Dyck path of length \(2(n - k)\); so

\[ D_n = \sum_{k=1}^{n} E_k D_{n-k}. \]

Solving these simultaneous recurrences gives the result.

### 5.5 Ballot numbers

An election is held with two candidates A and B, each of whom receives exactly \(n\) votes. In how many ways can the votes be counted so that A is never behind in the count?

It is easy to match these ballot numbers with Dyck paths. For \(n = 3\), the five counts are AAABBB, AABABB, AABBAB, ABAABB, and ABABAB.

This can be described another way. In a \(2 \times n\) array, we place the numbers 1, \ldots, \(2n\) in order against the candidates who receive those votes. This gives the representations shown in Figure 8.

![Figure 8: Tableaux](image)

Note that the numbers increase along each row and down each column.
5.6 Young diagrams and tableaux

The five objects shown are known as Young tableaux; they arise in the representation theory of the symmetric group and much related combinatorics.

A Young diagram (sometimes called a Ferrers diagram) consists of \( n \) boxes arranged in left-aligned rows, the number of boxes in each row being a non-decreasing function of the row number. This is simply a graphical representation of a partition of \( n \): for each partition \( n = a_1 + a_2 + \cdots \), with \( a_1 \geq a_2 \geq \ldots \), we take \( a_1 \) boxes in the first row, \( a_2 \) in the second, and so on. Now a Young tableau is a filling of the boxes with the numbers 1, 2, \ldots, \( n \) so that each row and each column is in increasing order. You may like to invent a ballot interpretation for the number of Young tableaux belonging to a given diagram.

This combinatorics is important in describing the representation theory of the symmetric group \( S_n \), the group of all permutations of \( \{1, \ldots, n\} \). It is known that the irreducible matrix representations of \( S_n \) over the complex numbers are in one-to-one correspondence with the partitions of \( n \) (that is, to the Young diagrams); the degree of a representation is equal to the number of Young tableaux belonging to the corresponding diagram. Thus, the five Young tableaux shown in the preceding section correspond to an irreducible representation of degree 5 of the group \( S_6 \).

There is a “hook length formula” for the number of Young tableaux corresponding to a given diagram. The hook associated with a cell consists of that cell and all those to its right in the same row or below it in the same column. The hook length of a cell is the number of cells in its hook. Now the number of Young tableaux associated with the diagram is equal to \( n! \) divided by the product of the hook lengths of all its cells.

Thus for the diagram with two rows of length 3, the formula gives

\[
\frac{6!}{4 \cdot 3 \cdot 2 \cdot 3 \cdot 2 \cdot 1} = 5.
\]

5.7 Permutation patterns

In this section, we regard a permutation of the numbers 1, \ldots, \( n \) to be simply an ordering of these numbers, rather than as a bijection on the set \( \{1, \ldots, n\} \). Now we can define what we mean for a permutation to contain a smaller permutation as a pattern: there should be elements in the big permutation which come in the same relative order as those in the small one. For example, the highlighted positions show that the permutation 521643 contains the pattern 231.
How many permutations avoid (that is, do not contain) a given pattern? This is an active research area for many people including some in this department.

A permutation avoids the pattern 12 if it doesn’t have two positions $i < j$ with the $i$th entry smaller than the $j$th; that is, the only such permutation is the decreasing one $n(n − 1)\ldots321$.

For patterns of length 3, things are more interesting. For any permutation of 123, the numbers of permutations of length 3 avoiding this pattern is $6! − 1 = 5$, since just one is forbidden.

**Proposition 5.1** For any permutation $\sigma$ of 123, the number of permutations of length $n$ avoiding $\sigma$ is the Catalan number $C_{n+1}$.

It is not at all obvious that these numbers are equal! By reversing, it is clear that the numbers for 123 and 321, for 132 and 231, and for 312 and 213 are equal. Also we can “turn a permutation upside down”, and see that the numbers avoiding 132 and 312 are the same. This still leaves two distinct possibilities, and it is not straightforward to show that they give the same answer.

### 5.8 Wedderburn–Etherington numbers

What happens if we count binary trees without the left-right distinction between the two children at each node? In other words, two binary trees will count as “the same” if a sequence of reversals of subtrees above various points converts one to the other.

For example, in Figure 1, all the trees except the one in the middle are equivalent. (We can swap the first and second by reflecting the subtree above the right-hand descendant of the root; and we can swap the first and by reflecting everything.) So the count of binary trees with four leaves is 2.

It can be shown that the recurrence relation for the number $W_n$ of binary trees with this convention (the **Wedderburn–Etherington numbers** is

$$W_n = \begin{cases} \frac{1}{2} \sum_{i=1}^{n-1} W_i W_{n-i} & \text{if } n \text{ is odd,} \\ \frac{1}{2} \left( \sum_{i=1}^{n-1} W_i W_{n-i} + W_{n/2} \right) & \text{if } n \text{ is even,} \end{cases}$$

and that the generating function $w(x)$ satisfies

$$w(x) = x + \frac{1}{2} (w(x)^2 + w(x^2)).$$

64
This is much more difficult to solve. Whereas $C_n$ is roughly $4^n$ (in the sense that the limit of $C_n^{1/n}$ as $n \to \infty$ is 4), $W_n$ is roughly $2.483 \ldots^n$ in the same sense.

**Exercises**

5.1. If you didn’t do Exercises 3.8 and 3.9 before, try them now.

5.2. Verify some of the formulae for Catalan objects in the notes, either by deriving a recurrence, or by finding bijections between the objects counted.

5.3. In the analysis of Dyck paths, adopt the convention that $D_0 = 1$ and $E_0 = 0$. Prove that, if $d(x)$ and $e(x)$ are the generating functions, then

$$xd(x) = e(x), \quad d(x) = 1 + e(x)d(x).$$

Hence derive formulae for $D_n$ and $E_n$.

5.4. Use the hook length formula to derive the formula for the Catalan number $C_n$.

5.5. Prove the recurrence relation and the equation for the generating function for the Wedderburn–Etherington numbers.
6 Gaussian coefficients

The $q$-analogue of a combinatorial formula is, loosely speaking, a formula involving a parameter $q$, which tends to the original formula as $q \to 1$. However, this is much too vague to be a definition, and it turns out that there are some very specific $q$-analogues which crop up in several different fields. Most important of these are the Gaussian, or $q$-binomial, coefficients, which we discuss in this chapter. Among other places, they come up in the following areas:

- Combinatorics of vector spaces over finite fields. It turns out that there are close analogies between sets and subsets, on one hand, and vector spaces and subspaces, on the other. The counting formulae replace the binomial coefficients by their $q$-analogues.
- Lattice paths. We know that the number of lattice paths from $(0,0)$ to $(m,n)$ (using only northward and eastward steps) is $\binom{m+n}{n}$. To count these paths by the area under them, we introduce a new variable $q$ to give a generating function, and the formula becomes the $q$-analogue of the binomial coefficient.
- Non-commutative geometry. I will not give a detailed account of this, but note a couple of occurrences. The simplest to describe is the binomial theorem for indeterminates $x,y$ which satisfy $yx = qxy$; the binomial coefficients in the expansion are replaced by their $q$-analogues.

There are several further applications of the $q$-binomial coefficients, among them quantum calculus and braided categories. I will not discuss these, but there is a very accessible book on quantum calculus if you are interested: V. Kac and P. Cheung, *Quantum Calculus*, Springer, New York, 2002.

6.1 The definition

Let $n$ and $k$ be non-negative integers. The *Gaussian* or *$q$-binomial coefficient* $\binom{n}{k}_q$ is defined to be

$$\binom{n}{k}_q = \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1) \cdots (q - 1)}.$$ 

In other words, we take the formula for the binomial coefficient, and replace each factor $m$ in either numerator or denominator by $q^m - 1$. 

66
For example,
\[
\binom{4}{2}_q = \frac{(q^4 - 1)(q^3 - 1)}{(q^2 - 1)(q - 1)} = (q^2 + 1)(q^2 + q + 1),
\]
a polynomial of degree 4 in \( q \). Putting \( q = 1 \), we obtain \( 2 \cdot 3 = 6 \), which is equal to \( \binom{4}{2} \).

**Proposition 6.1**

\[
\lim_{q \to 1} \binom{n}{k}_q = \binom{n}{k}.
\]

**Proof**  By L’Hôpital’s Rule, we have
\[
\lim_{q \to 1} \frac{q^a - 1}{q^b - 1} = \lim_{q \to 1} \frac{aq^{a-1}}{bq^{b-1}} = \frac{a}{b}.
\]
Now break the fractional expression giving the Gaussian coefficient into the product of \( k \) fractions, of which the \( i \)th tends to \( (n - i + 1)/(k - i + 1) \) as \( q \to 1 \). The result follows.

### 6.2 Vector spaces over finite fields

The number of elements in a finite field is necessarily a prime power, and there is a unique field (up to isomorphism) of any given prime power order. The field with \( q \) elements is denoted by \( \text{GF}(q) \), or sometimes \( \mathbb{F}_q \).

From elementary linear algebra, we know that an \( n \)-dimensional vector space \( V \) over \( \text{GF}(q) \) has a basis \( v_1, \ldots, v_n \) such that every vector \( v \) has a unique expression as a linear combination of basis vectors:

\[
v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n, \quad c_i \in \text{GF}(q).
\]

So, as usual, \( V \) can be identified with the space \( \text{GF}(q)^n \) of all \( n \)-tuples of elements of \( \text{GF}(q) \), with coordinatewise operations. In particular, we see:

**Proposition 6.2**  If \( V \) is an \( n \)-dimensional vector space over \( \text{GF}(q) \), then \( |V| = q^n \).

Now the connection with Gaussian coefficients is the following:
**Theorem 6.3** The number of $k$-dimensional subspaces of an $n$-dimensional vector space over $GF(q)$ is $\binom{n}{k}_q$.

**Proof** We specify a $k$-dimensional subspace by giving a basis, a linearly independent $k$-tuple $(v_1, \ldots, v_k)$ of vectors in $V$. The number of choices

- for $v_1$ is $q^n - 1$ (any vector except 0);
- for $v_2$ is $q^n - q$ (any vector except one of the $q$ multiples of $v_1$);
- for $v_3$ is $q^n - q^2$ (any vector except one of the $q^2$ linear combinations of $v_1$ and $v_2$);

and so on; the total number of choices is

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1}).$$

However, we have over-counted, since a given $k$-dimensional subspace has many bases. How many? The number is found from the same formula using $k$ instead of $n$, since we are working within a $k$-dimensional space.

So the number of $k$-dimensional subspaces is

$$\frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1})}.$$ 

Cancelling powers of $q$, this reduces to the formula for the Gaussian coefficient.

This result has another counting interpretation. Recall that a $k \times n$ matrix $A$ over a field $F$ is in *reduced echelon form* if

- if a row is not identically zero, the first non-zero element in it is 1;
- the “leading ones” in the non-zero rows occur further to the right as we go down the matrix;
- all the other elements in the column of a leading one are zero.

Now standard linear algebra shows that any matrix can be put into reduced echelon form by elementary row operations, which do not change the row space of the matrix. So any $k$-dimensional subspace of $F^n$ has a basis whose vectors are the rows of a matrix in reduced echelon form. Moreover, it is easy to see that the reduced echelon basis of a given subspace is unique. So:
Proposition 6.4 The number of $k \times n$ matrices over $\text{GF}(q)$ which have no zero rows and are in reduced echelon form is $\begin{bmatrix} n \\ k \end{bmatrix}_q$.

The definition of “reduced echelon” does not depend on properties of fields; any alphabet containing distinguished elements called 0 and 1 will do. So we have a counting interpretation of the Gaussian coefficients for arbitrary positive integers $q > 1$.

Example For $n = 4$ and $k = 2$, the matrices in reduced echelon form are shown. $*$ can be any element of the field. Beside each matrix is the number of matrices of this form over $\text{GF}(q)$.

\[
\begin{pmatrix}
1 & 0 & * & * \\
0 & 1 & * & *
\end{pmatrix} \quad q^4
\]
\[
\begin{pmatrix}
1 & * & 0 & * \\
0 & 0 & 1 & *
\end{pmatrix} \quad q^3
\]
\[
\begin{pmatrix}
1 & * & * & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \quad q^2
\]
\[
\begin{pmatrix}
0 & 1 & 0 & * \\
0 & 0 & 1 & *
\end{pmatrix} \quad q^2
\]
\[
\begin{pmatrix}
0 & 1 & * & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \quad q
\]
\[
\begin{pmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \quad 1
\]

We find, as before, that $\begin{bmatrix} 4 \\ 2 \end{bmatrix}_q = (q^2 + 1)(q^2 + q + 1)$.

6.3 Relations between Gaussian coefficients

The $q$-binomial coefficients satisfy an analogue of the recurrence relation for binomial coefficients.

Proposition 6.5 $\begin{bmatrix} n \\ 0 \end{bmatrix}_q = \begin{bmatrix} n \\ n \end{bmatrix}_q = 1$, $\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n - 1 \\ k - 1 \end{bmatrix}_q + q^k \begin{bmatrix} n - 1 \\ k \end{bmatrix}_q$ for $0 < k < n$.
Proof. This comes straight from the definition. Suppose that $0 < k < n$. Then

\[
\begin{align*}
\left[\begin{array}{c} n \\ k \end{array}\right]_q - \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q &= \left(\frac{q^n-1}{q^k-1} \right) \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q \\
&= q^k \left(\frac{q^{n-k}-1}{q^k-1} \right) \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q \\
&= q^k \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q.
\end{align*}
\]

The array of Gaussian coefficients has the same symmetry as that of binomial coefficients

**Proposition 6.6**

\[
\left[\begin{array}{c} n \\ k \end{array}\right]_q = \left[\begin{array}{c} n \\ n-k \end{array}\right]_q.
\]

The proof is an exercise from the formula. Note that, in the vector space interpretation, we have a different way to see this. Given an $n$-dimensional vector space $V$ over $\text{GF}(q)$, it has a dual space $V^*$, the space of linear maps from $V$ to $\text{GF}(q)$. Now any $k$-dimensional subspace of $V$ has an $(n-k)$-dimensional annihilator in $V^*$, and the correspondence between $k$-dimensional subspaces of $V$ and $(n-k)$-dimensional subspaces of $V^*$ is bijective.

From this we can deduce another recurrence relation.

**Proposition 6.7**

\[
\left[\begin{array}{c} n \\ 0 \end{array}\right]_q = \left[\begin{array}{c} n \\ n \end{array}\right]_q = 1, \quad \left[\begin{array}{c} n \\ k \end{array}\right]_q = q^{n-k} \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q + \left[\begin{array}{c} n-1 \\ k \end{array}\right]_q, \quad \text{for } 0 < k < n.
\]

Proof

\[
\begin{align*}
\left[\begin{array}{c} n \\ k \end{array}\right]_q &= \left[\begin{array}{c} n \\ n-k \end{array}\right]_q \\
&= \left[\begin{array}{c} n-1 \\ n-k-1 \end{array}\right]_q + q^{n-k} \left[\begin{array}{c} n-1 \\ n-k \end{array}\right]_q \\
&= \left[\begin{array}{c} n-1 \\ k \end{array}\right]_q + q^{n-k} \left[\begin{array}{c} n-1 \\ k-1 \end{array}\right]_q.
\end{align*}
\]

We come now to the $q$-analogue of the binomial theorem, which states the following.
Theorem 6.8 For a positive integer $n$, a real number $q \neq 1$, and an indeterminate $x$, we have

$$\prod_{i=1}^{n}(1 + q^{i-1}x) = \sum_{k=0}^{n} q^{k(n-1)/2} x^{k} \left[ \begin{array}{c} n \\ k \end{array} \right]_{q}.$$ 

Proof The proof is by induction on $n$; starting the induction at $n = 1$ is trivial. Suppose that the result is true for $n - 1$. For the inductive step, we must compute

$$\left( \sum_{k=0}^{n-1} q^{k(n-1)/2} x^{k} \left[ \begin{array}{c} n - 1 \\ k \end{array} \right]_{q} \right) (1 + q^{n-1}x).$$

The coefficient of $x^{k}$ in this expression is

$$q^{k(n-1)/2} \left[ \begin{array}{c} n - 1 \\ k \end{array} \right]_{q} + q^{(k-1)(n-2)+n-1} \left[ \begin{array}{c} n - 1 \\ k - 1 \end{array} \right]_{q}$$

$$= q^{k(n-1)/2} \left( \left[ \begin{array}{c} n - 1 \\ k \end{array} \right]_{q} + q^{n-k} \left[ \begin{array}{c} n - 1 \\ k - 1 \end{array} \right]_{q} \right)$$

$$= q^{k(n-1)/2} \left[ \begin{array}{c} n \\ k \end{array} \right]_{q}$$

by Proposition 6.7.

6.4 Lattice paths

We consider lattice paths from the origin to the point $(m,n)$, where $m$ and $n$ are non-negative integers, and the allowable steps are east and north. The number of paths is $\binom{m+n}{m}$, since we have to take $m+n$ steps altogether, of which $m$ must be easterly and $n$ northerly.

For each such path $p$, there is a certain area $A(p)$ enclosed between the path, the X-axis, and the line $x = m$. Figure 9 shows the six paths for $m = n = 2$ and the area enclosed in each case.

