MATH1081
DISCRETE MATHEMATICS

Section 4

Combinatorics

Counting is probably the earliest mathematical operation you learn, and it might seem rather strange to have part of a university course devoted to it. (In fact, it is easy to find enough material in this topic for a whole course at virtually any level you choose!) We are interested in counting objects in cases where we cannot just line them up and recite "one, two, three...", either because there are simply too many of them or because we want an answer in terms of an unknown parameter. The study of counting, in this sense, is often referred to as **enumeration** or **combinatorics**. Here are some examples of important combinatorial problems in mathematics and other fields.

- What is the chance of winning Lotto?
- How many different saturated hydrocarbon molecules can be formed containing n carbon atoms?
- How many ways are there to colour the faces of a cube, if three different colours are available?
- Given a list of cities and the airline routes between them, find the number of possible trips which visit each city once and return to the starting point.
- Given a string of 0s and 1s, how many strings of 0s and 1s differ from it in at most two places?

Problem. A restaurant menu lists seven different main courses and five different desserts.

- If I am going to order a main course or a dessert (but not both), how many options do I have?
- If I am going to order a main course and a dessert, how many options do I have?

Answers.

-
-

These problems illustrate two principles which underlie even the most sophisticated counting methods. Suppose that a choice can be made from m options, and another choice can be made from n options. Then

- the number of ways to make *either* the first choice *or* the second is $m + n$, assuming that all the options for the two choices are different;
- the number of ways of making the first choice *and* the second is mn, assuming that all combinations of the two choices are permissible.

Comments.

- **Note** that each of the above involves assumptions on any connections between the first and second choices. Here are some examples which violate these assumptions.
    - In term 2 last year 682 students studied MATH1081 and 511 studied MATH1231. The number of ways to choose a student who studied either MATH1081 or MATH1231 is *not* $682 + 511$. Why not?
    - The School of Mathematics offers 9 different first year courses in first term and 7 in second term. The number of ways in which a student can do one course from each term is *not* 9×7. Why not?
- We will sometimes expect you to use common sense in solving counting problems! For instance, in the problem on page 3 we would expect you to realise that a main course would not (usually?) be eaten for dessert (so there is no overlap between the two parts of the menu).

Problem. We wish to construct three–letter words from the twenty six letters of the English alphabet. How many possibilities are there, if

(a) any choice of letters is permitted;

(b) the word must contain one vowel;

(c) the word must contain *at least* one vowel;

(d) the word must come before EGG in alphabetical order?

Note that for this course a word is any string of letters from a given alphabet. It does not matter whether or not the word is in the dictionary, or whether or not it makes sense. For example, if we have an alphabet { A,C,T } consisting of just three letters, some of the "words" we can form are

CAT and TAC and CCTTT and ATACTCC .

Solution.

(a) We set out the working step by step.

1. Choose the first letter .... 26 ways
2. Choose the second letter .... 26 ways
3. Choose the third letter .... 26 ways

The total number of words is $26 \times 26 \times 26$, that is, 26^3.

(b) Note that we must choose not only the letters, but also where they go.

1. Choose a place for the vowel .... 3 ways
2. Choose the vowel .... 5 ways
3. Choose the first consonant .... 21 ways
4. Choose the second consonant .... 21 ways

The number of possible words is $3 \times 5 \times 21^2$.

(c) The number of words with *exactly* one vowel is $3 \times 5 \times 21^2$, as above. Next we count the words with two vowels,

1. Choose a place for the consonant .... 3 ways
2. Choose the consonant .... 21 ways
3. Choose the first vowel .... 5 ways
4. Choose the second vowel .... 5 ways

giving $3 \times 21 \times 5^2$ words, and those with three vowels,

1. Choose the first vowel .... 5 ways
2. Choose the second vowel .... 5 ways
3. Choose the third vowel .... 5 ways

giving a further 5^3. A word contains at least one vowel if it contains one **or** two **or** three, so the final answer is obtained by *adding* these three partial totals. The number of words is

$$3 \times 5 \times 21^2 + 3 \times 21 \times 5^2 + 5^3 .$$

Alternative solution. As in (a), the total number of words with three letters is 26^3. However we do not wish to count the words with *no* vowels, and there are 21^3 of these (why?). So the number of words with at least one vowel is

$$26^3 - 21^3 .$$

(d) A three–letter word comes before EGG in the dictionary if it begins with A, B, C or D and has any two further letters.

1. Choose the first letter, A to D . . . . 4 ways
2. Choose any second letter . . . . 26 ways
3. Choose any third letter . . . . 26 ways

But a word beginning with E will also come before EGG if its second letter is before G.

1. Choose the first letter E . . . . 1 way
2. Choose the second letter, A to F . . . . 6 ways
3. Choose any third letter . . . . 26 ways

And a word starting with EG will still be counted if the third letter is from A to F.

1. Choose the first letter E . . . . 1 way
2. Choose the second letter G . . . . 1 way
3. Choose the third letter, A to F . . . . 6 ways

Since we can choose a word in the first category **or** the second **or** the third, the total number of words to be counted is

$$4 \times 26^2 + 6 \times 26 + 6 \ .$$

Comments.

- For counting problems, always be sure to set out your working clearly, listing the choices to be made and the number of options in each case, then combine the subtotals by means of addition and/or multiplication to obtain the final answer.
- The two alternative answers for part (c) may be surprising as they do not appear to be the same! In fact they are equal, as you can confirm with the aid of a calculator. **Exercise**. Using basic algebra, confirm *without* using a calculator that the two answers are the same.
- Usually we would not solve a counting problem by going all the way back to addition and multiplication; we shall develop techniques by which substantial parts of a problem can be solved in a single step. The above problems illustrate two of these techniques.

Ordered selections with repetition. Suppose that we have n different items and we wish to make r selections from these items, where

- the same item may be selected more than once (for example, AABA is a permissible choice of four letters);
- the order of the items is significant (for example, ABCA and CBAA are regarded as different and therefore are both counted).

Then the number of possible choices is n^r.

Example. There are 100 students in a course. Each is to choose one of three elective subjects. There are no limits on the number of students taking each elective. In how many different ways can the choices be made?

Solution. Line up all the students in alphabetical order and ask them to choose an elective one by one. Each has three options to choose from, and different students are allowed to choose the same option. It makes a difference which students choose which elective, so the order of the choices is important. Therefore there are 3^{100} possibilities.

Subtraction. If a choice can be made in m different ways altogether, but n of these are not permitted for a particular problem, then the number of permissible ways to make the choice is $m - n$.

Example. A student has to choose a term 1 elective from a list of 12 courses and a term 2 elective from a list of 15. However these figures include 7 courses which are offered in both terms, and the student is not allowed to take the same course twice. How many options are open to the student?

Solution. Taking one course from each list,

1. Take a T1 course . . . . 12 ways
2. Take a T2 course . . . . 15 ways

there are $12 \times 15 = 180$ options. But this includes forbidden possibilities in which the same course is selected twice:

1. Take a course common to both lists . . . . 7 ways
2. Take the same course again . . . . 1 way

These 7 possibilities must be subtracted, so the actual number of options is 173.

Comment. We have assumed that if there are two courses A and B which can be taken in either term, then "A in T1, B in T2" and "B in T1, A in T2" are different options and are both counted. If we regarded these as the same option we would not want to count it twice and the problem would be different.

Problem. How many ten–letter words can be made from the English alphabet which contain at least one X, at least one Y and at least one Z?

Solution.

1. Choose a place for the X .... 10 ways
2. Choose a place for the Y .... 9 ways
3. Choose a place for the Z .... 8 ways
4. Choose another seven letters from 26 possibilities, with repetitions allowed and order significant .... 26^7 ways

So the number of words is $10 \times 9 \times 8 \times 26^7$.

But this solution is **completely wrong!!** Why?

We shall see later how to solve the preceding problem correctly.

There are two important principles that you need to pay attention to when solving combinatorial problems.

- Make sure you count everything!
- Make sure you don't count anything twice!

These may sound completely obvious, but, as the previous example illustrates, they are not always easy to put into practice.

Problem. Re–do the problems from page 5 if we add an extra condition: no letter may be used more than once in a word.

Solutions.

Ordered selections without repetition. Suppose we have n different items and we wish to select r of these, where

- no item may be selected more than once (for example, XYZZY is not allowed);
- the order of the items is significant (for example, XYZ, XZY, YXZ, YZX, ZXY and ZYX are all regarded as different and therefore are counted as six different words).

Such a selection is referred to as a **permutation** of r objects chosen from n (different) objects, and the number of possible choices is

$$P(n,r) = n(n-1)(n-2)\cdots(n-(r-1)) \ .$$

Example. A band of pirates has 13 members, including the captain's parrot. A crew photo is to be taken with seven pirates standing in a row and six sitting on chairs in front of them. Suppose that we don't care who stands where, but we do care who sits on which chair. In how many possible ways can pirates be seated on the chairs? What if the captain and the parrot must sit on the middle two chairs?