If we take a generating function in the variable $q$ for these areas, that is, a path with area $A$ contributes $q^{A}$ to the sum, we obtain

$$1 + q + 2q^{2} + q^{3} + q^{4} = \left[ \begin{array}{c} 4 \\ 2 \end{array} \right]_{q}.$$ 

This is quite general:
<table>
<thead>
<tr>
<th>Path</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image0" alt="Path 0" /></td>
<td>0</td>
</tr>
<tr>
<td><img src="image1" alt="Path 1" /></td>
<td>1</td>
</tr>
<tr>
<td><img src="image2" alt="Path 2" /></td>
<td>2</td>
</tr>
<tr>
<td><img src="image3" alt="Path 3" /></td>
<td>2</td>
</tr>
<tr>
<td><img src="image4" alt="Path 4" /></td>
<td>3</td>
</tr>
<tr>
<td><img src="image5" alt="Path 5" /></td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 9: Lattice paths
**Theorem 6.9** Let $\mathcal{P}$ be the set of lattice paths from $(0,0)$ to $(m,n)$ using northerly and easterly steps only. For $p \in \mathcal{P}$, let $A(p)$ be the area enclosed by $p$, the X-axis and the line $x=m$. Then

$$\sum_{p \in \mathcal{P}} q^A(p) = \binom{m+n}{m}_q.$$

**Proof** Call the left-hand side $F(m,n)$. It is clear that $F(0,n) = F(m,0) = 0$. Now consider $F(m,n)$. There are two cases:

- If the last step on the path $p$ is northerly, then it is a path from $(0,0)$ to $(m,n-1)$ followed by a northerly step, and the last step doesn’t change the area.
- If the last step is easterly, then $p$ is a path from $(0,0)$ to $(m-1,n)$ followed by an easterly step, which adds $n$ to the area.

So

$$F(m,n) = F(m,n-1) + q^n F(m-1,n).$$

Now a simple induction using **Proposition 6.5** gives the result.

**Exercises**

6.1. Show that the Gaussian coefficient $\binom{n}{k}_q$ is a polynomial in $q$ with degree $k(n-k)$.

6.2. Use the interpretation in terms of matrices in reduced echelon form to show the recurrence relation of **Proposition 6.5**.

6.3. Show that the polynomial in the first problem has coefficients which are symmetric, that is, if

$$\binom{n}{k}_q = \sum_{i=0}^{k(n-k)} a_i q^i,$$

then $a_i = a_{k(n-k)-i}$.

6.4. A matrix is said to be in *echelon form* if it satisfies the first two of the three conditions for reduced echelon form.

Show that, if $q$ is an integer greater than 2, the right-hand side of the $q$-binomial theorem with $x=1$ counts the number of $n \times n$ matrices in echelon form.
6.5. Let \( x \) and \( y \) be elements of an algebra, which satisfy \( yx = qxy \). Prove that

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} q^{n-k} y^k.
\]

6.6. Let \( X \) be the set of 1-dimensional subspaces of an \( n \)-dimensional vector space \( V \) over \( \text{GF}(q) \), where \( n \geq 3 \). For any 2-dimensional subspace \( W \) of \( V \), let \( B(W) \) be the set of 1-dimensional subspaces contained in \( W \), and let \( \mathcal{B} \) be the collection of all blocks or subsets of \( X \) arising in this way. Prove that \((X, \mathcal{B})\) is a \((2, q + 1, (q^n - 1)/(q - 1))\) Steiner system. (This system is known as the \((n - 1)\)-dimensional projective space over \( \text{GF}(q) \).)
7 Möbius inversion

In a certain school, 73% of the pupils take part in acting, 79% play billiards, and 71% play cricket. The percentage who do acting and billiards is 52%, acting and cricket 47%, billiards and cricket 41%, and 15% do all three. What proportion of the pupils do none of these three activities?

7.1 The Principle of Inclusion and Exclusion

This is a classic example of the Principle of Inclusion and Exclusion, or PIE for short. The answer is

\[ 100 - 73 - 79 - 71 + 52 + 47 + 41 - 15 = 2, \]

so 2% of the pupils do none of the activities. Here is the general statement.

**Theorem 7.1 (Principle of Inclusion and Exclusion)** Let \( A_1, \ldots, A_n \) be subsets of a finite set \( X \). For \( I \subseteq \{1, \ldots, n\} \), let

\[ A_I = \bigcap_{i \in I} A_i, \]

with the convention that \( A_{\emptyset} = X \). Then

\[ \left| X \setminus \bigcup_{i \in \{1, \ldots, n\}} A_i \right| = \sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|} |A_I|. \]

**Proof** The sum on the right-hand side can be regarded as being the sum, over all elements \( x \in X \), of a contribution from \( x \) which is the sum of appropriate signs for the subsets \( A_I \) which contain \( x \).

If \( x \) lies in none of the sets \( A_i \), then the only contribution from \( x \) corresponds to the term \( I = \emptyset \) in the sum, and so \( x \) contributes 1.

Let \( K = \{i \in \{1, \ldots, n\} : x \in A_i\} \), and suppose that \( K \neq \emptyset \). Let \( |K| = k \). Then the sets \( A_I \) containing \( x \) are just those with \( I \subseteq K \), and the contribution of \( x \) is

\[ \sum_{I \subseteq K} (-1)^{|I|} = \sum_{i=0}^{k} (-1)^i = (1 - 1)^k = 0, \]

by the Binomial Theorem.

So the sum counts precisely the points lying in no set \( A_i \), as claimed.
A special case of the theorem is often useful.

**Theorem 7.2** Let $A_1, \ldots, A_n$ be subsets of a finite set $X$. With notation as in the preceding theorem, suppose that, if $|I| = i$, then $|A_I| = f_i$. Then

$$
|X \setminus \bigcup_{i \in \{1, \ldots, n\}} A_i| = \sum_{i \in \{1, \ldots, n\}} (-1)^i \binom{n}{i} f_i.
$$

In this section, we will look at some consequences of PIE, and then see a wide generalisation of it known as M"obius inversion, which has connections with number theory and also with the $q$-binomial theorem which we meet in the last section of the notes.

### 7.2 Applications

I will give three applications here.

**Surjections**

The first involves counting functions. It is well known that the number of functions from an $m$-set to an $n$-set is $n^m$, since we get to choose the value of the function at each point. Also, the number of injective functions is $n(n-1) \cdots (n-m+1)$, since there are $n$ choices for the value of the function at the first point of its domain, $n-1$ at the second, $\ldots$, and $n-m+1$ at the last. (In particular, this formula gives 0 if $m > n$, as it should.) How many surjective functions are there?

**Theorem 7.3** The number of surjective functions from an $m$-set to an $n$-set is

$$
\sum_{i=0}^{n-1} (-1)^i \binom{n}{i} (n-i)^m.
$$

**Proof** We apply PIE to the set $X$ of all functions from an $m$-set to an $n$-set. Then $|X| = n^m$. Let $A_i$ be the set of functions which do not take the value $i$. Then $A_I$ is the set of functions which take no values in the set $I$, for $I \subseteq \{1, \ldots, n\}$; so $|A_I| = (n-i)^m$, where $|I| = i$. By the second form of PIE, the number of functions lying in none of the sets $A_i$ is

$$
\sum_{i=0}^{n-1} (-1)^i \binom{n}{i} (n-i)^m.
$$

But these are precisely the surjective functions!
Derangements

The second question asks the following.

$n$ letters are waiting to be put into $n$ addressed envelopes. A new secretary inserts the letters randomly into envelopes. What is the probability that no letter is in its correct envelope?

Most people’s intuition is that the answer is close to 0, or maybe close to 1. In fact, it is close to $e = 0.367879441\ldots$. We will derive an exact formula and then show that this limit is correct. A permutation with no fixed points is called a derangement.

**Theorem 7.4** The number of derangements of $\{1, \ldots, n\}$ is

$$d_n = n! \left( \sum_{i=0}^{n} \frac{(-1)^i}{i!} \right)$$

**Proof** We apply PIE to the set $X$ of all permutations of $\{1, \ldots, n\}$, with $|X| = n!$. Let $A_j$ be the set of permutations which fix the point $j$. Then, for $I \subseteq \{1, \ldots, n\}$, $A_I$ is the set of permutations fixing every point in $I$, and so permuting the remaining points; thus, if $|I| = i$, then $|A_I| = (n-i)!$. So the number of permutations in no set $A_i$ is

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! = n! \left( \sum_{i=0}^{n} \frac{(-1)^i}{i!} \right).$$

These permutations are precisely the derangements!

Now the probability that a given permutation is a derangement is $d_n/n!$, which from above is the sum of the first $n+1$ terms in the Taylor series for $e^x$ at $x = -1$. So the limit is indeed $e^{-1}$, as claimed. But in fact more is true:

**Corollary 7.5** For $n \geq 1$, $d_n$ is the nearest integer to $n!/e$.

**Proof** For $i \leq n$, we have $n!/i! = n(n-1)\cdots(i+1)$, an integer. For $i > n$, this fraction is $1/(n+1)\cdots i$, which is a decreasing function of $n$. Now the sum of an alternating series whose terms are decreasing in modulus is smaller than the modulus of the first term. So the difference between $n!/e$ and the sum of the first $n+1$ terms is

$$|n!/e - d_n| < 1/(n+1) \leq 1/2,$$

so indeed $d_n$ is the nearest integer to $n!/e$. (Note, however, that this is a very inefficient way to calculate $d_n$, since we would need to know $1/e$ to very high accuracy.)
There are several different ways to calculate $d_n$. One is by means of a recurrence relation:

$$d_n = (n-1)(d_{n-1} + d_{n-2}) \text{ for } n \geq 2.$$ 

This, together with the initial values $d_0 = 1$ and $d_1 = 0$, determines $d_n$ for all $n$. (But our general theory of C-finite series does not apply here since the coefficients in the recurrence are not constants!)

To see this, take a derangement $\pi$ of $\{1, \ldots, n\}$, and suppose that $\pi$ maps $n$ to $i$, where $i \neq n$. There are two cases:

(a) $\pi$ also maps $i$ to $n$. Then $\pi$ is the product of the transposition $(i, n)$ with a derangement on the remaining $n-2$ points, so there are $d_{n-2}$ derangements of this shape.

(b) $\pi$ maps $j$ to $n$, where $j \neq i$. Then the permutation $\pi'$ obtained by shortcircuiting $n$ (that is, $\pi'$ maps $j$ to $i$ and agrees with $\pi$ elsewhere) is a derangement of $\{1, \ldots, n-1\}$. There are $d_{n-1}$ derangements of this shape.

Now there are $n-1$ choices of $i$, and for each choice there are $d_{n-1} + d_{n-2}$ derangements mapping $n$ to $i$; so the result follows.

This relation allows the derangement numbers to be calculated, but is not so useful for finding a formula. However, it allows us to prove the shorter recurrence

$$d_n = nd_{n-1} + (-1)^n$$

by induction on $n$. (For the inductive step, assume that $d_{n-1} = (n-1)d_{n-2} + (-1)^{n-1}$. Then

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2}$$
$$= (n-1)d_{n-1} + d_{n-1} - (-1)^{n-1}$$
$$= nd_{n-1} + (-1)^n.$$ 

From this we can deduce our formula for $d_n$ by another easy induction: assuming that

$$d_{n-1} = (n-1)! \sum_{i=0}^{n-1} \frac{(-1)^i}{i!},$$

we have

$$d_n = n! \left( \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} + \frac{(-1)^n}{n!} \right),$$

as required.
Graph colourings

Given a set of $q$ colours, in how many different ways can we colour the vertices of a graph properly, that is, so that any two vertices joined by an edge must get different colours?

For example, for the graph

there are $q$ choices for the colour of the top vertex, $q - 1$ for the bottom vertex (it must be different from the top), $q - 2$ for the left vertex, and again $q - 2$ for the right vertex: altogether $q(q - 1)(q - 2)^2$ colourings.

In general the answer is a monic polynomial in $q$ with some very nice properties. The usual proof of this is graph-theoretic, involving operations of deletion and contraction of graphs. Here is a proof using PIE.

Let $G$ be a given graph with vertex set $V$ and edge set $E$. We let $X$ be the set of all colourings of the vertices of $G$, proper or improper; then $|X| = q^n$, where $n$ is the number of vertices. For every edge $e$, let $A_e$ be the set of colourings for which $e$ is improperly coloured (that is, its endpoints have the same colour).

Given a subset $I$ of the edge set $E$, how many colourings have the property that the edges in $I$ (and maybe others) are improperly coloured? This means that the vertices on any edge of $I$ have the same colour; so, if we consider the graph $G_I$ with vertex set $V$ and edge set $I$, then all the connected components of $G_I$ have the same colour. So there are $q^{c(I)}$ colourings, where $c(I)$ is the number of connected components of the graph $(V, I)$; that is, $|A_I| = q^{c(I)}$.

By PIE, the number of proper colourings (those with no improper edges) is

$$
\sum_{I \subseteq E} (-1)^{|I|}q^{c(I)}.
$$

This is clearly a polynomial in $q$. Moreover, the largest power of $q$ is $q^n$, occurring just when $I = \emptyset$; so it is a monic polynomial of degree $n$.

This polynomial is the chromatic polynomial of the graph $G$. The formula we have obtained for it is important in statistical mechanics, where it is known as the cluster expansion for the Potts model.
7.3 The Möbius function of a poset

In this section we will see a remarkable generalisation of PIE. The context is partially ordered sets.

A poset, or partially ordered set, consists of a set $A$ with a relation $\leq$ on $A$ which is

(a) reflexive: $a \leq a$ for all $a \in A$;
(b) antisymmetric: $a \leq b$ and $B \leq a$ imply $a = b$, for all $a, b \in A$;
(c) transitive: $a \leq b$ and $b \leq c$ imply $a \leq c$, for all $a, b, c \in A$.

An important combinatorial example consists of the case where $A$ is the set of all subsets of a finite set $S$, and $a \leq b$ means that $a$ is a subset of $b$. It turns out that the Inclusion-Exclusion principle can be formulated in terms of this poset, and then generalised so as to apply to any poset.

We begin with an observation which will not be proved here.

Theorem 7.6 Let $P = (A, \leq)$ be a finite poset with $n$ elements. Then we can label the elements of $A$ as $a_1, a_2, \ldots, a_n$ such that, if $a_i \leq a_j$, then $i \leq j$.

This is sometimes stated “Every poset has a linear extension”. The analogous result for infinite posets requires a weak form of the Axiom of Choice in its proof.

Now let $P = (A, \leq)$ be a poset. We define the incidence algebra of $P$ as follows: the elements are all functions $f : A \times A \to \mathbb{R}$ such that $f(a, b) = 0$ unless $a \leq b$. Addition and scalar multiplication are defined in the obvious way, and multiplication by the rule

$$fg(a, b) = \begin{cases} \sum_{a \leq c \leq b} f(a, c)g(c, b) & \text{if } a \leq b, \\ 0 & \text{if } a \nleq b. \end{cases}$$

If we number the elements of $A$ as in the preceding theorem, then we can represent a function from $A \times A$ to $\mathbb{R}$ by an $n \times n$ matrix; the definition of the incidence algebra shows that any function which lies in the algebra is upper triangular. The multiplication in the algebra is then just matrix multiplication, so the incidence algebra is a subalgebra of the algebra of all $n \times n$ real matrices.

We now define three particular elements of the incidence algebra.
(a) ι is the identity function:

\[ ι(a, b) = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b \end{cases} \]

represented by the identity matrix.

(b) ζ is the zeta function:

\[ ζ(a, b) = \begin{cases} 1 & \text{if } a \leq b, \\ 0 & \text{if } a \not\leq b. \end{cases} \]

(c) μ, the Möbius function, is the inverse of the zeta function: \( \mu ζ = ζ μ = ι \).

The zeta function is represented by an upper unitriangular matrix (that is, an upper triangular matrix with 1 on the diagonal) with integer entries; so it is invertible, and its inverse, the Möbius function, is also represented by an upper unitriangular matrix with integer entries. Its definition shows that, if \( a < b \), then

\[ \sum_{a \leq c \leq b} μ(a, c) = 0, \]

so that

\[ μ(a, b) = - \sum_{a \leq c < b} μ(a, c). \]

This gives a recursive method for calculating the Möbius function, as we will see.

From the definition, we immediately have the Möbius inversion formula:

**Theorem 7.7** Let \( P \) be a poset with Möbius function \( μ \). Then the following are equivalent:

(a) \( g(a, b) = \sum_{a \leq c \leq b} f(a, c) \) for all \( a \leq b \);

(b) \( f(a, b) = \sum_{a \leq c \leq b} g(a, c) μ(c, b) \) for all \( a \leq b \).

We tacitly assume that the functions \( f \) and \( g \) belong to the incidence algebra, that is, \( f(a, b) = g(a, b) = 0 \) unless \( a \leq b \). In fact, we never refer to any values of \( f(a, b) \) or \( g(a, b) \) except when \( a \leq b \).
7.4 Some examples

The preceding remark shows that the value of $\mu(a, b)$ depends only on the structure of the interval $[a, b] = \{c : a \leq c \leq b\}$.

Many important posets have a least element (which is usually called 0) and a “homogeneity property”: for any $a, b$ with $a \leq b$, there is an element $c$ such that the interval $[a, b]$ is isomorphic (as a partially ordered set) to the interval $[0, c]$. In a poset with this property, $\mu(a, b) = \mu(0, c)$, and we can regard the Möbius function as a one-variable function.

A chain

A chain, or linear order, is a poset in which every pair of elements is comparable. Any finite chain is isomorphic (as partially ordered set) to $\{0, 1, \ldots, n-1\}$ with the usual order. Its Möbius function is given by

$$\mu(a, b) = \begin{cases} 1 & \text{if } b = a, \\ -1 & \text{if } b = a + 1, \\ 0 & \text{otherwise.} \end{cases}$$

This follows immediately from the recursive method of computing $\mu$.

In this case, any interval $[a, b]$ is isomorphic to the interval $[0, b - a]$, so it would have sufficed to take $a = 0$; but the general case is simple enough.

Direct product

The direct product of posets $P_1 = (A_1, \leq_1)$ and $P_2 = (A_2, \leq_2)$ has set $A_1 \times A_2$ (Cartesian product), and

$$(a_1, a_2) \leq (b_1, b_2) \iff a_1 \leq_1 b_1 \text{ and } a_2 \leq_2 b_2.$$ 

It is easily checked that

$$\mu((a_1, a_2), (b_1, b_2)) = \mu_1(a_1, b_1)\mu_2(a_2, b_2),$$

where $\mu_1$ and $\mu_2$ are the Möbius functions of $P_1$ and $P_2$. This extends in a straightforward way to the direct product of any finite number of posets.