Solution. For the first problem there are thirteen different pirates and we wish to select six of them, with repetitions not allowed and order important. This can be done in $P(13, 6)$ ways.

For the second, note that in a row of six there are two middle chairs.

1. Choose one of the middle chairs for the captain .... 2 ways
2. Choose the other for the parrot .... 1 way
3. For the other chairs, choose 4 pirates from 11,
   with repetition not allowed and order important.................. $P(11, 4)$ ways

The number of arrangements is $2P(11, 4)$.

Factorials. A particularly important case of permutations is when we wish to find the number of ordered selections of *all* available objects, taken once each. If there are n objects, the number of such selections is

$$P(n,n) = n(n-1)(n-2)\cdots 3 \times 2 \times 1 = n!\ .$$

For any n, r, we can write $P(n,r)$ in terms of factorials:

$$\begin{aligned} P(n,r) &= \frac{n(n-1)\cdots(n-(r-1))(n-r)\cdots \times 2 \times 1}{(n-r)\cdots \times 2 \times 1} \\ &= \frac{n!}{(n-r)!}\ . \end{aligned}$$

This relation is not helpful for actual calculations (it just means we work out two big products which mostly cancel anyway), but it can be useful for theoretical purposes.

Exercise. Repeat the problems from the previous page if we are concerned with those standing as well as those sitting.

A **pack of cards** is very useful for practising combinatorial techniques. It consists of 52 cards, comprising thirteen *values* ace, two, three, . . . , ten, jack, queen and king in each of four *suits* spades, hearts, diamonds and clubs.

The spades and clubs are coloured black, while the hearts and diamonds are red.

A selection of cards from a pack is referred to as a *hand of cards*. We shall assume (as is true in most card games) that a hand may not contain the same card twice, and that the order of cards in a hand is not important.

	Spades	Hearts	Diamonds	Clubs
Ace	$\spadesuit$A	$\heartsuit$A	$\diamondsuit$A	$\clubsuit$A
Two	$\spadesuit$2	$\heartsuit$2	$\diamondsuit$2	$\clubsuit$2
Three	$\spadesuit$3	$\heartsuit$3	$\diamondsuit$3	$\clubsuit$3
Four	$\spadesuit$4	$\heartsuit$4	$\diamondsuit$4	$\clubsuit$4
Five	$\spadesuit$5	$\heartsuit$5	$\diamondsuit$5	$\clubsuit$5
Six	$\spadesuit$6	$\heartsuit$6	$\diamondsuit$6	$\clubsuit$6
Seven	$\spadesuit$7	$\heartsuit$7	$\diamondsuit$7	$\clubsuit$7
Eight	$\spadesuit$8	$\heartsuit$8	$\diamondsuit$8	$\clubsuit$8
Nine	$\spadesuit$9	$\heartsuit$9	$\diamondsuit$9	$\clubsuit$9
Ten	$\spadesuit$10	$\heartsuit$10	$\diamondsuit$10	$\clubsuit$10
Jack	$\spadesuit$J	$\heartsuit$J	$\diamondsuit$J	$\clubsuit$J
Queen	$\spadesuit$Q	$\heartsuit$Q	$\diamondsuit$Q	$\clubsuit$Q
King	$\spadesuit$K	$\heartsuit$K	$\diamondsuit$K	$\clubsuit$K

Problem. How many thirteen–card hands can be chosen from a pack of cards?

Solution. We must choose thirteen cards, with repetition not allowed, from a pack of 52 cards. First let's suppose that we choose the cards in a particular order: then the number of selections is

$$P(52, 13)\ .$$

Now imagine that we have a hand of 13 cards. The number of ways that we can select all 13 of them, *in order*, is

$$P(13, 13)\ .$$

Since we are not actually interested in the order of cards within a hand, this means that we initially counted every hand many times, in fact, $P(13, 13)$ times. So the correct number of thirteen–card hands is

$$\frac{P(52, 13)}{P(13, 13)} = \frac{P(52, 13)}{13!}\ .$$

Unordered selections without repetition. Suppose we have n different items and we wish to select r of these, where

- no item may be selected more than once (for example, $\diamondsuit$ K $\spadesuit$ 7 $\spadesuit$ A $\diamondsuit$ K is not allowed);
- the order of the items is not significant (for example, $\diamondsuit$ K $\spadesuit$ 7 $\spadesuit$ A and $\spadesuit$ 7 $\diamondsuit$ K $\spadesuit$ A are regarded as the same and therefore are not both counted).

Such a selection is called a **combination** of r objects chosen from n (different) objects, and the number of possible choices is

$$C(n,r) = \binom{n}{r} = \frac{P(n,r)}{r!} \, .$$

Example. Of the hands in the problem on the previous page, how many consist of seven cards in one suit and six in another?

Solution. As usual we set out the working carefully.

1. Pick a suit . . . . 4 ways
2. Choose seven cards from this suit,
   with repetition forbidden and order not important . . . . $C(13,7)$ ways
3. Pick a different suit . . . . 3 ways
4. Choose six cards from this suit . . . . $C(13,6)$ ways

The number of hands is $4 \times C(13,7) \times 3 \times C(13,6)$.

We could rephrase the solution of the previous problem.

1. Choose two suits from four . . . . $P(4, 2)$ ways
2. Pick seven cards from the first suit . . . . $C(13, 7)$ ways
3. Pick six cards from the second suit . . . . $C(13, 6)$ ways

So, the number of hands is $P(4, 2)C(13, 7)C(13, 6)$.

By a similar argument we might say that the number of twelve–card hands containing four cards in each of three suits is

$$P(4, 3)C(13, 4)C(13, 4)C(13, 4) = P(4, 3)C(13, 4)^3 .$$

Wrong! Why, and what is the right answer?

Exercise. How many thirteen–card hands contain four cards in each of two suits and five in a third?

Two questions. You will notice from the problems we have solved so far that it is frequently important, when faced with a counting problem, to ask yourself two questions concerning the selection of objects.

- Is repetition allowed?
- Is order important?

We have seen how to handle the cases where repetition is not allowed and order is or is not important; and also the case where repetition is allowed and order is important. The final possibility (repetition allowed and order not important) has a simple answer, but the reasoning behind it is a bit more subtle, so we shall study it later.

More on combinations. The numbers $C(n, r)$ satisfy a huge number of identities, many of which have combinatorial interpretations. First an easy one. If $0 \leq r \leq n$ we can write combinations in terms of factorials:

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!\,(n-r)!}$$

or as the following quotient:

$$\begin{aligned} C(n, r) &= \frac{n(n-1)(n-2)\cdots(n-(r-1))}{r!} \\ &= \frac{n}{1}\,\frac{n-1}{2}\,\frac{n-2}{3}\,\cdots\,\frac{n-(r-1)}{r}\ . \end{aligned}$$

The last expression is handy for calculation: it can be proved that if you do the arithmetic in the order indicated (start with n, divide by 1, multiply by $n-1$, divide by 2 and so on) then the intermediate results will always be integers and you will never need to work with fractions.

Lemma. *Two properties of* $C(n,r)$. Let n and r be integers with $0 \le r \le n$. We have

- $C(n,r) = C(n,n-r)$;
- $C(n+1,r) = C(n,r) + C(n,r-1)$, provided $r \neq 0$.

Proof. Consider the problem of choosing r different objects from n objects, where order is not significant. One way of doing this is simply to select the r objects; this can be done in $C(n,r)$ ways. Alternatively, we could select $n-r$ objects, throw them away, and keep the rest; this can be done in $C(n,n-r)$ ways. Since we have just solved the same problem by two methods, our two answers must be the same: that is,

$$C(n,r) = C(n,n-r) \ .$$

For the second identity, consider choosing r objects (repetitions not allowed, order not important) from $n+1$. Doing this in the obvious way, there are $C(n+1,r)$ possibilities. Alternatively, label one of the items X, and make choices in the following way: either

1. Reject X, and choose r objects from the remaining $n \ldots .$ $C(n,r)$ ways

or

1. Choose X. . . . . 1 way
2. Choose $r-1$ objects from the remaining n . . . . $C(n,r-1)$ ways

Using this procedure, the number of possible choices is $C(n,r)+C(n,r-1)$, which is therefore equal to $C(n+1,r)$.

Comments.

- This type of argument, where we prove that two expressions are equal by showing that they are both the answer to the same counting problem, is known as a *combinatorial proof*.
- Alternatively, both identities can be proved by writing $C(n,r)$ in terms of factorials as on page 23. **Exercise**. Do so!
- Where have you seen the second identity from the previous lemma before?