Subsets of a set

The poset of all subsets of $\{1, 2, \ldots, n\}$ can be represented as the direct product of $n$ copies of the 2-element chain $\{0, 1\}$; the subset $a$ is identified with the $n$-tuple $(a_1, \ldots, a_n)$, where

$$a_i = \begin{cases} 1 & \text{if } i \in a, \\ 0 & \text{if } i \notin a. \end{cases}$$
It follows from the two preceding paragraphs that the Möbius function is
\[
\mu(a, b) = \begin{cases} 
(-1)^{|b \setminus a|} & \text{if } a \subseteq b, \\
0 & \text{if } a \nsubseteq b.
\end{cases}
\]

In this case, if \( a \subseteq b \), then \([a, b]\) is isomorphic to \([\emptyset, b \setminus a]\), and we see the homogeneity property in action. So the following are equivalent:

(a) \( f(a) = \sum_{b \leq a} g(b); \)

(b) \( g(a) = \sum_{b \leq a} f(b)(-1)^{|a \setminus b|}. \)

With a little rearrangement, this is a generalisation of PIE, with cardinality replaced by an arbitrary function. The rearrangement proceeds in two steps. First, we “reverse the order” to see that the following are equivalent:

(a) \( f(a) = \sum_{b \geq a} g(b); \)

(b) \( g(a) = \sum_{b \geq a} f(b)(-1)^{|a \setminus b|}. \)

Now let \( A_1, \ldots, A_n \) be subsets of \( X \). For \( a \subseteq \{1, \ldots, n\} \), let \( f(a) \) be the number of elements of \( X \) which lie in all the sets \( A_i \) for \( i \in a \), and let \( g(a) \) be the number of elements lying in all the sets \( A_i \) for \( i \in a \) and no others. Clearly equation (a) above holds, and therefore (b) also holds. Putting \( a = \emptyset \), we recover the original form of PIE.

**The classical Möbius function**

The classical Möbius function from number theory is defined on the natural numbers; the partial order is given by \( a \leq b \) if \( a \) divides \( b \). Although this partial order is infinite, all intervals are finite, and it has the homogeneity property: if \( a \mid b \), then the interval \([a, b]\) is isomorphic to \([1, b/a]\).

This poset is isomorphic to the product of chains, one for each prime power. We have
\[
\mu(p^a, p^b) = \begin{cases} 
1 & \text{if } b = a, \\
-1 & \text{if } b = a + 1, \\
0 & \text{otherwise}.
\end{cases}
\]
Hence we have the general formula:

\[ \mu(m, n) = \begin{cases} (-1)^d & \text{if } m \mid n \text{ and } n/m \text{ is a product of } d \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases} \]

In particular, \( \mu(1, n) \) is the number-theorists’ Möbius function, which they write as \( \mu(n) \). We have the classical Möbius inversion formula, the equivalence of the following functions \( f, g \) on \( \mathbb{N} \):

(a) \( g(n) = \sum_{m \mid n} f(m) \);

(b) \( f(n) = \sum_{m \mid n} g(m) \mu(n/m) \).

Here is an application. Back in Section 3, we found a formula for the number \( a_n \) of monic irreducible polynomials of degree \( n \) over a field with \( q \) elements:

\[ q^n = \sum_{i \mid n} ia_i. \]

By Möbius inversion, we get

\[ na_n = \sum_{i \mid n} q^i \mu(n/i), \]

so

\[ a_n = \frac{1}{n} \sum_{i \mid n} q^i \mu(n/i). \]

From this we see that \( na_n \geq q^n - q^{n/2} - q^{n/3} - \cdots - q^n - nq^{n/2} \), so \( a_n \neq 0 \); that is, as we claimed, there does exist an irreducible polynomial of degree \( n \) over any finite field. [The crude inequality above covers all cases except that when \( q = n = 2 \), which can be done directly.]

**Subspaces of a vector space**

For our final example, let \( A \) be the set of all subspaces of an \( n \)-dimensional vector space over a field of order \( q \). If \( V \leq W \), the structure of the interval \([V, W]\) depends only on \( \dim(W) - \dim(V) \), and so is isomorphic to \([0, W/V]\).

Recall the \( q \)-binomial theorem:

\[ \prod_{i=1}^{n} (1 + q^{i-1}z) = \sum_{k=0}^{n} q^{k(k-1)/2} z^{\left[\frac{n}{k}\right]} \binom{n}{k}_q. \]
Putting \( z = -1 \), the left-hand side becomes 0; then we have
\[
(-1)^n q^{n(n-1)/2} = - \sum_{k=0}^{n-1} (-1)^k q^{k(k-1)/2} \binom{n}{k}.\]
This shows, recursively, that if \( \dim(V) = n \), then \( \mu(\{0\}, V) = (-1)^n q^{n(n-1)/2} \).

**Exercises**

7.1. Show that the recurrence relation
\[
f(0) = 1, \quad f(n) = nf(n - 1) + g(n) \text{ for } n > 0
\]
has solution
\[
f(n) = n! \left( \sum_{i=0}^{n} \frac{g(i)}{i!} \right).
\]

7.2. Let \( F(n) \) be the number of “words” (strings of letters) that can be made from \( n \) distinct letters (including the empty word). Show that
\[
F(0) = 1, \quad F(n) = nF(n - 1) + 1 \text{ for } n > 0,
\]
and hence show that \( F(n) \) is the nearest integer to \( n! e \) for \( n > 1 \).

7.3. Show that the exponential generating function of the derangement numbers is given by
\[
\sum_{n \geq 0} \frac{d_n x^n}{n!} = \frac{\exp(-x)}{1 - x}.
\]
Give an alternative proof of this formula as follows. An arbitrary permutation of \( \{1, \ldots, n\} \) can be specified by choosing a subset of the \( n \) points to be fixed by the permutation, and choosing a derangement on the remaining points; so
\[
n! = \sum_{k=0}^{n} \binom{n}{k} d_{n-k}.
\]

7.4. Let \( G \) be a tree (a connected graph with no cycles) on \( n \) vertices. Show that any set of \( i \) edges form a subgraph with \( n - i \) connected components. Hence show that the chromatic polynomial of \( G \) is \( q(q - 1)^{n-1} \), independent of the choice of tree (in other words, it is possible for different graphs to have the same chromatic polynomial).
Can you prove this directly?
7.5. Use similar reasoning to find the chromatic polynomial of a cycle.

7.6. A simple partially ordered set consists of a top and bottom element and \( n \) incomparable elements between:

![Diagram of a partially ordered set with a top, bottom, and \( n \) incomparable elements]

Calculate its Möbius function.

7.7. There is a partial order on the set of all partitions of \( \{1, \ldots, n\} \), defined as follows: if \( a \) and \( b \) are partitions, say that \( a \) refines \( b \) if every part of \( b \) is a union of parts of \( a \).

Can you find the Möbius function of this partial order?

7.8. Prove the following “approximate version” of Inclusion-Exclusion:

Let \( A_1, \ldots, A_n, A'_1, \ldots, A'_n \) be subsets of a set \( X \). For \( I \subseteq \mathbb{N} = \{1, \ldots, n\} \), let

\[
a_I = \left| \bigcap_{i \in I} A_i \right|, \quad d_I = \left| \bigcap_{i \in I} A'_i \right|.
\]

If \( a_I = d_I \) for all proper subsets \( I \) of \( \mathbb{N} \), then \( |a_N - d_N| \leq |X|/2^{n-1} \).

7.9. The following problem, based on the children’s game “Screaming Toes”, was suggested to me by Julian Gilbey.

\( n \) people stand in a circle. Each player looks down at someone else’s feet (i.e., not at their own feet). At a given signal, everyone looks up from the feet to the eyes of the person they were looking at. If two people make eye contact, they scream. What is the probability of at least one pair of people screaming?

Prove that the required probability is

\[
\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{(-1)^{k-1} (n)_{2k}}{(n-1)^{2k} 2^k k!}.
\]

where \((n)_j = n(n - 1) \cdots (n - j + 1)\).
8 Number partitions

In this section we are going to count the number of partitions of a natural number. This is related to, but not the same as, the number of partitions of a set, which we’ll consider in the next section.

8.1 Definition and generating function

A partition of a natural number \( n \) is a non-increasing sequence of positive integers whose sum is \( n \). So, for example, there are seven partitions of 5:

\[
5 = 4 + 1 = 3 + 2 = 3 + 1 + 1 = 2 + 2 + 1 = 2 + 1 + 1 + 1 = 1 + 1 + 1 + 1 + 1.
\]

We let \( p(n) \) denote the number of partitions of \( n \). Note that \( p(0) = 1 \).

Another interpretation is that \( p(n) \) is the number of conjugacy classes in the symmetric group \( S_n \). This depends on two facts about permutations: first, any permutation can be written as a product of disjoint cycles; and second, two permutations are conjugate in \( S_n \) if and only if they have the same cycle structure, where the cycle structure of a permutation in \( S_n \) is the partition of \( n \) listing the lengths of its cycles in non-increasing order. Thus, \( S_5 \) has seven conjugacy classes.

**Proposition 8.1** The generating function for the sequence \((p(0), p(1), \ldots)\) is

\[
\sum_{n \geq 0} p(n)x^n = \prod_{k \geq 1} (1 - k^2)^{-1}.
\]

**Proof** The right-hand side is \( \prod_{k \geq 1} (1 + x^k + x^{2k} + \cdots) \).

Given any partition of \( n \), having, say, \( a_1 \) parts of length 1, \( a_2 \) parts of length 2, and so on, gives rise to a term in \( x^n \) in this infinite product, taking \( x^{a_1} \) from the first bracket, \( x^{2a_2} \) from the second, and so on, since \( a_1 + 2a_2 + \cdots = n \). Conversely, every term in \( x^n \) arises in a unique way from such a partition.

We know that obtaining a recurrence relation for a sequence is equivalent to computing its multiplicative inverse in the ring of formal power series. Now the inverse of the generating function for the partition numbers is \( \prod_{k \geq 0} (1 - x^k) \), and so our next job is to compute this.
8.2 Euler’s Pentagonal Numbers Theorem

**Proposition 8.2** The coefficient of $x^n$ in $\prod_{k \geq 0} (1 - x^k)$ is equal to the number of partitions of $n$ into an even number of distinct parts, minus the number of partitions of $n$ into an odd number of distinct parts.

**Proof** This is similar to the previous result. Given a partition of $n$ into $r$ distinct parts $c_1, c_2, \ldots, c_r$, we obtain a term in $x^n$ by choosing $x^{c_i}$ from the bracket $(1 - x^{c_i})$ for $i = 1, \ldots, r$, and 1 from all the other brackets. The coefficient of this term is $(-1)^r$, so it counts as positive if $r$ is even and negative if $r$ is odd.

For example, the partitions of 8 into distinct parts are

$$8 = 7 + 1 = 6 + 2 = 5 + 3 = 5 + 2 + 1 = 4 + 3 + 1,$$

with three even and three odd, so the coefficient of $x^8$ in the product is zero.

A pentagonal number is one of the form $k(3k \pm 1)/2$, for some natural number $k$. The reason for the name is shown in the dot diagram for the pentagonal number $12 = 3 \cdot 8/2$:

```
  ●
 / \  ●
/   \ ●
●   ● ●
●   ● ●
●   ● ●
●   ● ●
```

For the choice of the plus sign in the formula, the explanation is not so clear, but is explained by the fact that $k(3k + 1)/2 = (-k)(-3k - 1)/2$, so these can be regarded as “pentagonal numbers with negative index”.

**Theorem 8.3 (Euler’s Pentagonal Numbers Theorem)** (a) If $n$ is not a pentagonal number, then the numbers of partitions of $n$ into an even or odd number of distinct parts are equal; if $n = k(3k \pm 1)/2$, the difference between these numbers is $(-1)^k$.

(b) 

$$\prod_{m \geq 1} (1 - x^m) = 1 + \sum_{k \geq 1} (-1)^k \left(x^{k(3k-1)/2} + x^{k(3k+1)/2}\right).$$
The series on the right is
\[ 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + \cdots, \]
and since this is the inverse of the generating function for the partition numbers, we immediately get the recurrence relation:

**Proposition 8.4** The partition numbers \( p(n) \) satisfy \( p(0) = 1 \) and, for \( n \geq 1 \),
\[ p(n) = p(n - 1) + p(n - 2) - p(n - 5) - p(n - 7) + p(n - 12) + p(n - 15) - \cdots, \]
where we include those terms on the right-hand side for which the argument of \( p \) is non-negative — the general terms are
\[ (-1)^{k-1} p(n - k(3k - 1)/2) + (-1)^{k-1} p(n - k(3k + 1)/2). \]

This is a very efficient relation for computation; the number of terms in the recurrence is not constant, but grows quite slowly with \( n \), being only of order \( \sqrt{n} \). So we can compute \( p(m) \) for all \( m \leq n \) with only about \( cn^{3/2} \) additions and subtractions. (Such computations, by McMahon, gave the data on which Ramanujan based his spectacular conjectures about congruences for partition numbers.)

We now turn to the proof of the theorem.

**Proof of Euler’s Pentagonal Numbers Theorem** To demonstrate Euler’s Theorem, we try to produce a bijection between partitions with an even and an odd number of distinct parts; we succeed unless \( n \) is a pentagonal number, in which case a unique partition is left out.

We represent a partition \( \lambda : n = c_1 + c_2 + \cdots \), with \( c_1 \geq c_2 \geq \cdots \), by a diagram \( D(\lambda) \), whose \( i \)th row has \( c_i \) dots aligned on the left. See Figure 10 below for the diagram of the partition \( 17 = 6 + 5 + 4 + 2 \). Such a diagram is essentially the same as a **Ferrers diagram** or **Young diagram**. The difference is that, instead of dots, these diagrams have boxes into which positive integers can be put.

Let \( \lambda \) be any partition of \( n \) into **distinct** parts. We define two subsets of the diagram \( D(\lambda) \) as follows:

- The **base** is the bottom row of the diagram (the smallest part).
- The **slope** is the set of cells starting at the east end of the top row and proceeding in a south-westerly direction for as long as possible.
Figure 10: Base and slope

Note that any cell in the slope is the last in its row, since the row lengths are all distinct. See Figure 10.

Now we divide the set of partitions of $n$ with distinct parts into three classes, as follows:

- **Class 1** consists of the partitions for which *either* the base is longer than the slope and they don’t intersect, *or* the base exceeds the slope by at least 2;
- **Class 2** consists of the partitions for which *either* the slope is at least as long as the base and they don’t intersect, *or* the slope is strictly longer than the base;
- **Class 3** consists of all other partitions with distinct parts.

Given a partition $\lambda$ in Class 1, we create a new partition $\lambda'$ by removing the slope of $\lambda'$ and installing it as a new base, to the south of the existing diagram. In other words, if the slope of $\lambda$ contains $k$ cells, we remove one from each of the largest $k$ parts, and add a new (smallest) part of size $k$. This is a legal partition with all parts distinct. Moreover, the base of $\lambda'$ is the slope of $\lambda$, while the slope of $\lambda'$ is at least as large as the slope of $\lambda$, and strictly larger if it meets the base. So $\lambda'$ is in Class 2.

In the other direction, let $\lambda'$ be in Class 2. We define $\lambda$ by removing the base of $\lambda'$ and installing it as a new slope. Again, we have a partition with all parts distinct, and it lies in Class 1. (If the base and slope of $\lambda$ meet, the base is one greater than the second-last row of $\lambda'$, which is itself greater than the base of $\lambda'$, which has become the slope of $\lambda$. If they don’t meet, the argument is similar.)

The partition shown in Figure 10 is in Class 2; the corresponding Class 1 partition is shown in Figure 11.
These bijections are mutually inverse. Thus, the numbers of Class 1 and Class 2 partitions are equal. Moreover, these bijections change the number of parts by 1, and hence change its parity. So, in the union of Classes 1 and 2, the numbers of partitions with even and odd numbers of parts are equal.

Now we turn to Class 3. A partition in this class has the property that its base and slope intersect, and either their lengths are equal, or the base exceeds the slope by 1. So, if there are \( k \) parts, then \( n = k^2 + k(k - 1)/2 = k(3k - 1)/2 \) or \( n = k(k + 1) + k(k - 1)/2 = k(3k + 1)/2 \). Figure 12 shows the two possibilities.

So, if \( n \) is not pentagonal, then Class 3 is empty; and, if \( n = k(3k - 1)/2 \) or \( n = k(3k + 1)/2 \), for some \( k \in \mathbb{N} \), then it contains a single partition with \( |k| \) parts. Euler’s Theorem follows.

### 8.3 Jacobi’s Triple Product Identity

Euler’s theorem can be deduced from a more general result called Jacobi’s Triple Product Identity.

**Theorem 8.5 (Jacobi’s Triple Product Identity)**

\[
\prod_{n>0}(1 + q^{2n-1}z)(1 + q^{2n-1}z^{-1})(1 - q^{2n}) = \sum_{l \in \mathbb{Z}} q^{l^2}z^l.
\]
The series on the right breaks my rules that formal power series should have only terms with non-negative powers of the variable. This just shows that formal power series are even more flexible than I previously described! You can check that the three infinite products on the left contribute only finitely much to each term $q^n z^l$, with $l$ positive, negative or zero.

By replacing $q$ by $q^{1/2}$ and moving the third term in the product to the right-hand side, the identity takes the form

$$
\prod_{n>0} \left(1 + q^{n-1/2} z \right)\left(1 + q^{n-1/2} z^{-1} \right) = \left(\sum_{l \in \mathbb{Z}} q^{l^2/2} z^l \right) \left(\prod_{n>0} \left(1 - q^n \right)^{-1} \right),
$$
in which form we will prove it. The proof here is a remarkable argument by Richard Borcherds, the Fields Medallist for his work on “Monstrous Moonshine”. He expounded this proof in a talk in the London Algebra Colloquium some years ago. This write-up from my Combinatorics textbook.