Lemma. Let $0 \leq r \leq n$. If $(x+y)^n$ is expanded and like terms collected, the coefficient of $x^r y^{n-r}$ is $C(n,r)$.

Proof. Imagine $(x+y)^n$ written out as a product of n factors,

$$(x+y)^n = (x+y)(x+y)(x+y)\cdots(x+y)\ .$$

We obtain a term $x^r y^{n-r}$ by choosing r of these factors to provide an x. To obtain a single term, clearly

- we cannot choose the same factor twice;
- the order of factors is not important;
- once we have chosen certain factors to contribute an x, all the others must contribute a y.

Therefore the number of terms $x^r y^{n-r}$ is $C(n,r)$, and the proof is complete.

Corollary. *The Binomial Theorem.* If n is a non–negative integer then

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

So, the numbers $C(n,r)$ are just those that appear in the Binomial Theorem, or in Pascal's Triangle. In fact, the second identity on page 24 is just the rule which defines Pascal's Triangle. Because of this connection, the numbers $C(n,r)$ are often known as the **binomial coefficients**.

We consider some counting problems involving sets and functions.

Problem.

(a) If S is a set with n elements and $0 \le r \le n$, how many r–element subsets has S?

(b) How many subsets has an n–element set altogether?

(c) How many subsets has an n–element set altogether?

(d) What do the answers to (b) and (c) tell you about binomial coefficients?

(e) How is this related to the Binomial Theorem?

Solution.

(a) In a set, repetition of elements is not allowed and order is unimportant. So the number of ways of choosing an r–element subset from an n–element set is $C(n, r)$.

Problem. If S has n elements and T has m elements, how many functions $f : S \to T$ are there? How many of these are one–to–one? How many are onto? How many are bijections?

Solution. Write $S = \{ s_1, s_2, \dots, s_n \}$. First we count all possible functions.

1. Choose $f(s_1)$ . . . . m ways
2. Choose $f(s_2)$ . . . . m ways

$\vdots$

n. Choose $f(s_n)$ . . . . m ways

We have made n choices from m objects, where the same object may be chosen repeatedly and the order of the objects is important. So the total number of functions is m^n.

Counting one–to–one functions uses the same ideas, except that now repetition is forbidden. (If an object is chosen twice then we have, for example, $f(s_1) = f(s_2)$, and so f is not one–to–one.) Therefore the number of one–to–one functions from S to T is $P(m, n)$.

Exercise. What is $P(m, n)$ if $m < n$? Does the previous problem make sense in this case?

Counting onto functions is harder and we'll leave it till later – see if you can work it out yourself before we get there!

We can only get a bijection between sets S and T if they have the same number of elements, that is, $m = n$. If this is the case then a one–to–one function is automatically onto as well; therefore the number of bijections from S to T is $P(n, n) = n!$ if $m = n$, and zero if $m \neq n$.

Recall that on page 12 we made a mistake.

Problem. How many ten–letter words can be made from the English alphabet which contain at least one X, at least one Y and at least one Z?

Answer. $10 \times 9 \times 8 \times 26^7$.

Let's now solve this problem correctly. First we'll use the subtraction rule: the total number of ten–letter words (with repeated letters allowed) is 26^{10}, and from this we need to subtract the number of words with no X *or* no Y *or* no Z.

It is best to treat this as a problem in counting the elements of sets. Consider a universal set $\mathcal{U} = \{\text{ ten–letter words }\}$, and three subsets of $\mathcal{U}$:

$$\begin{aligned} A &= \{\text{ ten–letter words with no X }\} \\ B &= \{\text{ ten–letter words with no Y }\} \\ C &= \{\text{ ten–letter words with no Z }\}\ . \end{aligned}$$

The words we don't want to count are those which are elements of $A \cup B \cup C$. We shall solve the problem by using the **inclusion/exclusion formula**

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

. . . continued

Now to choose a word in A we must pick ten letters from 25 possibilities (X is not used), with repetitions permitted and order important. The number of such words is 25^{10}, and the numbers of elements in B and in C are the same.

Similarly, words in $A \cap B$, in $A \cap C$ and in $B \cap C$ are to be chosen from 24 letters, so each of these sets contains 24^{10} words. Likewise, $A \cap B \cap C$ has 23^{10} elements.

Putting all this information back together, the number of words is

$$\begin{aligned} \left|A^c \cap B^c \cap C^c\right| &= \left|(A \cup B \cup C)^c\right| \\ &= |\mathcal{U}| - |A \cup B \cup C| \\ &= |\mathcal{U}| - |A| - |B| - |C| \\ &\qquad + |A \cap B| + |A \cap C| + |B \cap C| \\ &\qquad - |A \cap B \cap C| \\ &= 26^{10} - 3 \times 25^{10} + 3 \times 24^{10} - 23^{10}\ . \end{aligned}$$

Comments.

- The principle of inclusion/exclusion is a formula for counting the elements in a union of sets. So, you should always begin an inclusion/exclusion problem by defining a universal set and appropriate subsets.
- There is an inclusion/exclusion formula for any (finite) number of sets, and the coefficients are always ± 1. To count the number of elements in a union of sets, add the number in each set, then subtract the number in every possible intersection of two sets, then add the number in every possible intersection of three sets, and so on. For example,

$$\begin{aligned} |A_1 \cup A_2 \cup A_3 \cup A_4| = {} & |A_1| + |A_2| + |A_3| + |A_4| \\ & - |A_1 \cap A_2| - \cdots \\ & + |A_1 \cap A_2 \cap A_3| + \cdots \\ & - |A_1 \cap A_2 \cap A_3 \cap A_4| \ . \end{aligned}$$

- Inclusion/exclusion is commonly useful where we have to count items which have at least one of various properties, and where an item may have more than one of the properties. This is because to say that an item has one property *and* another gives more definite information about the item than to say that it has one property *or* another, and therefore makes it easier to count the possibilities.

Problem. How many thirteen–card hands chosen from a standard pack contain

- exactly five cards in some suit?
- two cards of the same value and colour (for example, $\diamondsuit\, 10$ and $\heartsuit\, 10$)?

Solution. Write $\mathcal{U} = \{\,\text{thirteen–card hands}\,\}$. For the first problem, let

$$S = \{\,\text{thirteen–card hands with exactly five spades}\,\}\ ,$$

and define H, D and C similarly. To select a hand in S,

1. Choose 5 spades from 13 .... $C(13,5)$ ways
2. Choose 8 other cards from 39 .... $C(39,8)$ ways

The sets H, D and C have the same number of elements as S. For sets like $S \cap H$ we have

1. Choose 5 spades from 13 .... $C(13,5)$ ways
2. Choose 5 hearts from 13 .... $C(13,5)$ ways
3. Choose 3 other cards from 26 .... $C(26,3)$ ways

Hands in a triple intersection such as $S \cap H \cap D$ must have at least fifteen cards, which is impossible. So $|S \cap H \cap D| = 0$ and $|S \cap H \cap D \cap C| = 0$, and the total number of hands is

$$4\,C(13,5)C(39,8) - 6\,C(13,5)^2C(26,3)\ .$$

For the second problem we must consider 26 sets such as

$$S_1 = \{\,\text{thirteen–card hands containing } \heartsuit\,\text{A and } \diamondsuit\,\text{A}\,\}\ ;$$

any intersection of seven or more of these will be empty. The number of hands is

$$\begin{aligned} 26\,C(50,11) &- C(26,2)C(48,9) \\ &+ C(26,3)C(46,7) - C(26,4)C(44,5) \\ &+ C(26,5)C(42,3) - C(26,6)C(40,1)\ . \end{aligned}$$

Explanation. A hand in the intersection of three of the sets S_k must contain six specified cards (for example, $\spadesuit$ 2, $\clubsuit$ 2 and $\heartsuit$ J, $\diamondsuit$ J and $\heartsuit$ Q, $\diamondsuit$ Q); to complete the hand we must choose 7 of the remaining 46 cards. This can be done in $C(46,7)$ ways; and the number of such three–way intersections is $C(26,3)$. This is why we have a term $C(26,3)C(46,7)$, and the other terms are justified similarly.

Exercise. Carefully write out an explanation for another of the terms in the above result.

Example. How many thirteen–letter words having no repeated letters contain at least one of the subwords CAT and DOG and FOX?

Solution. As a "suitable" word could contain more than one of the subwords, we use inclusion/exclusion. Let $\mathcal{U} = \{$ thirteen–letter words without repeated letters $\}$, and let C, D, F be the sets of words from $\mathcal{U}$ which contain CAT, DOG and FOX respectively.

Now imagine that we have available the 24 "elements" $\boxed{\text{CAT}}$, B, D, E, F, . . . , Z. To form a word in C, we choose 11 of these, including $\boxed{\text{CAT}}$, repetitions banned and order important.

1. Choose a place for $\boxed{\text{CAT}}$ . . . . 11 ways
2. Choose 10 letters from the remaining 23 . . . . $P(23, 10)$ ways

Therefore $|C| = 11P(23, 10) = |D| = |F|$.

To form words in $C \cap D$, consider the 22 "elements" $\boxed{\text{CAT}}$, $\boxed{\text{DOG}}$, B, E, F, H, . . . , Z, and choose 9 of them.

1. Choose places for $\boxed{\text{CAT}}$, $\boxed{\text{DOG}}$, order important, no repetition . . . . . $P(9, 2)$ ways
2. Choose 7 other letters from 20 . . . . $P(20, 7)$ ways

Hence $|C \cap D| = P(9, 2)P(20, 7) = |C \cap F|$.

Since DOG and FOX cannot appear in the same word, $|D \cap F| = |C \cap D \cap F| = 0$. So the number of words is

$$3 \times 11P(23, 10) - 2P(9, 2)P(20, 7) \ .$$

We return to a problem which was too hard for us earlier!

Problem. If S is a set with n elements and T is a set with m elements, how many functions are there from S **onto** T?

Solution. Write $S = \{ s_1, s_2, \ldots, s_n \}$, $T = \{ t_1, t_2, \ldots, t_m \}$, and let $\mathcal{U}$ be the set of **all** functions from S to T. The functions which are not onto, and which we therefore do not want to count, are those for which t_1 is missing from the range, or t_2 is missing from the range, and so on: that is, those in the union of the sets

$$F_k = \{ \text{functions } f : S \to T \text{ with } t_k \notin \text{range}(f) \}$$

for $k = 1, 2, \ldots, m$, and we can count these functions by inclusion/exclusion.

As on page 28, we choose a function in F_1 by selecting the n values $f(s_1), f(s_2), \ldots, f(s_n)$. In this case, however, t_1 cannot be chosen, so there are $m - 1$ options for each choice, and $(m-1)^n$ altogether. Similarly, the number of functions in $F_1 \cap F_2$ is $(m-2)^n$, and so on. The number of functions from S onto T is

$$\begin{aligned} &|\mathcal{U}| - |F_1| - \cdots + |F_1 \cap F_2| + \cdots - |F_1 \cap F_2 \cap F_3| - \cdots \\ &\qquad = m^n - m(m-1)^n + C(m,2)(m-2)^n - C(m,3)(m-3)^n + \cdots \\ &\qquad = \sum_{k=0}^{m} (-1)^k C(m,k)(m-k)^n \ . \end{aligned}$$

If, say, $n = 5$ and $m = 4$, the number of functions is $4^5 - 4 \times 3^5 + 6 \times 2^5 - 4 \times 1^5 = 240$.

Problem. How many words can be made by rearranging all the letters of the word WOOLLOOMOOLOO?

Solution. We wish to select every letter of the word once each and arrange them in order. However, the problem is complicated by the fact that not all of the letters of WOOLLOOMOOLOO are different. (RANDWICK would be much easier!) To begin, however, think of the letters as being all different: we could write the word as

$$\mathrm{W_1O_1O_2L_1L_2O_3O_4M_1O_5O_6L_3O_7O_8}\ .$$

Then the number of arrangements is just 13!. But this includes examples such as

$$\mathrm{W_1O_1O_2L_3L_2O_3O_4M_1O_5O_6L_1O_7O_8}$$

and

$$\mathrm{W_1O_1O_2L_2L_3O_3O_4M_1O_5O_6L_1O_7O_8}\ ,$$

which are actually all the same. There will be 3! words like this, obtained by rearranging the three Ls; therefore we must divide our previous answer of 13! by 3!. Similarly we must divide by 8! to take account of the repeated Os. As the available letters are one W, eight Os, three Ls and one M, the number of arrangements is

$$\frac{13!}{1!\,8!\,3!\,1!}\ .$$

The factors of 1! in the denominator can be omitted as they do not alter the number, but it's better to include them as they help to clarify the pattern.

What we have just seen is known as a problem of **permutations with repetitions**. Suppose that we have n_1 identical objects, n_2 other identical objects not identical to the first lot, and so on, finishing with n_m objects of an mth type. Then the total number of objects is $n = n_1 + n_2 + \cdots + n_m$, and the number of ways of arranging these n objects in a row is

$$\frac{n!}{n_1!\, n_2! \, \cdots \, n_m!} \ .$$

Exercise. How many ways are there to arrange the letters of WOOLLOOMOOLOO in a circle on the floor, if we can look at the circle from any direction we please?

Placing objects in sets. Consider the following problem. We have five different objects, say A, B, C, D and E, and we want to place them into three sets S, T and U. Two elements are to go into S, two into T and one into U. How many arrangements are possible?

Solution. Remember that in a set, order is not important.

1. Choose two elements for S: repetition is not allowed, and order is unimportant . . . . $C(5,2)$ ways
2. Choose two elements for T . . . . $C(3,2)$ ways
3. Choose one element for U . . . . 1 way

So the total number of arrangements is $C(5,2)C(3,2)$.

Now let's write this in terms of factorials:

$$C(5,2)C(3,2) = \frac{5!}{2!\,3!}\,\frac{3!}{2!\,1!} = \frac{5!}{2!\,2!\,1!}\ .$$

Do you notice anything? Can you explain it?

Another problem. This time we have five identical objects, say •, •, •, • and •, and we wish to place them in three sets S, T and U, with no restrictions on the number of items in each set (we can even have one or more sets empty if we wish). There is a clever argument for counting the number of distributions in this and similar problems.

S	T	U
• • • • •		
	• • • • •	
		• • • • •
• • • •	•	
• • • •		•
•		• • • •
• • •	•	•

Solution. Some of the possible arrangements of the five objects are shown in the accompanying table. We can view each of these distributions as an arrangement of five dots and two lines in seven places. To count the total number of such arrangements we simply choose two of the seven places for the lines. We cannot choose the same place twice, and the order in which we choose them is irrelevant, so the number of possibilities is $C(7, 2)$.

Alternatively, treat it as a "permutations with repetitions" problem. As on page 36, the number of words we can make by arranging the symbols $\bullet\bullet\bullet\bullet\bullet\,||$ is

$$\frac{7!}{5!\,2!} = C(7, 2)\ .$$

This argument enables us to fill in a gap in our technique.

Unordered selections with repetition. Suppose we have n different items and we wish to make r selections from these items, where

- the same item may be selected more than once;
- the order of items is not relevant.

For example, if the items available are the letters of the alphabet, then AAABNN is an allowable choice involving six selections, and is the same as BANANA. A different possible choice is ABBBNA.

In such a selection, the only matter of importance is the number of times each item is chosen. So we can visualise a selection by drawing a dot for each time the first item is selected, then a line to separate it from the next item, then a dot for each time the second item is selected, then another line, and so on. For example, if we are allowed to use all the letters of the alphabet then BANANA (or AAABNN) would be pictured as

$$\bullet\bullet\bullet|\bullet||||||||||||\bullet\bullet||||||||||||$$

If we have n items we shall require $n-1$ "dividers"; if we are making r selections we shall have r dots; these $n+r-1$ objects must be arranged in $n+r-1$ places, and all we need to do is to choose the places for the lines. We cannot choose any place twice, and the order of choices is irrelevant, so the number of possibilities is

$$C(n+r-1, n-1) \ .$$

Summary. Suppose that we have n different items and we wish to make r selections from these items. The number of ways of doing so depends on whether or not we may choose the same item more than once, and whether or not the order of selections is relevant.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Repetitions</td></tr>
<tr><td>allowed</td><td>not allowed</td></tr>
<tr><td rowspan="2">Order</td><td>relevant</td><td>n^r</td><td>$P(n,r)$</td></tr>
<tr><td>irrelevant</td><td>$C(n+r-1,n-1)$</td><td>$C(n,r)$</td></tr>
</table>

Problem. How many solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 = 26 \ ,$$

if x_1, x_2, x_3, x_4, x_5 are non–negative integers

(a) with no further restrictions?

(b) with $x_1 \geq 1$, $x_2 \geq 2$, $x_3 \geq 3$, $x_4 \geq 4$ and $x_5 \geq 5$?

(c) with every $x_k \leq 10$?

(d) with every $x_k \leq 6$?

(e) with every x_k even?

(f) with $x_k \equiv 1 \pmod 3$ for every k?