A level is a number of the form $n + \frac{1}{2}$, where $n$ is an integer. A state is a set of levels which contains all but finitely many negative levels and only finitely many positive levels. The state consisting of all the negative levels and no positive ones is called the vacuum. Given a state $S$, we define the energy of $S$ to be

$$
\sum \{ l : l > 0, l \in S \} - \sum \{ l : l < 0, l \notin S \},
$$

while the particle number of $S$ is

$$
|\{ l : l > 0, l \in S \}| - |\{ l : l < 0, l \notin S \}|.
$$

Although it is not necessary for the proof, a word about the background is in order!

Dirac showed that relativistic electrons could have negative as well as positive energy. Since they jump to a level of lower energy if possible, Dirac hypothesised that, in a vacuum, all the negative energy levels are occupied. Since electrons obey the exclusion principle, this prevents further electrons from occupying these states. Electrons in negative levels are not detectable. If an electron gains enough energy to jump to a positive level, then it becomes ‘visible’; and the ‘hole’ it leaves behind behaves like a particle with the same mass but opposite charge to an electron. (A few years later, positrons were discovered filling these specifications.) If the vacuum has no net particles and zero energy, then the energy and particle number of any state should be relative to the vacuum, giving rise to the definitions given.
We show that the coefficient of $q^m z^l$ on either side of the equation is equal to the number of states with energy $m$ and particle number $l$. This will prove the identity.

For the left-hand side this is straightforward. A term in the expansion of the product is obtained by selecting $q^{n-\frac{1}{2}z}$ or $q^{n-\frac{1}{2}z-1}$ from finitely many factors. These correspond to the presence of an electron in positive level $n-\frac{1}{2}$ (contributing $n-\frac{1}{2}$ to the energy and 1 to the particle number), or a hole in negative level $-(n-\frac{1}{2})$ (contributing $n-\frac{1}{2}$ to the energy and $-1$ to the particle number). So the coefficient of $q^m z^l$ is as claimed.

The right-hand side is a little harder. Consider first the states with particle number 0. Any such state can be obtained in a unique way from the vacuum by moving the electrons in the top $k$ negative levels up by $n_1, n_2, \ldots, n_k$, say, where $n_1 \geq n_2 \geq \ldots \geq n_k$. (The monotonicity is equivalent to the requirement that no electron jumps over another. The jumping process allows the possibility that some electrons jump from negative levels to higher but still negative levels, so $k$ is not the number of occupied positive levels.) The energy of the state is thus $m = n_1 + \ldots + n_k$. Thus, the number of states with energy $m$ and particle number 0 is equal to the number $p(m)$ of partitions of $m$, which is the coefficient of $q^m$ in $P(q) = \prod_{n>0}(1-q^n)^{-1}$, as we saw in the first section of this chapter.

Now consider states with positive particle number $l$. There is a unique ground state, in which all negative levels and the first $l$ positive levels are filled; its energy is
\[
\frac{1}{2} + \frac{3}{2} + \ldots + \frac{2l-1}{2} = \frac{1}{2}l^2,
\]
and its particle number is $l$. Any other state with particle number $l$ is obtained from this one by ‘jumping’ electrons up as before; so the number of such states with energy $m$ is $p(m-\frac{1}{2}l^2)$, which is the coefficient of $q^m z^l$ in $q^{l^2/2}z^l P(q)$, as required.

The argument for negative particle number is similar.

**Exercises**

8.1. Show that $\lim_{n \to \infty} p(n)^{1/n} = 1$. [Hint: where does the series $\sum_{n \geq 0} p(n)x^n$ converge?]

8.2. Why is the series of partition numbers not C-finite?
8.3. Deduce Euler’s Pentagonal Numbers Theorem from Jacobi’s Triple Product Identity. [Hint: put $q = x^{3/2}$, $z = -x^{-1/2}$.]
9 Set partitions and permutations

It could be said that the main objects of interest in combinatorics are subsets, partitions and permutations of a finite set. We have spent some time counting subsets; now we turn to partitions and permutations.

9.1 Partitions

For partitions, we are unable to write down a simple formula for the counting numbers, and have to rely on recurrence relations or other techniques.

The Bell number $B(n)$ is the number of partitions of a set of cardinality $n$.

There is no simple formula for $B(n)$, and even estimating its size is quite hard. But there is one thing we can say. Every permutation of $\{1, \ldots, n\}$ has a cycle decomposition, an expression as a product of disjoint cycles. The supports of the cycles partition $\{1, \ldots, n\}$. So we have a map from permutations to partitions, which is clearly surjective, since any subset supports a cycle. So $B(n) \leq n!$ for all $n$. We will refine this estimate shortly.

Here is a recurrence relation.

**Proposition 9.1** $B(0) = 1$ and

$$B(n) = \sum_{k=1}^{n} \binom{n-1}{k-1} B(n-k) \text{ for } n > 0.$$  

**Proof** The initial condition is “empty set theory”.

For the recurrence, look at the part of the partition which contains the element $n$. Suppose that this part has size $k$. Then there are $\binom{n-1}{k-1}$ ways to choose the remaining $k$ points, and then $B(n-k)$ ways to partition the points that are left over. Multiplying and summing over $k$ gives the result.

This recurrence allows us to work out the exponential generating function for the Bell numbers, which has a nice form:

**Proposition 9.2**

$$\sum_{n \geq 0} \frac{B(n)x^n}{n!} = \exp(\exp(x) - 1).$$
Proof Let \( F(x) \) be the function on the left of the equation. Then

\[
\frac{d}{dx} F(x) = \sum_{n \geq 1} \frac{B(n)x^{n-1}}{(n-1)!}
\]

\[
= \sum_{n \geq 0} \sum_{k=1}^{n} \frac{x^{k-1} B(n-k)x^{n-k}}{(k-1)! (n-k)!}
\]

\[
= \left( \sum_{l \geq 0} \frac{x^{l}}{l!} \right) \left( \sum_{m \geq 0} \frac{B(m)x^{m}}{m!} \right)
\]

\[
= \exp(x) F(x).
\]

(In the third line we replaced the variables \( n \) and \( k \) by \( l = k - 1 \) and \( m = n - k \). The given range of values of \( n \) and \( k \) corresponds to \( l \) and \( m \) taking all non-negative integer values.)

Now we have a first-order linear differential equation with initial condition, which is easily solved. We find that

\[
\frac{d}{dx} (\exp(-\exp(x))F(x)) = 0,
\]

so \( F(x) = A \exp(\exp(x)) \); and the initial value \( F(0) = 1 \) gives \( A = \exp(-1) \).

We learn something about the magnitude of Bell numbers from this theorem, which strengthens greatly our result that \( B(n) \leq n! \):

**Proposition 9.3** The ratio \( B(n)/n! \) tends to zero faster than any exponential function \( c^n \) for \( c > 0 \).

**Proof** The radius of convergence of a power series \( \sum a_n z^n \) is the number \( r \) such that the series converges for \( |z| < r \) and diverges for \( |z| > r \). (We do not specify the behaviour for \( |z| = r \).) This includes the cases \( r = 0 \) (the series diverges for all \( z \neq 0 \) and \( r = \infty \) (the series converges for all \( z \)).

We use two facts from analysis:

- a convergent power series defines an analytic function inside the disc of convergence, and the radius of convergence is the distance from the origin to the nearest singularity of the function;
- the radius of convergence of \( \sum a_n z^n \) is the reciprocal of \( \limsup_{n \to \infty} a_n^{1/n} \).
Now the sum of $\sum_{n \geq 0} B(n)z^n/n!$ is analytic for all $z$, so the radius of convergence is infinite; so $\lim_{n \to \infty} (B(n)/n!)^{1/n} = 0$. This is equivalent to the statement of the proposition.

Here is another application. For the partition numbers $p(n)$, we saw that

$$\sum_{n \geq 0} p(n)z^n = \prod_{k \geq 1} (1 - z^k)^{-1}.$$ 

The right-hand side is analytic inside the unit circle, but has singularities at every root of unity. So the radius of convergence is 1, whence $\limsup_{n \to \infty} p(n)^{1/n} = 1$. This means that $p(n)$ grows slower than any exponential function $c^n$ with $c > 1$. On the other hand, it can be shown that it grows faster than any polynomial.

We refine the Bell numbers in the same way that the binomial coefficients do for subset counting.

The Stirling number of the second kind, $S(n, k)$, is the number of partitions of an $n$-set into $k$ parts. Thus, $S(0, 0) = 1$ and $S(0, k) = 0$ for $k > 0$; and if $n > 0$, then $S(n, 0) = 0$, $S(n, 1) = S(n, n) = 1$, and $S(n, k) = 0$ for $k > n$. Clearly we have

$$\sum_{k=1}^{n} S(n, k) = B(n) \text{ for } n > 0.$$ 

The recurrence relation replacing Pascal’s is:

**Proposition 9.4**

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k) \text{ for } 1 \leq k \leq n.$$ 

**Proof** We divide the partitions up into two classes:

- Those in which $n$ is a singleton part. This part is simply added to a partition of $\{1, \ldots, n - 1\}$ with $k - 1$ parts, so there are $S(n - 1, k - 1)$ of this type.

- Those in which $n$ is in a part of size greater than 1. Removing $n$ from this part, we have a partition of $\{1, \ldots, n - 1\}$ with $k$ parts. Given such a partition, we can choose arbitrarily which of the $k$ parts we add $n$ to. So there are $kS(n - 1, k)$ of this type.

Adding gives the result.
The Stirling numbers can be set out in a triangle like Pascal’s. We will left-align this triangle (that is, take the column $k = 1$ to be vertical). The recurrence relation says that each entry is obtained by adding the entry to its north-west and the entry above multiplied by $k$:

\[
\begin{array}{cccc}
1 \\
1 & 1 \\
1 & 1 & 2 \\
1 & 1 & 3 & 1 \\
1 & 7 & 6 & 1
\end{array}
\]

You may spot some patterns here, especially if you continue the triangle a couple more steps. See the exercises for some of these.

It turns out that we can turn the recurrence into a statement about a generating function, but with a twist. Let

\[(x)_k = x(x - 1) \cdots (x - k + 1) \quad (k \text{ factors}).\]

Then we have

**Theorem 9.5**

\[x^n = \sum_{k=1}^{n} S(n,k)(x)_k \text{ for } n > 0.\]

**Proof** I will give two proofs of this result; the first one involves manipulations with the generating functions, the second one is more obviously combinatorial.

**First proof** Assume the result for $n - 1$. Then

\[
\begin{align*}
\sum_{k=1}^{n} S(n,k)(x)_k &= \sum_{k=1}^{n} S(n-1,k-1)(x)_{k-1}(x - k + 1) + \sum_{k=1}^{n} kS(n-1,k)(x)_k \\
&= x^{n-1} \cdot x - \sum_{k=2}^{n} S(n-1,k-1)(x)_{k-1}(k - 1) + \sum_{l=1}^{n-1} (l-1)S(n-1,l-1)x_{l-1} \\
&= x^n,
\end{align*}
\]

since the last two terms cancel. (We changed the variable to $l = k + 1$ in the third term, and omitted terms involving $S(n-1,0)$ and $S(n-1,n)$ whose value is zero.)
Second proof  Take an auxiliary set $X$ of size $x$, where $x \geq n$. Now $x^n$ is the total number of $n$-tuples of elements from $X$. We count these $n$-tuples another way.

Each $n$-tuple $(x_1, \ldots, x_n)$ defines a partition of $\{1, \ldots, n\}$, where we put $i$ and $j$ into the same part of the partition if $x_i = x_j$. Having chosen a partition, how do we recover the $n$-tuple? Suppose that the partition has $k$ parts $P_1, \ldots, P_k$, ordered by the smallest element (so $P_1$ is the part containing 1, $P_2$ the part containing the smallest element not in $P_1$, and so on). Now we have to choose a value in $X$ for the elements $x_i$ for $i \in P_1$, a different value for the elements $x_i$ with $i \in P_2$, and so on; in all, the number of choices of $k$ distinct elements of $X$ in order, which is $x(x-1) \cdots (x-k+1) = (x)_k$. Now there are $S(n,k)$ partitions with $k$ parts to consider.

Thus we have
\[ \sum_{k=1}^{n} S(n,k) x^k = x^n. \]

This proves the theorem when $x$ is a positive integer. But both sides are polynomials of degree $n$; if they take the same value for every positive integer $x \geq n$, then they must coincide as polynomials. So the result is proved.

It is possible to find the ordinary generating function for the index $n$:
\[ \sum_{n \geq k} S(n,k) y^n = \frac{y^k}{(1-y)(1-2y) \cdots (1-ky)}. \]

Also, the exponential generating function for the index $n$ is
\[ \sum_{n \geq k} S(n,k) \frac{x^n}{n!} = \frac{(\exp(x)-1)^k}{k!}. \]

Summing over $k$ gives the e.g.f. for the Bell numbers:
\[ \sum_{n \geq 0} \frac{B(n)x^n}{n!} = \exp(\exp(x)-1). \]

However, I will not give the proofs of these results here.
9.2 Permutations

The number of permutations of an \( n \)-set (bijective functions from the set to itself) is the factorial function \( n! = n(n-1) \cdots 1 \) for \( n \geq 0 \). The exponential generating function for this sequence is \( 1/(1-x) \), while the ordinary generating function has no analytic expression (it is divergent for all \( x \neq 0 \)). The recurrence relation for the factorials is

\[
0! = 1, \quad n! = n \cdot (n-1)! \quad \text{for} \quad n \geq 1.
\]

Any permutation can be decomposed uniquely into disjoint cycles. So we refine the count by letting \( u(n, k) \) be the number of permutations of an \( n \)-set which have exactly \( k \) cycles (including cycles of length 1). Thus,

\[
\sum_{k=1}^{n} u(n, k) = n! \quad \text{for} \quad n > 0.
\]

The numbers \( u(n, k) \) are the unsigned Stirling numbers of the first kind.

The reason for the name is that it is common to use a different count, where a permutation is counted with weight equal to its sign (as defined in elementary algebra, for example the theory of determinants).

The sign of a permutation has several definitions, all equivalent:

- The sign of a permutation is the determinant of the corresponding permutation matrix. But this definition has the flaw that determinants are usually defined using the concept of sign!

- The sign of a permutation \( g \) is \((-1)^p\), where \( g \) can be written as the product of \( p \) transpositions. This definition has the drawback that we need to show that the parity of the number of transpositions whose product is \( g \) is the same, for any expression.

- If \( g \) is a permutation of \( \{1, \ldots, n\} \) with \( k \) cycles in its cycle decomposition (including cycles of length 1), then the sign of \( g \) is \((-1)^{n-k}\). I regard this as the best definition.

I will sketch the proof that the second and third definitions agree – this also shows that the second definition works. Let \( g \) be an arbitrary permutation, decomposed into cycles, and let \( t \) be the transposition \((i, j)\). It is easy to show [do it!] that, if \( i \) and \( j \) are in different cycles of \( g \), then these two cycles are cut at these points and stitched into a single cycle in \( gt \); and conversely, if \( i \) and \( j \) are in the
same cycle of $g$, then this cycle is divided into two cycles of $gt$. So the parity of $n - k$ is changed (up or down) by 1 when we multiply by a transposition, and hence the sign (according to the third definition) is reversed. Now the identity has $n$ cycles, and so has sign $+1$. Any permutation is a product of transpositions; if it is a product of $p$ transpositions, then its sign is $(-1)^p$ (after $p$ reversals).

Now the signed Stirling number of the first kind is defined to be

$$s(n, k) = (-1)^{n-k} u(n, k).$$

Note that $(-1)^{n-k}$ is the sign of every permutation counted by $u(n, k)$; so we are just counting permutations “according to their sign”.

We have

$$\sum_{k=1}^{n} s(n, k) = 0 \text{ for } n > 1.$$  

This is related to the algebraic fact that, for $n > 1$, the permutations with sign $+$ form a subgroup of the symmetric group of index 2 (that is, containing half of all the permutations), called the alternating group. (The second definition shows that the sign map is a homomorphism from the symmetric group to the multiplicative group $\{+1, -1\}$; for $n > 1$, a transposition has sign $-1$, so the map is onto. Its kernel is the alternating group.)

We will mainly consider signed Stirling numbers below, though it is sometimes convenient to prove a result first for the unsigned numbers.

As usual we take $s(n, 0) = 0$ for $n > 0$ and $s(n, k) = 0$ for $k > n$.

Here is the recurrence relation.

**Proposition 9.6** We have $s(n, n) = 1$, $s(n, 1) = (-1)^{n-1}(n-1)!$, and

$$s(n, k) = s(n-1, k-1) - (n-1)s(n-1, k) \text{ for } 1 \leq k \leq n.$$  

**Proof** $s(n, n)$ counts just the identity permutation, which has sign 1. On the other hand, $s(n, n)$ counts the cyclic permutations. Since we may choose the starting point of a cycle arbitrarily, we can assume that the first entry in the cycle is 1; the other entries are then an arbitrary permutation of \{2, $\ldots$, $n\}, and the sign of a cycle is $(-1)^{n-1}$.

Now divide the permutations with $k$ cycles into two classes, according to their action on the point $n$.

- In the first class, $n$ is fixed, so the fixed point is adjoined to a permutation of \{1, $\ldots$, $n-1\} with $k-1$ classes; there are $u(n-1, k-1)$ such permutations.
In the second class, \( n \) occurs in some cycle \((\ldots, j, n, i, \ldots)\) of length greater than 1. (We may have \( i = j \) here.) By simply bypassing \( n \), replacing this cycle by \((\ldots, j, i, \ldots)\), we obtain a permutation on \( \{1, \ldots, n-1\} \) also with \( k \) cycles. Conversely, given such a permutation, we can choose to insert \( n \) anywhere in any of the cycles; there are thus \( n-1 \) ways this can be done, so the number of permutations is \((n-1)u(n-1,k)\).

Thus \( u(n,k) = u(n-1,k-1) + (n-1)u(n-1,k) \), and putting in the signs gives the result. (Permutations counted by \( u(n,k) \) and \( u(n-1,k-1) \) have the same sign, while those counted by \( u(n-1,k) \) have the opposite sign.)

We can produce a triangle of Stirling numbers of the first kind similarly to what we did for those of the second kind. This time, the multiplier on the southerly arrow is the row number of the source row, rather than the column number as before, and with the sign changed:

\[
\begin{array}{c}
1 \\
-1 \\
-2 \\
2 \\
-3 \\
-6 \\
1 \\
11 \\
-6 \\
1
\end{array}
\]

From this, we find a generating function:

**Theorem 9.7**

\[
\sum_{k=1}^{n} s(n,k)x^k = (x)_n.
\]

Putting \( x = 1 \) in this equation shows that indeed the sum of the signed Stirling numbers is zero for \( n > 1 \).

**Proof** The proof of this can be done by manipulating summations in the same way that we did for the other kind of Stirling numbers.

If you have met the **Orbit-Counting Lemma** for permutation groups, a nicer proof is possible. I will outline this here.

Suppose that a finite group \( G \) acts on a finite set \( T \). We denote by \( \text{fix}(g) \) the number of points of the set \( T \) which are fixed by the element \( g \in G \). We say that
two elements \(x, y\) of \(T\) are \textit{in the same orbit} if there is an element of \(G\) carrying \(x\) to \(y\). This is an equivalence relation on \(X\). (The identity, inverse and closure laws for \(G\) show respectively the reflexive, symmetric and transitive laws for the relation.) The equivalence classes are called, naturally enough, \textit{orbits}.

Now the Orbit-counting Lemma asserts that the number of orbits is

\[
\frac{1}{|G|} \sum_{g \in G} \text{fix}(g).
\]

I won’t prove this here, but will now apply it.

We take the group to be the symmetric group \(S_n\), and its action is on the set of all \(n\)-tuples of elements from an auxiliary set \(X\) with \(|X| = x\). Which \(n\)-tuples are fixed by a permutation \(g\)? Such a tuple must be constant on each cycle of \(g\); so the number is \(x^{c(g)}\), where \(c(g)\) is the number of cycles of \(g\). Thus, the number of orbits of \(G\) is

\[
\frac{1}{n!} \sum_{g \in G} x^{c(g)} = \frac{1}{n!} \sum_{k=1}^{n} u(n, k)x^k.
\]

On the other hand, two \(n\)-tuples lie in the same orbit of the symmetric group if and only if they contain each element of \(X\) the same number of times. Thus the orbits correspond to unordered selections of \(n\) things from a set of size \(x\). We saw that the number of such selections is

\[
\binom{n + x - 1}{n} = \frac{x(x+1) \cdots (x+n-1)}{n!}.
\]

Equating the two expressions and multiplying by \(n!\) gives

\[
x(x+1) \cdots (x+n-1) = \sum_{k=1}^{n} u(n, k)x^k.
\]

Now replacing \(x\) by \(-x\) and multiplying by \((-1)^n\) gives

\[
x(x-1) \cdots (x-n+1) = \sum_{k=1}^{n} s(n, k)x^k.
\]

We have proved the identity when \(x\) is a positive integer (since we need a set of cardinality \(x\) in the argument). But both sides are polynomials; so if the two expressions are equal for every positive integer \(x\), then they are equal as polynomials, as required.

Note that this is the inverse of the relation we found for the Stirling numbers of the second kind. So the matrices formed by the Stirling numbers of the first and second kind are inverses of each other.
9.3 The Möbius function of the partition lattice

In this section we will use our knowledge of the Stirling numbers of the first kind to calculate the Möbius function of the poset of partitions.

Let \( A \) be the set of all partitions of \( \{1, \ldots, n\} \), so that \( |A| = B(n) \), the Bell number. For \( a, b \in A \), we put \( a \leq b \) if \( a \) refines \( b \), that is, \( a \) splits the parts of \( b \), but each part of \( a \) is contained in a single part of \( b \). For example, \( a = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6\}\} \) refines \( b = \{\{1, 2\}, \{3, 4, 5, 6\}\} \), but does not refine \( \{\{1, 2\}, \{3, 4\}, \{5, 6\}\} \).

This poset is called the partition lattice, and denoted by \( P(n) \).

**Theorem 9.8** Suppose that \( b \) has \( r \) parts \( p_1, \ldots, p_r \), and that \( a \) splits \( p_i \) into \( m_i \) smaller parts for \( i = 1, \ldots, r \). Let \( m \) be the sum of the \( a_i \). Then

\[
\mu(a, b) = (-1)^{m-r}(m_1 - 1)!(m_2 - 1)! \cdots (m_r - 1)!. 
\]

**Proof** We begin with a couple of reductions.

**First reduction** The partitions which refine \( b \) form the cartesian product of partition lattices \( P(n_1) \times \cdots \times P(n_r) \). So it is enough to compute the Möbius function for the case when \( b \) has a single part, and multiply the results together.

**Second reduction** The partitions in the interval \([a, b]\) do not split the parts of \( a \), which behave like atoms. So, if \( b \) has a single part and \( a \) has \( m \) parts, then this interval is isomorphic to \( P(m) \).

We conclude that it is enough to compute \( \mu(a, b) \), where \( b \) is the partition of an \( m \)-set with a single part, and \( a \) is the partition of this set into parts of size 1. If this number is \( f(m) \), then the answer to our original problem is \( f(m_1) \cdots f(m_r) \).

We are going to show that \( f(m) = (-1)^{m-1}(m - 1)! \). The theorem will follow from this.

Note that our proposed value of \( f(m) \) is the Stirling number of the first kind \( s(m, 1) \). This is the clue that Stirling numbers are involved!

We review a couple of facts about Stirling numbers:

- \( s(m, k) \) is \((-2)^{m-k}\) times the number of permutations of an \( m \)-set with \( k \) cycles.
- \( s(m, 1) = (-1)^{m-1}(m - 1)! \).
Let $a$ and $b$ be the partition into singletons and the partition with a single part of a set of cardinality $m$. Then $\mu(a, b)$ is the negative of the sum of the values of $\mu(a, c)$, for $a \leq c < b$.

Consider a partition $c$ of $\{1, \ldots, m\}$ with $k$ parts, having sizes $a_1, \ldots, a_k$. For $k > 1$, then $\mu(a, c)$ is the product of $(-1)^{a_i - 1}(a_i - 1)!$ over all $i$. But this number is the number of permutations having the part of size $a_i$ as a single cycle, multiplied by the appropriate sign. So when we take the product over $i$, we obtain the number of permutations with the given parts as cycles, multiplied by the sign of such permutations. Summing over all partitions of size $k$ then gives us $s(m, k)$.

So the sum over all such $c$ is

$$\sum_{k=2}^{m} s(m, k) = -s(m, 1),$$

whence $\mu(a, b) = s(m, 1)$, as required. The proof is complete.

**Exercises**

9.1. Show that $S(n, 2) = 2^{n-1} - 1$, $S(n, n - 1) = n(n - 1)/2$, and $S(n, n - 2) = n(n - 1)(n - 2)(3n - 5)/24$.

9.2. Show that $s(n, n - 1) = -n(n - 1)/2$. What is $s(n, n - 2)$?

9.3. Show that conjugacy is an equivalence relation in any group.

The next exercise describes a deep connection between partitions, Stirling numbers of the second kind, and the exponential function on one hand, and between permutations, Stirling numbers of the first kind, and the logarithm function on the other hand. The inverse relation between the exponential and logarithm functions reflects the inverse relation between the two kinds of Stirling numbers that we saw at the end of section 9.2.

9.4. (a) Prove that the following are equivalent for sequences $(a_1, a_2, \ldots)$ and $(b_1, b_2, \ldots)$, with exponential generating functions $A(x)$ and $B(x)$ respectively:

   (i) $b_n = \sum_{k=1}^{n} S(n, k)a_k$ for $n \geq 1$;

   (ii) $B(x) = A(\exp(x) - 1)$. 

105
(b) Prove that the following are equivalent for sequences \((a_1, a_2, \ldots)\) and \((b_1, b_2, \ldots)\), with exponential generating functions \(A(x)\) and \(B(x)\) respectively:

(i) \(b_n = \sum_{k=1}^{n} s(n,k) a_k\) for \(n \geq 1\);

(ii) \(B(x) = A(\log(1 + x))\).
10  Stirling’s formula

One very important part of enumerative combinatorics, which we have hardly
touched on in this course, is asymptotic enumeration: given a combinatorial func-
tion $F(n)$, how rapidly does it grow (compared, say, to familiar analytic func-
tions)? We assume all our functions below are non-negative, a natural assumption
for counting functions!

We write $F(n) \sim G(n)$ (where $G(n) \neq 0$ for all $n$) to mean that $F(n)/G(n) \to 1$
as $n \to \infty$; in other words, the rate of growth of $F(n)$ is the same as that of $G(n)$.
In the proof, we also use the “small oh” notation: $F(n) = o(G(n))$ means that
$F(n)/G(n) \to 0$ as $n \to \infty$. (A weaker form, which we will not be using, is the
“big oh” notation: $F(n) = O(G(n))$ means that there exists a constant $C$ such that
$F(n) \leq CG(n)$ for all sufficiently large $n$.)

Perhaps the most important asymptotic result is Stirling’s formula, which gives
the asymptotics of the number of permutations of $\{1, \ldots, n\}$. We give the proof as
an illustration.

**Theorem 10.1**

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

**Proof**  Consider the graph of the function $y = \log x$ between $x = 1$ and $x = n$,
together with the piecewise linear functions shown in Figure 13.

Let $f(x) = \log x$, let $g(x)$ be the function whose value is $\log(m+1)$ for
$m \leq x < m + 1$, and let $h(x)$ be the function defined by the polygon with vertices $(m, \log m)$,
for $1 \leq m \leq n$. Clearly

$$\int_1^n g(x) \, dx = \log 2 + \cdots + \log n = \log n!.$$  

The difference between the integrals of $g$ and $h$ is the sum of the areas of triangles with base 1 and total height $\log n$; that is, $\frac{1}{2} \log n$.

Some calculus (given at the end of this proof) shows that the difference be-
tween the integrals of $f$ and $h$ tends to a finite limit $c$ as $n \to \infty$.

Finally, a simple integration shows that

$$\int_1^n f(x) \, dx = n \log n - n + 1.$$  

We conclude that

$$\log n! = n \log n - n + \frac{1}{2} \log n + (1 - c) + o(1),$$

107
so that

\[ n! \sim C n^{n+1/2} e^{-n} \]

To identify the constant \( C \), we can proceed as follows. Consider the integral

\[ I_n = \int_0^{\pi/2} \sin^n x \, dx. \]

Integration by parts shows that

\[ I_n = \frac{n-1}{n} I_{n-2}, \]

and hence

\[ I_{2n} = \frac{(2n)! \pi}{2^{2n+1} (n!)^2}, \]

\[ I_{2n+1} = \frac{2^{2n} (n!)^2}{(2n+1)!}. \]

On the other hand,

\[ I_{2n+2} \leq I_{2n+1} \leq I_{2n}, \]

108
from which we get
\[
\frac{(2n + 1)\pi}{4(n + 1)} \leq \frac{2^{4n}(n!)^4}{(2n)!(2n + 1)!} \leq \frac{\pi}{2},
\]
and so
\[
\lim_{n \to \infty} \frac{2^{4n}(n!)^4}{(2n)!(2n + 1)!} = \frac{\pi}{2}.
\]
Putting \(n! \sim Cn^{n+1/2}/e^n\) in this result, we find that
\[
\frac{C^2e}{4} \lim_{n \to \infty} \left(1 + \frac{1}{2n}\right)^{(2n+3/2)} = \frac{\pi}{2},
\]
so that \(C = \sqrt{2\pi}\).

The last part of this proof is taken from Alan Slomson’s *An Introduction to Combinatorics*. The equation for \(\pi/2\) is equivalent to Wallis’ product formula:
\[
\frac{\pi}{2} = \frac{2 \cdot 2}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \cdots}
\]
However, convergence is slow; putting \(n = 10^6\) gives \(\pi\) to five places of decimals.

Here is the argument we skipped over above.

Let \(F(x) = f(x) - h(x)\). The convexity of \(\log x\) shows that \(F(x) \geq 0\) for all \(x \in [m, m + 1]\). For an upper bound we use the fact, a consequence of Taylor’s Theorem, that
\[
\log x \leq \log m + \frac{x - m}{m} \leq \log m + \frac{1}{m}
\]
for \(x \in [m, m + 1]\). Then
\[
F(x) = \log x - \log m - \log \left(1 + \frac{1}{m}\right)(x - m) \leq \frac{1}{m} - \log \left(1 + \frac{1}{m}\right) \leq \frac{1}{2m^2},
\]
where the last inequality comes from another application of Taylor’s Theorem which yields \(\log(1 + x) \geq x - x^2/2\) for \(x \in [0, 1]\). Now \(\sum(1/m^2)\) converges, so the integral is bounded.

Here is an application. If \(n\) is even, how big is the central binomial coefficient \(\binom{2n}{n}\)? We know that it lies between \(2^n/(n+1)\) and \(2^n\), since it is the largest of the \(n + 1\) binomial coefficients \(\binom{n}{k}\), which sum to \(2^n\).
Using Stirling’s formula, we have

\[
\binom{n}{n/2} = \frac{n!}{((n/2)!)^2} \sim \sqrt{2\pi nn^e} \cdot \frac{e^n}{\pi n(n/2)^n} = \sqrt{\frac{2}{\pi n}} \cdot 2^n.
\]

**Exercises**

10.1. Let \( p \) be a rational number satisfying \( 0 < p < 1 \). Suppose that \( n \to \infty \) taking values so that \( pn \) is an integer. Show that

\[
\log \left( \frac{n}{pn} \right) \sim H(p)n,
\]

where \( H \) is the *entropy function* given by

\[
H(p) = -p\log p - (1-p)\log(1-p).
\]

10.2. *Cayley’s Theorem* asserts that the number \( T_n \) of trees on a given vertex set \( \{1, \ldots, n\} \) is given by \( T_n = n^{n-2} \) for \( n \geq 1 \).

Prove that the radius of convergence of the exponential generating function

\[
\sum_{n \geq 1} \frac{T_n x^n}{n!}
\]

is \( 1/e \).
11 Solutions to the exercises

1.1. (a) You might discover that \(1001 = \binom{14}{4}\), and conclude that you were allowed to choose any four toppings from a list of 14. In fact, it was not so organised: there was a list of 25 toppings and you could choose any subset, so the actual number of combinations was 33554432. I guess they felt that 1001 was a big enough number.

(b) Of course the correct answer is \(2^8 - 1 = 255\). (The \(-1\) is because, if you didn’t want any items, you presumably would not be having the meal deal in the first place.) Probably \(40312 = 8! - 8\) was used just because it was a larger number!

The Advertising Standards Authority was asked to rule on this advertisement, and “considered that the number quoted in the advertisement was not necessarily so exaggerated as to be misleading”. The sequence with \(n\)th term \(n! - n\) is listed in the On-Line Encyclopedia of Integer Sequences as McCombinations (sequence number A005096).

1.2. Use the fact that

\[(1 + x)^{m+n} = (1 + x)^m \cdot (1 + x)^n.\]

The coefficient of \(x^k\) on the left is \(\binom{m+n}{k}\); on the right it is

\[\sum \binom{m}{i} \binom{n}{k-i}.\]

1.3. Let \(S_n(a, b)\) be the sum of the binomial coefficients \(\binom{n}{k}\) with \(k \equiv a \pmod{b}\).

First note that, for \(n \geq 1\),

\[2^n = (1 + 1)^n = \sum_{k=0}^{n} \binom{n}{k} = S_n(0, 2) + S_n(1, 2),\]

\[0 = (1 - 1)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} = S_n(0, 2) - S_n(1, 2).\]

So \(S_n(0, 2) = S_n(1, 2) = 2^{n-1}\).

In particular, \(S_n(0, 4) + S_n(2, 4) = S_n(0, 2) = 2^{n-1}\).
Now follow the hint:

$$(1 + i)^n = \sum_{k=0}^{n} i^k \binom{n}{k} = S_n(0, 4) + iS_n(1, 4) - S_n(2, 4) - iS_n(3, 4).$$

Equating real parts, we find that

$$S_n(0, 4) - S_n(2, 4) = \Re((1 + i)^n).$$

To evaluate the right-hand side, use the modulus-argument form:

$$1 + i = \sqrt{2}e^{\pi i/4}.$$

So

$$\Re((1 + i)^n) = 2^{n/2}\cos(n\pi/4).$$

Now the cosine depends on the congruence of $n \mod 8$, taking the values

$$1, 1/\sqrt{2}, 0, -1/\sqrt{2}, -1, -1/\sqrt{2}, 0, 1/\sqrt{2}$$

for $n \equiv 0, 1, \ldots, 7 \pmod{8}$. Now adding the two equations and dividing by 2 gives the answer. For example, if $n \equiv 5 \pmod{8}$, we find that $S_n(0, 4) = (2^{n-1} - 2^{(n-1)/2})/2$.

As a check, for $n = 13$, we have

$$\binom{13}{0} + \binom{13}{4} + \binom{13}{8} + \binom{13}{12} = 1 + 705 + 1269 + 13 = 2016 = (4096 - 64)/2.$$

1.4. A little experimentation suggests the following. Suppose that we compute rows $0, 1, \ldots, 2^d - 1$ of Pascal’s triangle mod 2, obtaining the triangle $A$. The bottom row of $A$ consists entirely of 1s. (This is the induction hypothesis.) Then the next row is 1, 0, 0, \ldots, 0, 1, and so the next $2^d$ rows produce a copy of $A$ on the left and one on the right, with zeros in the inverted triangle between.

$$A \rightarrow \begin{array}{c} A \\ 0 \\ A \end{array}$$

In particular, row $2^{d+1} - 1$ consists entirely of 1s, so this statement is proved for all $d$ by induction.