. . . *continued*

Solution. We can picture a solution as a row of 26 dots, with lines separating the dots into values for x_1, x_2, x_3, x_4, x_5. For example,

$$\bullet\bullet\bullet\bullet\bullet\bullet\,|\,|\,\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\,|\,\bullet\bullet\bullet\bullet\bullet\bullet\bullet\,|$$

would represent the solution

$$x_1 = 6\,,\ x_2 = 0\,,\ x_3 = 13\,,\ x_4 = 7\,,\ x_5 = 0\,.$$

Therefore, this can be reinterpreted as the type of problem we have just introduced.

(a) There are no restrictions on placing the dots, so the number of solutions is $C(30, 4)$.

(b) We'll need at least one dot in the first sector, two in the second, three in the third, four in the fourth and five in the fifth. So we set aside 15 dots and consider what happens to the remaining 11. We arrange these in a row together with 4 dividers separating them into five categories; this can be done in $C(15, 4)$ ways. We then put one of the "reserved" dots into the first category, two into the second and so on: this guarantees that the required inequalities will hold. Hence, the number of possible solutions is $C(15, 4)$.

(c) These arrangements can be counted by the method of inclusion/exclusion. Write $\mathcal{U}$ for the set of *all* solutions to the given equation, and for $k = 1, 2, 3, 4, 5$ let

$$S_k = \{\,\text{solutions in which } x_k > 10\,\}\ .$$

To find a solution in S_1 we reserve 11 dots for the first sector, then distribute the other 15 arbitrarily into five sectors: there are $C(19, 4)$ ways of doing so. Similarly, the number of solutions in $S_1 \cap S_2$ is $C(8, 4)$. Any intersection of three or more sets, for example, $S_1 \cap S_2 \cap S_3$, is empty as it contains only those solutions in which $x_1 + x_2 + x_3$ is already at least 33, which is impossible. By inclusion/exclusion, the number of solutions with every $x_k \leq 10$ is

$$\begin{aligned}\left|S_1^c \cap S_2^c \cap S_3^c \cap S_4^c \cap S_5^c\right| &= |(S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5)^c| \\ &= |\mathcal{U}| - |S_1| - \cdots + |S_1 \cap S_2| + \cdots \\ &= C(30, 4) - C(5, 1)C(19, 4) + C(5, 2)C(8, 4)\ .\end{aligned}$$

(d) Using inclusion/exclusion again, we find that there are

$$C(30, 4) - C(5, 1)C(23, 4) + C(5, 2)C(16, 4) - C(5, 3)C(9, 4)$$

solutions in this case. **Exercise**. Check this, giving your working in full detail.

. . . continued

However, problem (d) has a simple and interesting *alternative solution*. Arrange five groups of six dots with lines between them.

$$\bullet\bullet\bullet\bullet\bullet\bullet \mid \bullet\bullet\bullet\bullet\bullet\bullet \mid \bullet\bullet\bullet\bullet\bullet\bullet \mid \bullet\bullet\bullet\bullet\bullet\bullet \mid \bullet\bullet\bullet\bullet\bullet\bullet$$

This gives 30 dots altogether; to adjust this to 26, we add four "antidots". For example, the arrangement

$$\mid \; \mid \circ \circ \mid \circ \mid \circ$$

would yield the solution

$$x_1 = 6\,,\ x_2 = 6\,,\ x_3 = 4\,,\ x_4 = 5\,,\ x_5 = 5\,.$$

The number of ways of placing four antidots into five locations is clearly the same as the number of ways of placing four dots into five locations, that is, $C(8,4)$.

Note that in adding at most four antidots to a group of six dots we shall still have a non–negative number of dots in each sector, and so we shall obtain a valid solution of the equation. This method would *not* work for question (c).

Exercise. Check that our two answers for this problem are the same.

(e) To ensure that every x_k is even we shall place thirteen "objects" and four "dividers" in a row, where each "object" is a *pair* of dots. The number of possible solutions is $C(17,4)$.

(f) In this case we shall reserve one dot for each section and then allocate dots in groups of 3, so that each x_k is always 1 modulo 3. The number of solutions is $C(11,4)$.

A **multinomial** is an expression like

$$(x+y+z)^{100} \quad \text{or} \quad (2a+3b+5c+7d+11e)^{100}$$

– in other words, pretty much like a binomial (page 26) except that there can be any number of terms inside the brackets.

Problem. If the multinomial

$$(m+a+t+h)^{1081}$$

is expanded and terms collected,

- how many different terms will there be?
- what will be the coefficient of $m^{268}a^{270}t^{271}h^{272}$?

Solution. The powers of m, a, t, h will be non–negative integers whose sum is 1081; in other words, this is a "dots and lines" problem like that on page 42. The number of terms is $C(1084, 3)$.

The coefficient of $m^{268}a^{270}t^{271}h^{272}$ will be the number of different ways in which such a term can arise; that is, the number of different arrangements of 268 ms, 270 as, 271 ts and 272 hs. So this is a "WOOLLOOMOOLOO" problem, and the required coefficient is

$$\frac{1081!}{268!\,270!\,271!\,272!}\ .$$

Exercise. If $(2a + 3b + 5c + 7d + 11e)^{100}$ is expanded,

- how many different terms will there be?
- what will be the coefficient of $a^{18}b^{19}c^{20}d^{21}e^{22}$?

Problem. Estimate the number of primes up to 1000.

Solution. One approach to this question is as follows. We'll employ inclusion/exclusion to count all the numbers up to 1000 which are not multiples of 2 or 3 or 5 or 7. This will include all the primes(?), though it will also include composites such as 11×13. Let $\mathcal{U} = \{\,1, 2, \ldots, 1000\,\}$ and write

$$S_m = \{\,\text{multiples of } m \text{ up to } 1000\,\}\ , \quad \text{so} \quad |S_m| = \left\lfloor \frac{1000}{m} \right\rfloor\ ;$$

then the number of integers up to 1000 which are not multiples of $2, 3, 5$ or 7 is

$$\begin{aligned}
|\mathcal{U}| - |S_2 \cup S_3 \cup S_5 \cup S_7| &= |\mathcal{U}| - |S_2| - |S_3| - |S_5| - |S_7| \\
&\qquad + |S_2 \cap S_3| + \cdots - |S_2 \cap S_3 \cap S_5| - \cdots + |S_2 \cap S_3 \cap S_5 \cap S_7| \\
&= |\mathcal{U}| - |S_2| - |S_3| - |S_5| - |S_7| \\
&\qquad + |S_6| + \cdots - |S_{30}| - \cdots + |S_{210}| \\
&= 1000 - 500 - 333 - 200 - 142 \\
&\qquad + 166 + 100 + 71 + 66 + 47 + 28 - 33 - 23 - 14 - 9 + 4 \\
&= 228\ .
\end{aligned}$$

In this total we have not counted $2, 3, 5$ and 7, which are prime, so we must add 4; and we have counted 1, which is not, so we may subtract 1. Thus there are no more than 231 primes up to 1000.

Comment. In fact, there are 168 primes up to 1000. We have counted 63 extra numbers, ranging from $121 = 11 \times 11$ to $989 = 23 \times 43$. Let $\pi(x)$ denote the number of primes up to x. There is no known useful exact formula for $\pi(x)$, but there are many inequalities and approximate formulae, of which the following is a small selection. First we have

$$\pi(x) \simeq \frac{x}{\log x} \ ,$$

where we have written log for the natural (base e) logarithm, denoted ln in first–year calculus. This approximation gets closer and closer as x increases, because

$$\frac{\pi(x)}{x/\log x} \to 1 \quad \text{as } x \to \infty \ .$$

If you want to know just how close to 1 this ratio is, we have the inequalities

$$\frac{1}{3} < \frac{\pi(x)}{x/\log x} < \frac{4}{3} \ ,$$

true for all $x \geq 2$. A more complicated but very accurate approximation is

$$\pi(x) \simeq \int_2^x \frac{dt}{\log t} \ .$$

For some details about this see MATH1131/1141 Calculus, chapter 8, problem 28.

Exercise. By substituting $x = 1000$ into these formulae, see how they compare for accuracy with our estimate $\pi(1000) \leq 231$.

The **Pigeonhole Principle** is not a counting technique, but rather a proof technique which makes use of counting methods.

Theorem. *The Pigeonhole Principle*. Let n be a positive integer. If n boxes contain (altogether) more than n objects, then there must be a box containing more than one object.

This result is traditionally stated as "if there are n pigeonholes with more than n pigeons living in them, then there must be a pigeonhole with more than one pigeon living in it" – hence the name. I do not know how pigeons originally got into the act.

Examples.

- Among any 27 people, there must be two or more whose surnames have the same initial. **Proof**. We have 26 "boxes" (letters of the alphabet), and 27 "objects" (people). Put each person in the box corresponding to their initial. Since there are more objects than boxes, there must be a box with more than one object in it. That is, there are two or more people with the same initial.