The fractal limit would be obtained by taking a black triangle, dividing it into four equal pieces and removing the inverted piece in the middle, and iterating this procedure on the three remaining triangles.
1.5. The statement of the theorem follows easily by induction from the statement of the hint. To prove this we use a trick which I learned from Bruce Sagan.

Let \( a = ps + c \) and \( b = pt + d \), where \( 0 \leq c,d \leq p - 1 \). Let \( \sigma \) be a permutation of \( \{1, \ldots, a\} \) having \( s \) cycles of length \( p \) and \( c \) fixed points: for example,

\[
\sigma = (1,2,\ldots,p)(p+1,\ldots,2p)(s-1)p+1,\ldots,sp)
\]

with the points \( sp+1,\ldots,sp+c \) fixed.

Consider how \( \sigma \) permutes the \( b \)-element subsets of \( \{1,\ldots,a\} \). Since \( p \) is prime, any such subset is either fixed or lies in a cycle of length \( p \). So \( \binom{a}{b} \) is congruent mod \( p \) to the number of sets fixed by \( \sigma \). Such a fixed set can be obtained by choosing \( t \) of the \( p \)-cycles of \( \sigma \) and \( d \) of the fixed points (since a fixed set must contain all or no points from each cycle); so the number is \( \binom{s}{t}\binom{c}{d} \).

The assertion is proved.

In the case when \( p = 2 \), the digits in the base 2 expansions are either 0 or 1. We have \( \binom{0}{0} = \binom{1}{0} = \binom{1}{1} = 1 \) while \( \binom{0}{1} = 0 \). So a binomial coefficient \( \binom{n}{k} \) is odd if and only if, when \( n \) and \( k \) are written to base 2, each digit of \( n \) is greater than or equal to the corresponding digit of \( k \).

For example, in base 2, we have \( 6 = 110 \) and \( 3 = 001 \). The units digit shows us that \( \binom{6}{3} \) is even.

1.6. The easiest way to do this exercise is to “reverse-engineer” the counting interpretation of the multinomial coefficient: given a set of \( n \) elements, it is the number of ways we can put them into \( r \) labelled boxes \( B_1,\ldots,B_r \) such that the number of objects in box \( i \) is \( k_i \). To see this, first order the objects (this can be done in \( n! \) ways), then put the first \( k_1 \) in box \( B_1 \), the next \( k_2 \) in box \( B_2 \), \ldots, and the last \( k_r \) in box \( B_r \). Now we have overcounted by a factor which is the number of different orders in which the elements of the boxes could have been put in, which is \( k_1!k_2!\cdots k_r! \).

Now we can prove the Multinomial Theorem by counting. In the expansion of \( (x_1 + \cdots + x_r)^n \), every term is a monomial of degree \( n \). The monomial \( x_1^{k_1}x_2^{k_2}\cdots x_r^{k_r} \) is obtained by choosing \( x_1 \) from \( k_1 \) of the \( n \) brackets, \( x_2 \) from \( k_2 \) brackets, \ldots, \( x_r \) from \( k_r \) brackets; this can be done in

\[
\binom{n}{k_1,k_2,\ldots,k_r}
\]
ways.

To give an analytic proof, it would be necessary to prove the Pascal-like recurrence for the multinomial coefficients, which is as follows. \( \binom{n}{k_1, k_2, \ldots, k_r} \) is the sum of \( r \) terms of the form \( \binom{n-1}{l_1, l_2, \ldots, l_r} \), where in the \( i \)th term we have \( l_j = k_j \) for \( j \neq i \) and \( l_i = k_i - 1 \).

1.7. Given a subset \( \{a_1, \ldots, a_k\} \) of \( \{1, \ldots, n\} \), we can always write it so that \( a_1 < a_2 < \ldots < a_k \). The set contains no two consecutive numbers if and only if \( a_{i+1} \geq a_i + 2 \) for all \( i \) (all the gaps are at least 2). We define the bijection by shrinking the gaps by one:

\[
\{a_1, a_2, \ldots, a_k\} \mapsto \{a_1, a_2 - 1, \ldots, a_k - (k - 1)\}.
\]

The largest possible element in such a set would be \( n - (k - 1) = n - k + 1 \). To reverse the bijection, take a \( k \)-subset of \( \{1, \ldots, n - k + 1\} \), arrange it in order, and add \( i - 1 \) to the \( i \)th term to increase all the gaps by one.

In the case of the National Lottery, the required probability is

\[
\frac{\binom{44}{6}}{\binom{49}{6}} = 0.5048\ldots,
\]

or roughly one-half, a result that many people find surprising.

The solution to Exercise 1.8 will be deferred for a while.

2.1. This says, “if we colour the 2-sets of an \( n \)-set red and blue, we get either a red 2-set or a \( l \)-set all of whose 2-subsets are blue, provided that \( n \geq l \); while \( n = l - 1 \) does not suffice for the conclusion. It is now almost obvious: if any 2-set is red, we have won; if not, then we have won provided that we have at least \( l \) points in our set.

The more general case is exactly the same.

2.2. Proof of the lemma: if we sum, over all people \( P_i \), the number of times that \( P_i \) shakes hands, we obtain an even number, since each handshake contributes two to the sum. So the number of odd summands must be even.

Now consider the proof of \( R_2(3,4) \leq 10 \). We observed that, if a point \( p \) has four red neighbours, then we succeed (since if the set of neighbours has a red edge, then we get a red triangle including \( p \), whereas if not then we have four points with all edges blue); and if \( p \) has six blue neighbours, we also succeed, since by
\( R_2(3, 3) = 6 \) we have either a red or a blue triangle in the neighbour set, and the blue triangle extends to a blue 4-set containing \( p \). Thus, if we have a colouring without either a red triangle or a blue 4-set, then any point has at most 3 red neighbours and at most 5 blue neighbours. In the case \( n = 9 \), since \( 3 + 5 = 9 - 1 \), this is only possible if “at most” is replaced by “exactly” in both places. But then there are nine vertices, each lying on three red edges; the Handshaking Lemma is contradicted.

2.3. Take an octagon with the eight edges and four diagonals red. It is easy to see that there are no red triangles. Colour the remaining edges blue. If there is a blue 4-set, then (by cyclic symmetry) we can assume it contains vertex 0 (if the vertices are numbered 0, ..., 7 in the obvious way). The blue neighbours of 0 are 2, 3, 5, 6, and the blue edges within this set are \{2, 5\}, \{3, 6\}, and \{3, 5\}; there is no blue triangle in the neighbourhood of 0, and so no blue 4-set in the whole configuration. So 8 is too small to force one or other of the two figures we are looking for, and we conclude that \( R_2(3, 4) \geq 9 \).

2.4. First we have to observe that the statement we are asked to prove does really imply the conclusion. For suppose that we have found finitely many rational numbers \( p_i/q_i \) such that \( |\alpha - p_i/q_i| < 1/q_i^2 \) for \( i = 1, \ldots, r \). We have to show that there is another one. Let \( q' \) be the largest of \( q_1, \ldots, q_r \). Since \( \alpha \) is irrational, we can choose \( n \) so large that all the differences \( |\alpha - p_i/q_i| \) are greater than \( 1/nq' \). Then, if the statement in the problem is correct, there will be a rational \( p/q \) with \( q < n+1 \) such that \( |\alpha - p/q| < 1/nq \), and so \( p/q \) cannot be equal to any \( p_i/q_i \).

Now to prove the statement, divide the unit interval into \( n \) intervals each of length \( 1/n \). The \( n+1 \) numbers \( c_1, \ldots, c_{n+1} \) must lie in these intervals (since \( \alpha \) is irrational, no \( c_i \) can lie on the boundary of an interval). So by the Doocot Principle, two of them, say \( c_i \) and \( c_j \), lie in the same interval; that is, \( |c_i - c_j| < 1/n \). Suppose that \( i > j \).

We have \( c_i = i\alpha - h_i \) for some integer \( h_i \); so

\[
|c_i - c_j| = |(i\alpha - h_i) - (j\alpha - h_j)| = |q\alpha - p|,
\]

where \( q = i - j \) and \( p = h_i - h_j \). Thus we have \( |q\alpha - p| < 1/n \), and so \( |\alpha - p/q| < 1/(nq) \). Finally, \( i, j \in \{1, \ldots, n+\}, \) so \( q = i - j \leq n \).

2.5. (a) I begin with something which is hopefully “obvious” geometrically: given a finite set of points in general position, we can find a subset of them which form a convex polygon containing the remainder in its interior.

Consider five points in general condition. If they form a convex 5-gon, choose any four, and we are done. If they form a convex 4-gon with one point in the
interior, again we are done. So suppose that the points form a triangle \((p_1, p_2, p_3)\) with two points \(p_4\) and \(p_5\) in its interior.

\[
\begin{tikzpicture}
\fill[pink] (0,0) circle (2pt) node[below] {\(p_1\)};
\fill[red] (1,2) circle (2pt) node[above] {\(p_3\)};
\fill[red] (3,0) circle (2pt) node[below] {\(p_2\)};
\fill[blue] (-1,1) circle (2pt) node[above] {\(p_4\)};
\fill[blue] (1,-1) circle (2pt) node[below] {\(p_5\)};
\draw (0,0) -- (1,2) -- (3,0) -- (0,0);
\end{tikzpicture}
\]

Since the points are in general position, the line \(p_4p_5\) does not pass through a vertex of the triangle, and so it must meet two sides, say \(p_1p_2\) and \(p_1p_3\): then \((p_2, p_4, p_5, p_3)\) is the required convex quadrilateral.

(b) If \(n\) points form a convex polygon, then none of them can lie in the interior of the triangle formed by three others, so all of the \(4\)-subsets form convex quadrilaterals.

Now suppose that we have at least \(R_4(5,n)\) points in general position. Colour the \(4\)-subsets red and blue as follows: a \(4\)-set is red if it forms a convex quadrilateral, blue otherwise. By part (a), there is no \(5\)-set all of whose \(4\)-sets are blue. So by Ramsey’s theorem, there is an \(n\)-set all of whose \(4\)-sets are red; by what was just said, it forms a convex \(n\)-gon.

2.6. The divisibility conditions state that, for \(i = 0, 1, \ldots, t - 1\), \(\binom{t+1-i}{t-i} = t+1-i\) divides \(\binom{2t+2-i}{t-i}\). Now

\[
\frac{1}{t+1-i}\binom{2t+2-i}{t-i} = \frac{1}{t+2}\binom{2t+2-i}{t+1-i},
\]

so we require that \(t+2\) divides \(\binom{2t+2-i}{t+1-i}\). But this binomial coefficient has \(t+2\) (a prime) as a factor in the numerator, which is not cancelled by any factor in the denominator; so the divisibility must hold.

2.7. This exercise is routine checking to show that any two points lie in a unique block.

2.8. So is this one.
**Remark** In each of these two exercises, it is legitimate to take the starting systems to be the “trivial” Steiner triple system (STS for short) with three points and a single block. Then Exercise 2.8 gives a 7-point STS, and Exercise 2.9 a 9-point STS. Then combining these we can produce systems of orders 15, 19, 21, 27, … Not all orders are covered by these constructions: e.g. we don’t obtain a 13-point STS. This has to be built directly!

3.1. Easiest to work backwards. We have

\[(1 - 4x)^{-1/2} = \sum_{n \geq 0} \binom{-1/2}{n} (-4)^n,\]

so our first task is to work out the binomial coefficient. We have

\[
\binom{-1/2}{n} = \frac{(-1)(-3) \cdots (-2n - 1)}{2^n n!} = (-1)^n \frac{1 \cdot 2 \cdot 3 \cdots (2n)}{2^n n! \cdot 2 \cdot 4 \cdots (2n)} = (-1)^n \frac{(2n)!}{(n!)^2} = (-1)^n \binom{2n}{n},
\]

so the coefficient of \(x^n\) is indeed \(\binom{2n}{n}\).

Now the convolution product on the left of the second display is the coefficient of \(x^n\) in \((1 - 4x)^{-1/2} \cdot (1 - 4x)^{-1/2} = (1 - 4x)^{-1}\), which is \(4^n\), as claimed.

3.2. In the equation \(e^{x+y} = e^x \cdot e^y\), the term of degree \(n\) on the left is \((x+y)^n/n!\). On the right we have the convolution product of two exponential series, so the term of degree \(n\) is

\[\sum_{k=0}^{n} \frac{x^k}{k!} \cdot \frac{y^{n-k}}{(n-k)!}.\]

So

\[\frac{(x+y)^n}{n!} = \sum_{k=0}^{n} \frac{x^k}{k!} \cdot \frac{y^{n-k}}{(n-k)!},\]

and multiplying by \(n!\) gives the Binomial Theorem (with an explicit formula for the coefficients).
3.3. Let’s do it in general: suppose that \( F_0 = c \) and \( F_1 = d \). Then we have
\[
A + B = c, \\
A\alpha + B\beta = d,
\]
which can be solved to get
\[
A = \frac{d - c\beta}{\alpha - \beta}, \quad B = \frac{c\alpha - d}{\alpha - \beta}.
\]
Now the solution to the problem is found by putting \( (c, d) = (0, 1), (1, 1) \) and \( (1, 2) \) respectively.

3.4. We’ve just done it!

3.5. Using the convolution product, and the fact that \( (1 - x)^{-1} = \sum_{n \geq 0} x^n \), the coefficient of \( x^n \) in the product \( A(x)(1 - x)^{-1} \) is
\[
\sum_{k=0}^{n} a_k \cdot 1,
\]
the \( n \)th partial sum of the sequence \( (a_n) \).

3.6. (a) Consider
\[
F(x) = (1 - x)^{-1}(1 - x^2)^{-1} \cdots = (1 + x + x^2 + \cdots)(1 + x^2 + x^4 + \cdots) \cdots
\]
We get a term in \( x^n \) from the product by choosing \( x_{a_1} \) from the first bracket, \( x^{2a_2} \) from the second, and so on, for every expression for \( n \) in the form \( n = a_1 + 2a_2 + \cdots \). But these expressions exactly correspond to partitions of \( n \); the one shown corresponds to taking \( a_1 1s, a_2 2s, \) and so on.

(b) We have
\[
F(x)^{-1} = (1 - x)(1 - x^2)(1 - x^3) \cdots
\]
Terms in \( x^n \) come from choosing \(-x^k\) from various brackets, where the values of \( k \) add up to \( n \). So we take all partitions of \( n \) into distinct parts. Now such a partition with an even number of parts contributes +1, while one with an odd number contributes −1. So
\[
b_n = (# \text{ partitions of } n \text{ into an even number of distinct parts})
\]
\[
- (# \text{ partitions of } n \text{ into an odd number of distinct parts}).
\]
(c) The first few cases are

\[
\begin{array}{ll}
1 & b_1 = -1 \\
2 & b_2 = -1 \\
3 = 2 + 1 & b_3 = 0 \\
4 = 3 + 1 & b_4 = 0 \\
5 = 4 + 1 = 3 + 2 & b_5 = +1 \\
6 = 5 + 1 = 4 + 2 = 3 + 2 + 1 & b_6 = 0 \\
7 = 6 + 1 = 5 + 2 = 4 + 3 = 4 + 2 + 1 & b_7 = +1 \\
\end{array}
\]

(If you play Killer Sudoku, you will be good at this!)

You might guess that all the coefficients are 0, +1 or −1, but you may not yet spot the pattern. We discuss this in Chapter 8.

3.7. We want a power series \( f(x) = \sum_{n \geq 1} a_n x^n \) with \( a_1 = 1 \) satisfying \( f(x) + f(x)^2 = x \).

If you recall the recurrence for the Catalan numbers (Proposition 3.5), you will see a connection. There we had \( F(x) = x + F(x)^2 \), and found that \( F(x) = \frac{1}{2}(1 - \sqrt{1 - 4x}) \). A little thought shows that \( f(x) = -F(-x) = \frac{1}{2}(\sqrt{1+4x} - 1) \).

Alternatively, regard the equation \( f + f^2 = x \) as a quadratic in \( f \), and solve it.

3.8. This was covered subsequently in lectures (see Chapter 5 of the notes).

3.9. The clown remains dry if and only if, at any stage, the number of red balls drawn does not exceed the number of blue balls drawn. So it is the ballot problem again, and the number of ways this can happen is

\[
C_{n+1} = \frac{1}{n+1} \binom{2n}{n}.
\]

But the total number of ways of drawing the balls is \( \binom{2n}{n} \), since there are \( 2n \) drawings and the red balls occur in \( n \) positions in the sequence. So the required probability is \( 1/(n+1) \).

A challenge here is to find an easy proof of this, not requiring the derivation of the formula for the Catalan numbers.

3.10. I will not give the solution here; it is not easy, but I think there are ample hints in the question.
4.1. It is clear that all the even values are equal, as are all the odd values; so the solution is

\[ a_n = \begin{cases} c & \text{if } n \text{ is even}, \\ d & \text{if } n \text{ is odd}. \end{cases} \]

The general method says: the generating function is \( p(x)/(1-x^2) \), where \( p \) is a linear polynomial. This can be written in partial fraction form as \( a/(1-x) + b/(1-x) \), and so the coefficient of \( x^n \) is

\[ a + b(-1)^n = \begin{cases} a + b & \text{if } n \text{ is even}, \\ a - b & \text{if } n \text{ is odd}. \end{cases} \]

Matching up these allows \( a \) and \( b \) to be expressed in terms of \( c \) and \( d \), and vice versa.

4.2.

\[
\begin{align*}
F_n^2 &= (F_{n-1} + F_{n-2})^2 \\
&= F_{n-1}^2 + F_{n-2}^2 + 2F_{n-1}F_{n-2} \\
&= F_{n-1}^2 + F_{n-2}^2 + 2(F_{n-2} + F_{n-3})F_{n-2} \\
&= F_{n-1}^2 + 3F_{n-2}^2 + 2F_{n-2}F_{n-3} \\
&= F_{n-1}^2 + 3F_{n-2}^2 + (F_{n-1}^2 - F_{n-2}^2 - F_{n-3}^2) \\
&= 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2.
\end{align*}
\]

If we assume that \( F_n^2 = aF_{n-1}^2 + bF_{n-2}^2 + cF_{n-3}^2 \), then using the first few values of \( F_n^2 \), namely 0, 1, 1, 4, 9, 25, we get

\[
\begin{align*}
4 &= a + b, \\
9 &= 4a + b + c, \\
25 &= 9a + 4b + c
\end{align*}
\]

and these equations can be solved to give the same result.

4.3. The generating function is \( p(x)/(1 - 5x + 6x^2) = c/(1 - 2x) + d/(1 - 3x) \), so the solution is \( a_n = c \cdot 2^n + d \cdot 3^n \).

For the first initial conditions we get \( c + d = 1, 2c + 3d = 0 \), giving \( c = 3, d = -2 \). For the second, \( c + d = 0, 2c + 3d = 1 \), giving \( c = -1, d = 1 \).

Now the recurrence relation is linear, and so a linear combination of solutions will again be a solution. Thus, to find the solution with \( a_0 = p \) and \( a_1 = q \), we add \( p \) times the first solution and \( q \) times the second to get

\[ a_n = (3p - q)2^n + (q - 2p)3^n. \]

120
(You may have seen this linearity principle in a course on differential equations . . . )

5.1. See Solutions 3.

5.2. There are plenty of exercises here, of widely varying difficulties. Here is the argument for dissections of a polygon.

Let $f(n)$ be the number of dissections of an $n$-gon; we have to prove that $f(n) = C_{n+1}$. Now choose a vertex $v$ of the $n$-gon, and divide the dissections up as follows:

- Dissections where $v$ lies on no diagonal. This means that the corner containing $v$ is cut off by a diagonal from the predecessor of $v$ to the successor of $v$. Removing this corner, we have a dissection of an $(n-1)$-gon, so there are $f(n-1)$ of this type.

- Taking $v = v_0$, suppose that the first diagonal from $v$, as we move in the positive direction round the polygon, joins it to $v_k$, where $2 \leq k \leq n-2$. This diagonal divides the original polygon into two parts, a $(k+1)$-gon and a $(n-k+1)$-gon, each dissected into triangles. There are $f(n-k+1)$ dissections of the $(n-k+1)$-gon. However, a bit of care is required for the $(k+1)$-gon, since there are no diagonals of the dissection containing $v$ (the diagonal we chose has become an edge of the new polygon). So, as in the first case, we have cut off $v$ and dissected the remaining $k$-gon, which is possible in $f(k)$ ways.

So

$$f(n) = f(n-1) + \sum_{k=2}^{n-2} f(k) f(n-k+1).$$

Using this it is not too hard to show, by induction, that $f(n) = C_{n+1}$.

5.3. We have (as in the method we used in the lecture) $D_n = E_{n+1}$, which implies that $xd(x) = e(x)$. Also, by considering the first return to the X-axis, we have

$$D_n = \sum_{k=1}^{n} E_k D_{n-k},$$

which gives the second relation.

From these relations we find that

$$e(x) = x + e(x)^2,$$

and since $E_0 = 0$, $E_1 = 1$, this gives $E_n = C_n$ for all $n$, from which $D_n = C_{n+1}$. 121
5.4. The hook lengths of the cells in a $2 \times n$ rectangle are $n+1, n, \ldots, 2$ (in the first row) and $n, n-1, \ldots, 1$ (in the second). Thus the hook length formula gives

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

5.5. A Wedderburn–Etherington object (WE-object for short) of size $n$ is obtained by combining one of size $k$ with one of size $n-k$, but we don’t care about the order.

If $n$ is odd, we can just take $k$ from 1 to $(n-1)/2$, and obtain

$$W_n = \sum_{k=1}^{(n-1)/2} W_k W_{n-k}.$$ 

If $k$ is even, we can take $k$ from 1 to $(n-2)/2$ to obtain all the terms except that with $k = n/2$. For these terms, we have to choose two WE-objects from a set of size $W_{n/2}$, with the order unimportant; this can be done in $\binom{W_{n/2}+1}{2} = (W_{n/2}^2 + W_{n/2})/2$ ways. Adding these gives the stated recurrence relation for $W_n$.

Now let $w(x)$ be the generating function. We see that $w(x)$ contains a term $x$ which does not occur in the right-hand side, so we add this in first. Now $w(x)^2/2$ counts all the quadratic terms correctly, leaving $\sum W_{n/2}x^n/2$ (summed over even $n$) which is $w(x^2)/2$.

If you are interested in how to proceed from here to a first look at the asymptotics: the equation we obtained for $w(x)$ is rather non-standard, because of the $w(x^2)$ term. But, roughly, the growth of $W_n$ is about $c^n$, where $1/c$ is the radius of convergence of the power series for $w(x)$. Now the equation can be written as a quadratic in $w(x)$, namely

$$w(x)^2 - 2w(x) + (w(x^2) + 2x),$$

with solution $w(x) = 1 - \sqrt{1 - 2x - w(x^2)}$ (perversely regarding $w(x^2)$ as known).

We seem to be no better off, but in fact we are. Inside the unit circle we have $|x| < 1$, and so $|x^2| < |x|$; so, when $x$ approaches the singularity of $w$, the function $w(x^2)$ is still safely inside its circle of convergence, and can be approximated well by the first few terms in its Taylor series.

So the procedure is: Calculate the first few terms using the recurrence (obtaining 1, 1, 1, 2, 3, 6, 11, 23, . . .). Then the singularity occurs when $1 - 2x - w(x^2) = 0$, and this equation can be solved numerically to give $x \approx 0.403 \ldots$. The reciprocal of this number is 2.483 . . ., so $W_n$ is very roughly $(2.483 \ldots)^n$. 

122
6.1. The proof is by induction on \( n \). The result is true when \( k = 0 \) or \( k = n \) (since the Gaussian coefficient is 1 in these cases). Assume the result for \( n - 1 \). Now look at the recurrence relation in Proposition 6.5. By the induction hypothesis, the first term on the right is a polynomial of degree \((k - 1)(n - k)\), while the second is a polynomial of degree \(k + k(n - 1 - k) = k(n - k)\). The second term is larger than the first, so the whole thing is a polynomial of degree \(k(n - k)\).

6.2. We divide the set of \( k \times n \) matrices in reduced echelon form with no zero rows into two classes:

- Matrices in which the leading one in the last row is in the last column. Such matrices have the form
  \[
  A = \begin{pmatrix}
  A' & 0 \\
  0 & 1
  \end{pmatrix},
  \]
  where \( A' \) is a \((k - 1) \times (n - 1)\) matrix in reduced echelon form. So there are \(q^{\binom{n-1}{k-1}}\) of these.

- The rest. In these, the last row is completely arbitrary, so they have the form
  \[
  A = (A^\dagger \ast),
  \]
  where \( \ast \) denotes a column whose entries are arbitrary, and \( A^\dagger \) is a \( k \times (n - 1) \) matrix in reduced echelon with no zero rows. So there are \( q^k \binom{n-1}{k} \) of these.

Adding gives the result.

This proof, of course, is valid only for prime power \( q \). But the Gaussian coefficients as defined by the formula, and the numbers of reduced echelon form matrices, coincide for all prime power \( q \) (since they satisfy the same recurrence and initial conditions) and are both polynomials in \( q \) (shown in the preceding question for the formula, and obvious for the numbers of matrices); so they must coincide as polynomials.

6.3. As we did in the lectures for \( n = 4 \) and \( k = 2 \), the reduced echelon matrices can be divided into \( \binom{n}{k} \) classes, each class determined by the set of \( k \) columns where the leading ones occur; the size of a class corresponding to a set \( A \) of columns is \( q^{m(A)} \), where \( m(A) \) is the number of asterisks in the figure, that is, the number of positions in the matrix which occur after the leading one in their row.
and not in the column of another leading 1. We “pair up” the sets $A$ in such a way that, if $A'$ is paired with $A$, then $m(A') = k(n - k) - m(A)$. Note that there are $k(n - k)$ positions in the matrix which are not in the column with index in $A$; of these, $m(A)$ lie after the leading one in their row, and so $k(n - k) - m(A)$ lie before the leading one. Now “turning the matrix upside down” interchanges these two sets of positions. That is, if we put $A' = \{ n + 1 - a : a \in A \}$ (the set of columns of the leading ones in the inverted matrix), then we do indeed have $m(A') = k(n - k) - m(A)$.

An alternative, and perhaps simpler, proof uses the interpretation counting area under lattice paths. Recall that if a lattice path $p$ from $(0, 0)$ to $(k, n - k)$ has an area $a(p)$ below it inside the $k \times (n - k)$ rectangle. The area above this path is $k(n - k) - a(p)$. Using this instead corresponds to rotating the path and rectangle $108^\circ$.

6.4. First note that there is a bijection between $k \times n$ matrices in reduced echelon form with no zero rows, and $n \times n$ matrices in reduced echelon form with rank $k$: just add $(n - k)$ zero rows.

For each matrix in reduced echelon form, there are $0 + 1 + \cdots + (k - 1) = k(k - 1)/2$ positions above the leading ones, which must be zero in reduced echelon matrices, but may be arbitrary in echelon form. So the number of matrices of rank $k$ in echelon form is

$$q^{k(k-1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

Now the $q$-binomial theorem with $x = 1$ shows that the total number of matrices in echelon form is

$$\sum_{k=0}^{n} q^{k(k-1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=1}^{n} (1 + q^{i-1}).$$

6.5. When we multiply out $(x + y)^n$, each term is a string of length $n$ of $x$s and $y$s. Such a string containing $k$ $y$s can be brought into the form $x^{n-k}y^k$ by jumping $y$s over $x$s, each such jump contributing a factor of $q$. So the coefficient of this term is $q^J$, where $J$ is the number of such jumps required. Let $C(n, k)$ be the sum of these coefficients over all strings with $k$ $y$s. Then $C(n, k)$ is a polynomial in $q$, and

$$(x + y)^n = \sum_{k=0}^{n} C(n, k)x^{n-k}y^k.$$  

We have to identify the coefficients $C(n, k)$ with Gaussian coefficients.
Clearly \( C(n, 0) = C(n, n) = 1 \), since there is just one term in each case and no jumps are needed. Also we have

\[
(x + y)^n = (x + y)^{n-1}(x + y) = \left( \sum_{k=0}^{n-1} C(n-1, k)x^{n-1-k}y^k \right) (x + y).
\]

Multiplying by \( x \), we have to jump this \( x \) over all \( k \) ys to reach the required form, giving a factor of \( q^k \). Multiplying by \( y \), no jumps are required. So we have

\[
C(n, k) = q^k C(n - 1, k) + C(n - 1, k - 1).
\]

Thus the coefficients \( C(n, k) \) satisfy the same recurrence and initial conditions as the Gaussian coefficients, and so are equal to them.

6.6. \( |X| \) is the number of 1-dimensional subspaces of a \( n \)-dimensional vector space, and is equal to \( \binom{n}{1}_q = (q^n - 1)/(q - 1) \). Similarly, the cardinality of each block is \( \binom{2}{1}_q = q + 1 \).

Given any two points, or distinct 1-dimensional subspaces \( U_1 \) and \( U_2 \) of \( V \), their span \( U_1 + U_2 \) is 2-dimensional, and is the unique 2-dimensional subspace containing \( U_1 \) and \( U_2 \).

So we have a Steiner system, as claimed.

7.1. The question doesn’t specify the value of \( g(0) \), so I will take it to be 1. (The first part of the argument below shows that we need to do this.)

We are given the formula, so we can verify it by induction. For \( n = 0 \) there is a single term in the sum, namely \( 0!g(0)/0! = 1 \). [This is why we have to take \( g(0) = 1 \).] Suppose that the result is true for \( n - 1 \). Then

\[
f(n) = nf(n-1) + g(n)
\]

\[
= n(n-1)! \left( \sum_{i=0}^{n-1} \frac{g(i)}{i!} \right) + n! \frac{g(n)}{n!}
\]

\[
= n! \left( \sum_{i=0}^{n} \frac{g(i)}{i!} \right),
\]

as required.
7.2. There is one word made from no letters, namely the empty word.

If $F(n - 1)$ words can be made from $n - 1$ letters. If now $n$ letters are available, then one possibility is the empty word; any other word starts with some letter $a$, which is followed by a word made from the remaining $n - 1$ letters. So $F(n) = 1 + nF(n - 1)$, as required.

Applying the result of the preceding exercise with $g(n) = 1$ for all $n$, we see that

$$F(n) = F(n) = n! \left( \sum_{i=0}^{n} \frac{1}{i!} \right),$$

which is the truncated exponential series multiplied by $n!$. So we have

$$F(n) - n!e = \sum_{i \geq n+1} \frac{n!}{i!},$$

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

$$\leq \frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots,$$

$$= \frac{1}{n},$$

summing a geometric series in the last step. So $F(n) - n!e < 1/2$ for $n \geq 1$, which shows that $F(n)$ is the nearest integer to $n!e$.

7.3. We use the rule for multiplying the series $\exp(-x)$ and $(1 - x)^{-1}$. The coefficient of $x^n$ is

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

which agrees with our inclusion-exclusion formula for $d(n)/n!$.

For the alternative proof, the statement about choosing a permutation is clearly true: every permutation is produced exactly once by this procedure. So the displayed formula holds. Now, if $D(x)$ is the exponential generating function for the derangement numbers $d_n$, then the coefficient of $x^n$ in $D(x) \exp(x)$ is

$$\sum_{k=0}^{n} \frac{1}{k!} \cdot \frac{d_{n-k}}{(n-k)!} = \frac{n!}{n!} = 1,$$

and so $D(x) \exp(x) = \sum_{n \geq 0} x^n = 1/(1 - x)$. 126
7.4. A tree (as defined) has \( n - 1 \) edges and one component. Each time an edge is removed, the number of components increases by one; so the sum of the numbers of edges and components is always \( n \).

For one proof, note that a set of \( k \) edges in a tree forms a graph with \( n - k \) connected components. So our formula for the chromatic polynomial gives

\[
\sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} q^{n-k} = q(q-1)^n.
\]

For the other proof, we can list the vertices of a tree (starting at a leaf) in such a way that each vertex has only one neighbour among its predecessors in the list. So, to colour the tree with \( q \) colours, we have \( q \) choices for the first leaf, and then \( q-1 \) choices for each subsequent vertex, giving the result \( q(q-1)^{n-1} \).

7.5. This example is much easier using inclusion-exclusion than using deletion-contraction. The whole graph has just one connected component, but if we remove one edge then we have a tree, so a set of \( k \) edges has \( n - k \) components if \( k < n \). Thus the formula gives

\[
P_{C_n}(q) = (-1)^n q + \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} q^{n-k} = (q-1)^n + (-1)^n (q-1) = (q-1)((q-1)^{n-1} - (-1)^{n-1}).
\]

As a sanity check, if \( n \) is even, the \( n \)-cycle has just two colourings with two colours, while if \( n \) is odd, it has none.

7.6. We know that \( \mu(x,x) = 1, \mu(x,y) = -1 \) if \( y \) covers \( x \) (i.e. \( y \) is immediately above \( x \) in the diagram), and \( \mu(x,y) = 0 \) if \( s \not\leq y \). The only case that remains is when \( x \) is the bottom element and \( y \) is the top; then

\[
\mu(x,y) = -\mu(x,x) + \sum_{i=1}^{n} \mu(x,z_i),
\]

where \( z_1, \ldots, z_n \) are the intermediate points; so

\[
\mu(x,y) = -(1-n) = n-1.
\]
7.7. There is a lot to do to answer this question completely!

First, of course, \( \mu(a, b) = 0 \) if \( a \) and \( b \) are partitions for which \( a \) does not refine \( b \); and \( \mu(a, a) = 0 \).

Now, suppose that \( b \) has \( m \) parts \( P_1, \ldots, P_m \), and that \( a \) splits the \( i \)th part \( P_i \) into \( r_i \) pieces. Then the interval between \( a \) and \( b \) is isomorphic to the Cartesian product of the partition lattices corresponding to sets of sizes \( r_1, \ldots, r_m \). So we only have to calculate \( \mu(a, b) \) in the case where \( b \) has a single part and \( a \) is the partition into singletons. This a number \( f(n) \) which depends only on \( n \), the number of points.

For \( n = 1, 2, 3 \), we find that the poset is a single point, a chain of length 1, and the poset of question 7.6 with 3 intermediate elements respectively, so the values of \( f(n) \) are 1, -1, 2 respectively.

For \( n = 4 \), things start getting more complicated. The poset has 15 elements; as well as the bottom \( a \) and top \( b \), there are 6 partitions into 3 parts (each one \( x \) of these covers \( a \), so \( \mu(a, x) = -1 \)); and 7 partitions into two parts (four of type 1|234 and three of type 12|34); the product rule gives \( \mu(a, x) = 2 \) for the first type, \((-1)^2 = 1\) for the second. So

\[
\mu(a, b) = -(1 - 6 + 4 \cdot 2 + 3 \cdot 1) = -6.
\]

Perhaps this is enough data for you to make a guess at the result – can you prove your guess? [See Chapter 9 for the solution.]

7.8. Let \( b_I \) be the number of elements lying in \( A_i \) for \( i \in I \) and not for \( i \notin I \), and similarly for \( b'_I \). Suppose without loss that \( a_N > a'_N \), and let \( a_N - a'_N = \varepsilon \). Then a backwards induction shows that \( a_I - a'_I = (-1)^{n-|I|}\varepsilon \). Now all the sets \( I \) with \( |I| \equiv n \, (\text{mod} \, 2) \) satisfy \( a_I = a'_I + \varepsilon \geq \varepsilon \). There are \( 2^{n-1} \) such sets, all mutually disjoint, so containing at least \( 2^{n-1}\varepsilon \) points altogether. But their union is contained in \( X \), so \( 2^{n-1}\varepsilon \leq |X| \), as required.