- In the game of bridge a hand consists of 13 cards. Show that if I play 1000 hands of bridge (dealt from a full pack every time) then there must be two with the same suit distribution, that is, the same number of spades, the same number of hearts, the same number of diamonds and the same number of clubs.

  Investigation. Usually, the hard part of a pigeonhole principle problem lies in deciding what we should take to be the "objects" and what the "boxes". A good way to think about this is to remember that the conclusion we obtain from the pigeonhole principle is always the same: there is a box containing more than one object; and to work out how we can make this correspond to our desired conclusion.

  In the present case, we want to show that there is a suit distribution shared by more than one hand. So, we shall take the "boxes" to be suit distributions, and the "objects" to be the hands.

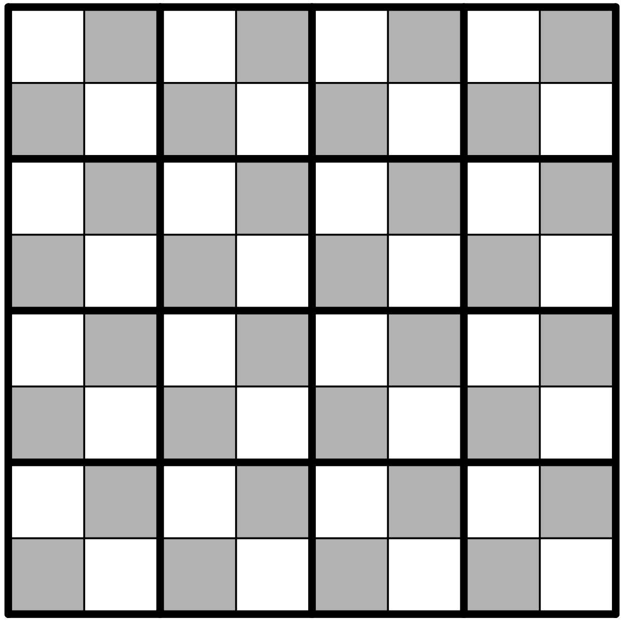
  Solution. Consider placing 1000 "objects" (hands of cards) into "boxes" representing suit distributions. The number of boxes is the same as the number of non–negative integer solutions to the equation

  $$s + h + d + c = 13 \ .$$

  Thus there are $C(16, 3) = 560$ boxes; since there are more than this number of objects, there must be a box containing more than one object, that is, a suit distribution I received in two or more hands.

- On a (normal 8×8) chessboard I choose 17 squares and colour them red. However I do this, I must have coloured two squares which are adjacent either horizontally, vertically or diagonally.

  Proof. Divide the chessboard into sixteen 2×2 sections as shown by the thick lines:

  these will be our "pigeonholes". The seventeen red squares are the "pigeons". Therefore there is a 2×2 section containing two red squares, and it is clear that these two are adjacent to each other.

Example. Another example: if a million people live in a city, there must be at least two with the same birthday, because there are only 366 possible birthdays (including 29 February).

It is clear that in this example we could have said much more: there will not only be two but many people having the same birthday!

Theorem. *The Generalised Pigeonhole Principle*. Let n and k be positive integers. If n boxes contain (altogether) more than nk objects, then there must be a box containing more than k objects.

Examples.

- Improving on the above example, we can calculate that 1000000 is more than 2732 times 366, and so in a city of a million people there must be 2733 or more who share the same birthday.

- 888 people take a pack of cards each and deal themselves thirteen cards. Show that there must be eight people who have the same number of jacks as each other, and the same number of queens, and the same number of kings. Can we say for certain that there will be nine such people? **Solution**. There are five possibilities for the number of jacks in a hand (anywhere from 0 to 4), and the same for queens and kings. These three numbers do not have to be all different, and their order is important, so the total number of possibilities is $5^3 = 125$. Imagine 125 boxes labelled, for example, "no jacks, two queens, two kings" and put each of the 888 people into the box which describes their hand. As there are more than 7 times 125 people, there must be a box containing more than 7 people. So there must be eight people holding the same type of hand. However there need not be nine such people since, for example, the 888 people could consist of exactly eight having each of 111 different types of hand (and none with the other 14 types).

Exercise. Prove that if 33 squares on a chessboard are coloured red, there must be three squares forming a red "L" (possibly upside down or back to front).

We can also consider "inverse" pigeonhole problems in which we ask how many "objects" there must be in order to guarantee a certain number in some class.

Example. Professor X is a bit eccentric and wants all the students in his MATH1801 tutorial to have the same birthday. However, for reasons of efficiency the university will not let him have fewer than 30 students in his class. How many students must be enrolled in MATH1801 for it to be certain that he can find 30 students all with the same birthday?

Solution. He will need $N = 366 \times 29 + 1$ students. If there are this many "pigeons" (students – sorry!) and 366 "pigeonholes" (birthdays), then N is more than 29 times 366, so there must be a birthday shared by more than 29 students, that is, by 30 or more. So there must be at least 10615 students enrolled in order for Professor X to be sure of getting his way.

Exercise. Professor Y is even more eccentric. He wants all the students in his class to have the same birthday as his own, not just the same as each other. How many students must be enrolled to be sure that this can be done?

The Pigeonhole Principle may sound so obvious that it is hard to imagine using it for anything of significance. However, it does have important applications in mathematics and computer science.

Efficiency of sorting algorithms. Suppose we have an algorithm which accepts as input a list of n different numbers, and returns a list containing the same numbers, sorted into increasing order. Suppose also that the algorithm uses only one type of instruction:

- compare the numbers in positions k_1 and k_2; if they are in the right order do nothing, if they are in the wrong order, swap them.

Then we must allow at least

$$n \log_2 \left(\frac{n}{e}\right)$$

steps to ensure that all lists of n numbers are sorted correctly.

Example. To illustrate such an algorithm, suppose that we have a list of just four numbers $[a, b, c, d]$. If the first instruction is to compare positions 1 and 2 we shall obtain the list

$$[a, b, c, d] \quad \text{or} \quad [b, a, c, d] \ ,$$

depending on whether $a \leq b$ or $a > b$. If the next instruction is to compare the elements in positions 2 and 3 we could end up with

$$[a, b, c, d] \quad \text{or} \quad [a, c, b, d] \quad \text{or} \quad [b, a, c, d] \quad \text{or} \quad [b, c, a, d] \ ,$$

and so on.

Proof. Suppose that the input list contains n elements, and that the algorithm takes m steps. What the algorithm does to any particular list can be described by a sequence of m actions chosen from two options, such as

$$\text{swap, no swap, no swap}, \ldots, \text{no swap, swap.}$$

There are 2^m such sequences. Given a set of n (different) numbers, there are $n!$ orders in which these elements may be listed.

Suppose that $n! > 2^m$, and apply the Pigeonhole Principle: take the pigeons to be the possible input lists, and the pigeonholes to be the "swap/no swap" sequences. Since there are more lists than sequences, there must be two different input lists which are treated in the same way. If both were sorted correctly this is impossible, because by starting with the final sorted list and undoing all the swaps we see that the two input lists must have been the same after all. So to ensure that all lists are sorted correctly we must have $2^m \geq n!$, that is,

$$m \geq \log_2(n!) \ .$$

We can use calculus methods to prove that

$$\log_2(n!) > n \log_2\left(\frac{n}{e}\right)$$

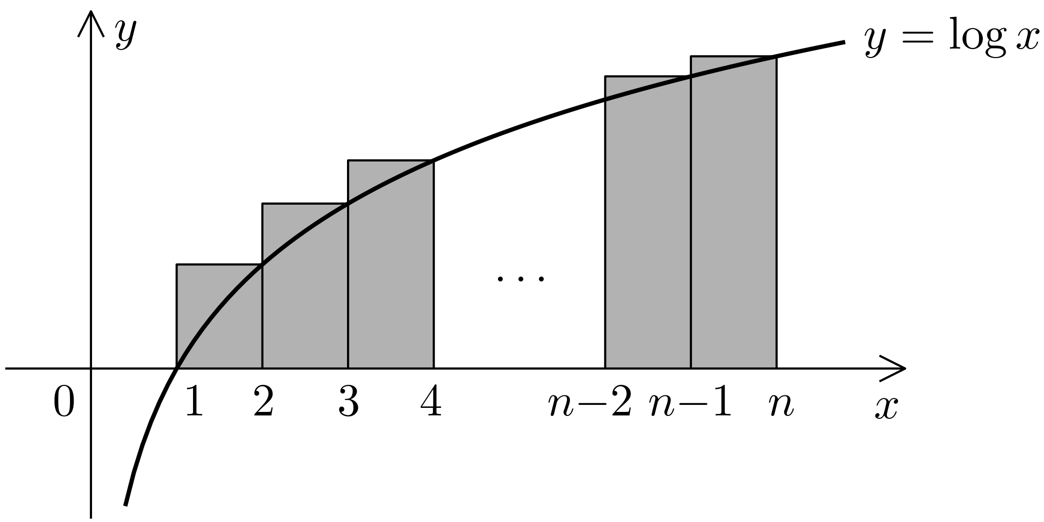
for any positive integer n (**exercise** – see the next page), and this completes the proof.

Comment. There are sorting procedures which *do not* rely solely on comparisons of individual elements; the foregoing result does not apply to such methods.

Example. If a list of $1\,000\,000$ tax records is sorted by any comparison–based method, the number of comparisons required is at least

$$1\,000\,000 \log_2\left(\frac{1\,000\,000}{e}\right) = 18\,400\,000 \ .$$

Exercise. The diagram shows the graph $y = \log x$, where log denotes the natural (base e) logarithm.



- Explain why the area of the shaded region is $\log(n!)$.
- Calculate the area under the graph from $x = 1$ to $x = n$.
- Deduce that

$$\log(n!) > n \log\left(\frac{n}{e}\right) \ .$$

- Deduce that

$$\log_2(n!) > n \log_2\left(\frac{n}{e}\right) \ .$$

Error–correcting codes. A message consisting of 0s and 1s is to be sent broken up into eight–bit "words". There are 256 possible words, but we do not wish to use all of them, as then a single error would result in an incorrect or indecipherable message being received. Instead, we wish to be sure that if any word contains up to two errors we will still know what the original word was. How many eight–bit words can we actually use?

Solution. Each correct word can be transformed into 8 other words by altering one letter, and into $C(8,2) = 28$ further words by altering two letters. We can visualise a correct word as being the centre of a "cluster" of 37 words from which the original word can be reconstructed. However, the "clusters" must not overlap as then we would not know which correct word was originally transmitted. For example, it would not be satisfactory if 11111111 and 11111000 were both correct, for then a received word 11111110 could be the former with an error in the last place, or the latter with errors in the sixth and seventh places. So, if there are n correct words then we must have n non–overlapping clusters of 37 words for a total of $37n$ different words: as there are only 256 eight–bit words altogether we must have $37n \leq 256$. Therefore $n \leq 6$, and we cannot use more than six words.

Comments.

- A system in which we only have 6 allowable words is obviously pretty useless! If we want to use this kind of procedure in practice, we have to modify our requirements. Two possible ways to do this are to use longer words (for example, 16 bits instead of 8 bits); or to demand less error–correcting capability (for example, only insist that words with one error can be corrected).

- The above reasoning does not quite use the pigeonhole principle. For a more formal argument, let the pigeons be pairs (c, w) such that c is a correct word and w is a word which differs from c in at most two places; let the pigeonholes be all words w, whether correct or not; and put the pigeon (c, w) into the pigeonhole w.

  Exercise. Suppose that we have n correct words. Count the pigeons, count the pigeonholes and finish the argument.

- The above argument does not actually guarantee that we can even find six suitable words! (How many 2×2 squares will fit into a 5×5 square without overlapping?)
  Exercise. Find six words, or prove that they cannot be found.

RECURRENCE RELATIONS

Suppose that we wish to solve a counting problem involving an unknown parameter n. For example,

- how many n–letter words contain the subword XY?
- how many saturated hydrocarbon molecules are there with n carbon atoms?

A powerful method of solving some problems of this type is to set up and solve a **recurrence relation**.

That is, let a_n be the answer to the problem; find an equation relating a_n to the previous values a_{n-1}, a_{n-2} and so on; and use this equation to derive a formula for a_n directly in terms of n.

Example. A vending machine accepts \$1 and \$2 coins. In how many ways can one insert \$$n$ into the machine? For example, if $n = 5$ there are eight ways:

$$2+2+1\,,\quad 2+1+2\,,\quad 1+2+2\,,$$
$$2+1+1+1\,,\quad 1+2+1+1\,,\quad 1+1+2+1\,,\quad 1+1+1+2\,,$$
$$1+1+1+1+1\,.$$

Solution. Let a_n be the number of ways of feeding the machine \$$n$. To do this, if $n \geq 2$, we can

- insert a \$1 coin, then another \$$(n-1)$: there are a_{n-1} ways of doing this; or
- insert a \$2 coin, then another \$$(n-2)$: there are a_{n-2} ways of doing this.

Therefore

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

This equation will give us the value of a_n in terms of previous values, provided that $n \geq 2$. If $n = 0$ or $n = 1$ we have to find the values of a_n directly from the information in the problem. In fact, it is not hard to see that we have the **initial conditions**

$$a_0 = 1 \quad \text{and} \quad a_1 = 1\ .$$

We shall see later how to find an explicit formula for a_n.

Comment. If the initial condition $a_0 = 1$ worries you, you could leave a_0 undefined and calculate a_1 and a_2 as the initial conditions.

Exercise. The vending machine malfunctions and will not accept two consecutive \$2 coins. Now how many ways are there to insert $\$n$ into the machine?

Specifically, how many ways are there to pay the machine \$10?

Solution.

A very famous sequence of numbers is the **Fibonacci sequence**: start with $0, 1$ and then obtain each element of the sequence by adding the previous two. That is,

$$F_0 = 0\,,\ F_1 = 1 \quad \text{and} \quad F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2.$$

Notice that the recurrence relation is the same as that on page 62, though the initial conditions are different. The solution to the first vending machine problem can be written as

$$a_n = F_{n+1}\ .$$

The first few terms of the Fibonacci sequence are

$$0\,,\ 1\,,\ 1\,,\ 2\,,\ 3\,,\ 5\,,\ 8\,,\ 13\,,\ 21\,,\ 34\,,\ 55\,,\ 89\,,\ 144\,,\ \ldots\ .$$

The Fibonacci sequence has many amazing properties. Here are a few:

- F_n is the nearest integer to $\dfrac{1}{\sqrt{5}}\left(\dfrac{1+\sqrt{5}}{2}\right)^n$.
- $F_0^2 + F_1^2 + F_2^2 + \cdots + F_n^2 = F_n F_{n+1}$.
- If F_n is prime then n is prime, with one exception.

An important topic in mathematical computer science is the study of **formal languages**. A language, in the mathematical sense, is any set of words formed from a given alphabet. The study of "natural" languages is difficult because of their lack of precision, so we are thinking rather of things like computer programming languages. Many languages are defined *recursively* by giving rules for constructing meaningful statements from the simplest possible components.

In studying languages it is often convenient to have a word containing *no* letters! This is referred to as the *empty word* and is denoted by λ.

Example. Consider an alphabet $\{ a, b \}$, and define a language L as follows:

- λ and a are in L;
- if w is in L then bw and abw are in L.

Then

$$\begin{aligned} L &= \{ \lambda, a, b, ab, ba, aba, bb, abb, bab, abab, \dots \} \\ &= \{ \text{words which contain only } a\text{s and } b\text{s} \\ &\qquad \text{and which} \qquad\qquad\qquad\qquad\qquad\qquad \} . \end{aligned}$$

Problem. If L is the language on the previous page, find a recursion giving the number of words of length n in L.

Solution. Let a_n be the number of words of length n in L. For $n \geq 2$, such a word is not one of the basis words and therefore must have been derived by using one of the two recursive rules. That is, such a word is

- bw, where w is a word in L of length $n-1$; or
- abw, where w is a word in L of length $n-2$.

So, the number of such words is $a_{n-1} + a_{n-2}$, and the sequence $\{ a_n \}$ satisfies the recurrence relation

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

This argument, of course, cannot tell us the number of words of length 0 or 1, and these have to be calculated separately to give the initial conditions

$$a_0 = 1 \ , \quad a_1 = 2 \ .$$

From earlier work, we see that

$$a_n =$$

We have seen examples where we can describe the solution to some combinatorial problem recursively. Usually, however, we prefer to obtain the solution explicitly – that is, find a formula for a_n directly in terms of n, and not in terms of a_{n-1}, a_{n-2} and so on. Doing this is referred to as *solving* the recurrence.

Terminology. An equation which gives a_n in terms of $a_{n-1}, a_{n-2}, \ldots, a_{n-k}$ is called a **recurrence relation of order** k. If the recurrence relation has the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} \ ,$$

or, equivalently,

$$a_n - c_1 a_{n-1} - c_2 a_{n-2} - \cdots - c_k a_{n-k} = 0 \ ,$$

it is said to be **linear**. The coefficients $c_1, c_2, \ldots, c_k$ may in principle depend on the variable n, but we shall concentrate on the case of **constant coefficients**. The above recurrence, with zero on the right hand side, is called **homogeneous**; we shall also look at the **inhomogeneous** case

$$a_n - c_1 a_{n-1} - c_2 a_{n-2} - \cdots - c_k a_{n-k} = f(n) \ .$$

A recurrence of order k cannot, by itself, determine the values of $\{ a_n \}$ since it is necessary to specify separately the values of $a_0, a_1, \ldots, a_{k-1}$. These are known as **initial conditions**. The question of solving a recurrence subject to given initial conditions is called an **initial value problem**.