7.9. Each of the \( n \) players chooses at random one of the remaining \( n - 1 \), so there are \( (n-1)^n \) possible configurations, which we assume are all equally likely.

How many choices will result in \( k \) pairs \( \{a_i, b_i\} \) looking at each other for \( i = 1, \ldots, k \) (and maybe others)? The remaining \( n - 2k \) players could look at anyone, so there are \( (n-1)^{2k} \) such choices.

How many ways can we choose \( k \) disjoint pairs? These involve altogether \( 2k \) players, who can be chosen in \( n(n-1) \cdots (n-2k+1) = (n)_{2k} \) ways (in order). We pair up the first and second, the third and fourth, and so on. But the same pairing could result from \( 2^{k}k! \) different choices of \( 2k \) players, since we could choose the pairs in any of the \( k! \) orders, and the players within each pair in either of 2 possible orders.
So there are \((n - 1)^{n-2k}(n)_{2k}/2^k k!\) choices in which at least a given collection of \(k\) pairs are looking at each other (and maybe more).

PIE gives the number of configurations in which no pairs are looking at each other as

\[
\sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k n(n-1)^{n-2k}(n)_{2k}/2^k k!,
\]

and hence the number in which at least one pair are looking at each other is

\[
\sum_{k=1}^{\lfloor n/2 \rfloor} (-1)^{k-1} n(n-1)^{n-2k}(n)_{2k}/2^k k!,
\]

since the \(k = 0\) term is just \((n - 1)^n\). Dividing this number by \((n - 1)^n\) gives the required probability.

For \(n = 2, 3, 4, 5\), the probabilities are 1, 3/4, 17/27, 145/256 respectively.

8.1. This is covered in Chapter 9 of the notes. It follows from the fact that the radius of convergence of a power series \(\sum a_n z^n\) is the reciprocal of the distance from the origin to the nearest singularity. The series \(\sum p(n) z^n\) is equal to the infinite product \(\prod (1-z)^{-1}\), which is analytic inside the unit disc but has singularities at all roots of unity.

8.2. Euler’s Pentagonal Numbers Theorem gives us a recurrence relation which, although it is linear with constant coefficients, does not have finite length.

In fact, we know that the leading term in the asymptotics of a C-finite sequence is \(n^{k-1} \alpha^n\), where \(\alpha\) is the root with largest modulus of the characteristic polynomial. If the partition numbers were C-finite, then \(\alpha\) would be 1 (by the preceding question), and so the growth would be like a power of \(n\). But this is not the case.

8.3. Make the substitution suggested and see what happens!

8.4. (a) \(p(n,k)\) is the number of Young diagrams with first column of length \(k\). So it is equal to the number of Young diagrams with first row of length \(k\), which is the number of partitions with largest part \(k\).

(b) These partitions have the form \((n-r,r)\), where \(r\) runs from 1 to \([n/2]\) (since \(n-r \geq r\)).

(c) These partitions have the form \((n-r-s,r,s)\) with \(n-r-s \geq r \geq s\). Then \(r-s \leq [2n/3]\), and for each value, we have \([\lfloor (r+s)/2 \rfloor]\) values of \(r\) and \(s\). So the required number is

\[
\sum_{k=2}^{\lfloor 2n/3 \rfloor} \lfloor k/2 \rfloor.
\]
Careful analysis shows that this is a quadratic polynomial for each residue class of \( n \mod 8 \).

Alternatively, use a special case of the next result.

(d) It is slightly easier to deal with partitions with largest part at most \( k \), since the sum or difference of two PORC functions is PORC. Now the generating function for these partitions is \( \prod_{i=1}^{k} (1 - x^i)^{-1} \), so the numbers are the solutions of a linear recurrence relation of length \( 1 + 2 + \cdots + k = k(k+1)/2 \). All the roots of the denominator have modulus 1, and indeed are \( k! \)-th roots of unity; so the solution has the form

\[
\sum_{i=0}^{k} k!p_i(n)\omega^i,
\]

where \( \omega \) is a primitive \( k! \)-th root of unity. It is easy to see that this is PORC, since the powers of \( \omega \) return to their original values when \( n \) increases by \( k! \).

9.1. For a partition with two parts, pick any subset of \( \{1, \ldots, n\} \) other than the empty set and the whole set to be the first part; its complement is the second. We have to divide by two since we could have chosen the parts in the other order. So

\[
S(n, 2) = \frac{(2^n - 2)}{2} = 2^{n-1} - 1.
\]

For a partition with \( n - 1 \) parts, one part must have size 2 and all the rest are singletons. There are \( n(n-1)/2 \) ways to choose the part of size 2, and then everything is determined. So \( S(n, n-1) = n(n-1)/2 \). There are two kinds of partitions with \( n - 2 \) parts:

- one part of size 3 and all the others singletons; there are \( n(n-1)(n-2)/6 \) of these.
- two parts of size 2 and all the others singletons. There are \( n(n-1)(n-2)(n-3)/24 \) ways of choosing a set of four points, and three ways of splitting it into two pairs, so there are \( n(n-1)(n-2)(n-3)/8 \) of these.

Summing gives the result.

9.2. A partition with \( n - 1 \) cycles is a transposition; there are \( n(n-1)/2 \) of these, and the sign is \( (-1)^{n-(n-1)} = -1 \). A partition with \( n - 2 \) cycles is either a 3-cycle or the product of two 2-cycles. The number of 3-cycles is twice the number of 3-sets, since there are two cycles on each 3-set; so there are \( n(n-1)(n-2)/3 \) of these. In the other case, the answer is the same for permutations as for partitions, that is, \( n(n-1)(n-2)(n-3)/8 \). Adding gives \( n(n-1)(n-2)(3n-1)/24 \).
9.3. This exercise really refers to Chapter 8 where we discussed conjugacy in the symmetric group. The general argument goes like this. Recall that \( g \) and \( h \) are conjugate in a group \( G \) if \( x^{-1}gx = h \) for some \( x \in G \).

- \( 1^{-1}g1 = g \), so \( g \) is conjugate to \( g \).
- If \( g \) is conjugate to \( h \), then there exists \( x \) with \( x^{-1}gx = h \). Then \( (x^{-1})^{-1}hx^{-1} = g \), so \( h \) is conjugate to \( g \).
- Suppose that \( g \) is conjugate to \( h \), and \( h \) to \( k \). Then there exist \( x \) and \( y \) with \( x^{-1}gx = h \) and \( y^{-1}hy = k \). So \( (xy)^{-1}g(xy) = k \), using the associative law and the fact that \( (xy)^{-1} = y^{-1}x^{-1} \). Thus \( g \) is conjugate to \( k \).

9.4. (a) The easiest proof uses the fact stated without proof in the notes, that

\[
\sum_{n \geq 0} S(n,k)\frac{x^n}{n!} = \frac{(\exp(x) - 1)^k}{k!}.
\]

[Can you prove this?]

Now, if \( b_n = \sum_{k=1}^{n} S(n,k)a_k \), then

\[
B(x) = \sum_{n \geq 1} \frac{b_n x^n}{n!}
\]

\[
= \sum_{n \geq 0} \sum_{k=1}^{n} S(n,k)a_k \frac{x^n}{n!}
\]

\[
= \sum_{k \geq 1} a_k \left( \sum_{n \geq k} \frac{S(n,k)x^n}{n!} \right)
\]

\[
= \sum_{k \geq 1} a_k (\exp(x) - 1)^k
\]

\[
= A(\exp(x) - 1).
\]

The argument reverses.

(b) This is easiest done by using the inverse relation between the two types of Stirling numbers. Thus,

\[
\left( b_n = \sum_{k=1}^{n} S(n,k)a_k \right) \iff \left( a_n = \sum_{k=1}^{n} s(n,k)b_k \right).
\]
Now the first equation is equivalent to $B(x) = A(\exp(x) - 1)$, which by the inverse relation between the exponential and logarithm functions is equivalent to $A(x) = B(\log(1 + x))$.

10.1. It follows from Stirling’s formula and the continuity of the logarithm function that

$$\log n! \sim n \log n - n.$$ 

Thus

$$\log \left( \frac{n}{pn} \right) = \log n! - \log(pn)! - \log((1 - p)n)!$$

$$= n \log n - n - pn \log p - pn \log n + pn - (1 - p)n \log(1 - p) - (1 - p)n \log n + (1 - p)n$$

$$= n(-p \log p - (1 - p) \log(1 - p)),$$

as required.

10.2. The radius of convergence is the reciprocal of the lim sup of $(T_n/n!)^{1/n}$. We have

$$\left( \frac{T_n}{n!} \right)^{1/n} \sim (2\pi n)^{-1/n} n^{-2/n} \frac{n}{n/e}$$

$$\rightarrow e$$

as $n \rightarrow \infty$, since $n^{1/n} \rightarrow 1$ as $n \rightarrow \infty$. [Can you prove this?] So the radius of convergence is $1/e$, as required.


12 Revision problems

The final section contains some further problems.

1. Let \( n, k, l \) be positive integers with \( n \geq k \geq l \). Prove that

\[
\binom{n}{k} \binom{k}{l} = \binom{n}{l} \binom{n-l}{k-l}
\]

(a) by using the formula for binomial coefficients;
(b) by a counting argument;
(c) by calculating a term of \((x + (y + z))^n = ((x + y) + z)^n\).

2.

(a) In how many ways can 12 sweets be distributed among 5 children?
(b) In how many ways can this be done if each child is to have at least one sweet?
(c) [Harder!] In how many ways can this be done if each child is to have at least one and at most three sweets?

3. By considering \((1 + \omega)^n\), where \( \omega = (-1 + \sqrt{-3})/2 \) is a cube root of unity, calculate

\[
\sum_{i=0}^{\lfloor n/3 \rfloor} \binom{n}{3i}
\]

(Hint: Separate the congruence classes for \( n \mod 6 \).)

4. Prove the following:

(a) \( \sum_{i=0}^{k} \binom{n+i}{i} = \binom{n+k+1}{k} \).

(b) \( \sum_{k=1}^{n} k \binom{n}{k} = n2^{n-1} \).

5. Prove the following about Fibonacci numbers. I take the convention that \( F_0 = F_1 = 1 \). Then \( F_n = F_{n-1} + F_{n-2} \) for \( n \geq 2 \).

(a) The number of compositions (ordered sums) for \( n \) as a sum of 1s and 2s is \( F_n \).
(b) Given \( n \) seats in a line, the number of ways of choosing a subset of them including no two consecutive positions is \( F_{n+1} \). If the \( n \) seats are in a circle (with \( n \geq 3 \), the number is \( F_n + F_{n-2} \).

(c) \( F_n^2 - F_{n-1}F_{n+1} = (-1)^{n-1} \).

(d) \( \sum_{k=0}^{n} F_k = F_{n+2} - 1 \).

(e) \( \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i} = F_n. \) [Hint: Part (a).]

6. Show that the number of ways of covering a \( 2 \times n \) board with dominoes (\( 1 \times 2 \) rectangles) is \( F_n \). [Harder] What about a \( 3 \times n \) board?

7. Show that the sequence with \( n \)th term \( a_n = F_n^2 \) satisfies a linear recurrence relation in which the \( n \)th term depends on the three preceding terms for \( n \geq 3 \).

8. Let \( X = \mathbb{Z}/(13) \) (the integers mod 13), and let \( B_1 = \{0, 3, 4\} \) and \( B_2 = \{0, 6, 8\} \). Show that every non-zero element \( x \in X \) can be written uniquely as \( x = y - z \) where \( y, z \in B_i \) for some \( i \). Hence show that the 26 translates of \( B_1 \) and \( B_2 \) mod 13 are the blocks of a Steiner system \( S(2, 3, 13) \).

9. This exercise shows that \( R_2(3, 3, 3) = 17 \). We have \( R_2(3, 3, 3) \leq 17 \), so we need to construct a colouring of the 2-subsets of a 16-set with three colours and no monochromatic triangles.

   Take \( X \) to be the set of all complementary pairs \( \{x, \bar{x}\} \) of binary vectors of length 5. (For example \( \{11010, 00101\} \)). There are \( 2^5/2 = 16 \) pairs. Note that mod-2 addition is well defined on these pairs: if \( x + y = z \), then \( \bar{x} + \bar{y} = \bar{z}, \bar{x} + \bar{y} = \bar{z} \), etc. Join two pairs by a red edge if their sum is a pair of vectors with 1 and 4 ones; a green edge if their sum is a pair of vectors with 2 and 3 ones, and the 2 ones (or zeros) are consecutive in the cyclic order; and by a blue edge otherwise.

   For example, the edge from \( \{11010, 00101\} \) to \( \{11110, 00001\} \) is red; from \( \{11010, 00101\} \) to \( \{10110, 01001\} \) is green; and from \( \{11010, 00101\} \) to \( \{01110, 10001\} \) is blue.

   Show that there is no monochromatic triangle. (Hint: use symmetry to show that, if there is such a triangle, then there is one containing \( \{00000, 11111\} \).)

10. Prove that

\[
S(n,k) = \frac{1}{k!} \sum_{i=0}^{k-1} (-1)^i \binom{k}{i} (k-i)^n.
\]

Check this formula directly for \( k \leq 3 \).
11. Prove that the number of permutations of \( \{1, \ldots, n\} \) having \( a_i \) cycles of length \( i \) for \( i = 1, 2, \ldots \) is
\[
\frac{n!}{1^{a_1} a_1! 2^{a_2} a_2! \cdots}.
\]
For which value of \( k \) is the number of \( k \)-cycles (fixing the other \( n-k \) points) maximum?

The next questions are based on work of Abdullahi Umar, and show that the orders of certain inverse semigroups are given by combinatorial numbers that we have met in this course.

Let \( X = \{1, \ldots, n\} \). A **partial permutation** of \( X \) is a bijective map from a subset of \( X \) to another subset of the same cardinality. The set of all partial permutations forms an **inverse semigroup** (under the operation of composition: if \( f \) and \( g \) are partial permutations, then \( f \circ g \) is defined on \( \text{Range}(f) \cap \text{Domain}(g) \cdot f^{-1} \), the set of points in the domain of \( f \) whose images under \( f \) lie in the domain of \( g \). [Maps are written on the right here, but this has no effect on the counting problems.] The semigroup of all partial permutations is denoted by \( P(X) \) and called the **symmetric inverse semigroup**.

A partial permutation is **monotone** if \( x < y \) implies \( xf < yf \) (where \( < \) refers to the usual order on \( \{1, \ldots, n\} \), and is **decreasing** if \( xf \leq x \) for all \( x \in \text{Domain}(f) \). We let \( P_m(X), P_d(X) \) denote the sets of monotone and decreasing maps respectively, and \( P_{md}(X) = P_m(X) \cap P_d(X) \). Moreover, \( P_s(X) \) denotes the set of **strictly decreasing** partial permutations (satisfying \( xf < x \) for all \( x \in \text{Domain}(X) \)), and \( P_{ms}(X) = P_m(X) \cap P_s(X) \). All of these sets are sub-semigroups of \( P(X) \). [Can you prove this?]

For proofs of the results in the next exercise, and much more of the same kind, see the paper “Combinatorial results for semigroups of order-preserving partial transformations” by A. Laradji and A. Umar, *J. Algebra* 278 (2004), 342–459. Further generalisations are suggested in my notes on inverse semigroups on the course web page.

12. (a) Show that \( |P(X)| = \sum_{k=0}^{n} \binom{n}{k}^2 k! \).

(b) Show that \( |P_m(X)| = \binom{2n}{n} \), the central binomial coefficient.
(c) Show that $|P_d(X)| = B(n+1)$ and $|P_s(X)| = B(n)$, where $B(n)$ is the $n$th Bell number.

(d) Show that $|P_{md}(X)| = C_{n+1}$ and $|P_{ms}(X)| = C_n$, where $C_n$ is the $n$th Catalan number.

13. Let $M(n,q)$ be the semigroup of linear maps on an $n$-dimensional vector space $V$ over a field of order $q$, and $P(n,q)$ be the semigroup of bijective linear maps between subspaces of $V$ (the linear analogue of the symmetric inverse semigroup). Prove that $|M(n,q)| = |P(n,q)|$, and write a formula for this number.

14. Let $P$ denote the partially ordered set of subgroups of the symmetric group $S_4$, ordered by inclusion. Calculate $\mu(\{1\}, S_4)$.

15. Prove that the number $p(n)$ of partitions of $n$ grows faster than any polynomial in $n$.

16. Let $B$ be the infinite matrix whose $(n,k)$ entry is $\binom{n}{k}$ for $n,k \geq 0$. What are the entries of $B^2$?

17. A total preorder of a set $X$ is a relation $\leq$ on $X$ satisfying

- for all $x,y \in X$, either $x \leq y$ or $y \leq x$ (or maybe both);
- the transitive law holds: for all $x,y,z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

Show that the relation $\equiv$ given by $x \equiv y$ if $x \leq y$ and $y \leq x$ is an equivalence relation, whose equivalence classes are totally ordered by the given preorder.

Hence show that the number of preorders on a set is given by

$$P(n) = \sum_{k=1}^{n} S(n,k)k!.$$ 

Prove also that

$$P(n) = \sum_{k \geq 1} \frac{k^n}{2k+1}.$$ 

18. A function $F$ on the natural numbers is said to be multiplicative if

$$\gcd(m,n) = 1 \Rightarrow F(mn) = F(m)F(n).$$
(a) Suppose that $F$ and $G$ are multiplicative. Show that the function $H$ defined by

$$H(n) = \sum_{k|n} F(k)G(n/k)$$

is multiplicative.

(b) Show that the Möbius and Euler functions are multiplicative.

(c) Let $d(n)$ be the number of divisors of $n$, and $\sigma(n)$ the sum of the divisors of $n$. Show that $d$ and $\sigma$ are multiplicative.