If initial conditions are not known, the best we can do is to find the **general solution** of the recurrence; this will involve k unknown constants.

The problem of solving a **first or second order** (or higher) **linear recurrence with constant coefficients** is very similar to that of solving a second order linear differential equation with constant coefficients. For a start we shall look at the first order homogeneous case.

Theorem. The initial value problem

$$a_n = ra_{n-1} \quad \text{with} \quad a_0 = A$$

has solution

$$a_n = Ar^n \ .$$

Proof. Exercise! An easy induction.

Second order recurrences. Consider the second order linear recurrence with constant coefficients

$$a_n + pa_{n-1} + qa_{n-2} = 0 \ . \qquad (*)$$

We shall solve this by intelligent guesswork! Since $a_n = r^n$ is a solution of the rather similar first order equation on page 68, we might hope that it will also be a solution of the second order equation – though we don't really have a clue what value we should take for the constant r. Let's try it anyway! For $a_n = r^n$ to be a solution of the recurrence we need

$$r^n + pr^{n-1} + qr^{n-2} = 0 \ ,$$

so either $r = 0$, which gives the uninteresting solution $a_n = 0$, or

$$r^2 + pr + q = 0 \ . \qquad (**)$$

Since p and q are known, we can find feasible non-zero values of r simply by solving a quadratic.

Comments.

- The quadratic $(**)$ is called the **characteristic equation** of the recurrence $(*)$.
- A similar method will work for recurrences of higher order. In general, a recurrence of order k will have a characteristic equation whose left hand side involves a polynomial of degree k.

If the characteristic equation of the recurrence $(*)$ has roots α and β, and if A and B are any constants, then

$$a_n = A\alpha^n + B\beta^n$$

is also a solution of the recurrence. To see this, remember that

$$\alpha^2 + p\alpha + q = 0 \quad \text{and} \quad \beta^2 + p\beta + q = 0 \ ,$$

and substitute the formula for a_n back into $(*)$. We get

$$\begin{aligned} a_n + pa_{n-1} + qa_{n-2} &= (A\alpha^n + B\beta^n) + p(A\alpha^{n-1} + B\beta^{n-1}) + q(A\alpha^{n-2} + B\beta^{n-2}) \\ &= A(\alpha^n + p\alpha^{n-1} + q\alpha^{n-2}) + B(\beta^n + p\beta^{n-1} + q\beta^{n-2}) \\ &= A\alpha^{n-2}(\alpha^2 + p\alpha + q) + B\beta^{n-2}(\beta^2 + p\beta + q) \\ &= 0 \ . \end{aligned}$$

It can be proved that if $\alpha \neq \beta$ then the above is actually the general solution of the recurrence, that is, any solution has this form. If $\alpha = \beta$ the above formula gives

$$a_n = A\alpha^n + B\alpha^n = C\alpha^n \ ,$$

and we would not expect this to be the general solution as it contains only one arbitrary constant. In fact, we need a second (independent) solution $a_n = n\alpha^n$ to make up the general solution.

If α and β are complex conjugates (and $\alpha \neq \beta$) then the form of the general solution given above is still valid, but it involves complex expressions and we would normally prefer to replace these by real expressions.

Theorem. *Solving second–order recurrences*. Consider the recurrence

$$a_n + pa_{n-1} + qa_{n-2} = 0 \ ;$$

let the solutions of the equation $r^2 + pr + q = 0$ be α and β.

- If α, β are real and $\alpha \neq \beta$, then the general solution of the recurrence is

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

- If $\alpha = \beta$, then the general solution of the recurrence is

$$a_n = A\alpha^n + Bn\alpha^n \ .$$

- If α and β are complex conjugates we can write them in polar form as $\alpha, \beta = re^{\pm i\theta}$, and the recurrence has general solution

$$a_n = r^n(C\cos n\theta + D\sin n\theta) \ .$$

Comment. The "complex" case is included for completeness but is not examinable.

Hard exercise. Prove the above theorem. *Hints*. First substitute $a_n = \alpha^n b_n$ into the recurrence and simplify to obtain a recurrence for b_n. Then substitute $c_n = b_n - b_{n-1}$ and find a recurrence for c_n. You should find that this recurrence is easy to solve. Then substitute back to find b_n (not easy!) and hence a_n. Along the way it will be helpful to remember that $\alpha + \beta = -p$ and $\alpha\beta = q$. For the third case, use Euler's formula to simplify $(re^{i\theta})^n$.

Examples.

- Find the general solution of the recurrence

$$a_n - 5a_{n-1} - 14a_{n-2} = 0 \ .$$

Also, find the solution which satisfies the initial conditions $a_0 = 4$ and $a_1 = 1$.

Solution. The recurrence has characteristic equation

$$r^2 - 5r - 14 = 0 \ ,$$

with roots $r = 7$ and $r = -2$. Since these roots are real and unequal, the general solution of the recurrence is

$$a_n = A\,7^n + B(-2)^n \ .$$

Up to this point, the constants A and B are unknown. However, substituting $n = 0$ and $n = 1$ into the general solution, and using the given values of a_0 and a_1, we find that

$$A + B = 4 \quad \text{and} \quad 7A - 2B = 1 \ .$$

Therefore $A = 1$, $B = 3$, and the required solution is

$$a_n = 7^n + 3(-2)^n \ .$$

Note. Make sure you understand that $(-2)^n$ is not the same as -2^n !

- Solve the initial value problem

$$a_n = 6a_{n-1} - 9a_{n-2}\ , \quad a_0 = -5\ , \quad a_1 = 6\ .$$

Solution. First we rewrite the recurrence in the usual form

$$a_n - 6a_{n-1} + 9a_{n-2} = 0\ .$$

The characteristic equation

$$r^2 - 6r + 9 = 0$$

has two equal roots $r = 3, 3$ and so the general solution is

$$a_n = A3^n + Bn3^n = (A + Bn)3^n\ .$$

Substituting $n = 0$ and $n = 1$ gives

$$A = -5 \quad \text{and} \quad 3(A + B) = 6\ ,$$

so $A = -5$ and $B = 7$. Therefore, the solution of the initial value problem is

$$a_n = (7n - 5)3^n\ .$$

Exercises. Solve the following initial value problems:

- $a_n + 4a_{n-1} + 4a_{n-2} = 0$, $a_0 = 1$, $a_1 = 10$;
- $a_n + 7a_{n-1} - 8a_{n-2} = 0$, $a_0 = 7$, $a_1 = -11$.

Example. We can find a general formula for the Fibonacci numbers F_n by solving the initial value problem

$$F_n = F_{n-1} + F_{n-2}\ , \quad F_0 = 0\ , \quad F_1 = 1\ .$$

The characteristic equation $r^2 - r - 1 = 0$ can be solved by the quadratic formula to give

$$r = \frac{1 \pm \sqrt{5}}{2}\ ,$$

and since there are two unequal real roots we have

$$F_n = A\left(\frac{1+\sqrt{5}}{2}\right)^n + B\left(\frac{1-\sqrt{5}}{2}\right)^n$$

for some constants A and B. From the initial conditions,

$$A + B = 0 \quad \text{and} \quad A\left(\frac{1+\sqrt{5}}{2}\right) + B\left(\frac{1-\sqrt{5}}{2}\right) = 1\ ,$$

and a bit of careful algebra yields the solution $A = -B = 1/\sqrt{5}$. Hence

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

Comment. For any value of n, the second term in the general solution is (easily!) smaller than $\frac{1}{2}$. But it is obvious from the recurrence that F_n is always an integer, so

$$F_n \quad \text{is the nearest integer to} \quad \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n ,$$

as claimed on page 64.

Higher order recurrences can be solved in the same way. The only difference is that it may be difficult to solve the characteristic equation.

Example. Solve the recurrence

$$a_n + 2a_{n-1} - 13a_{n-2} + 10a_{n-3} = 0$$

subject to the initial conditions $a_0 = 4$, $a_1 = -8$, $a_2 = 10$.

Solution. The characteristic equation is

$$r^3 + 2r^2 - 13r + 10 = 0 \ ;$$

by trial and error $r = 1$ is a root, so we can factorise to give $(r-1)(r^2 + 3r - 10) = 0$ and then solve the quadratic to find the three **characteristic roots** $r = 1,\ 2, -5$. Therefore the recurrence has general solution

$$a_n = A + B2^n + C(-5)^n \ .$$

Substituting $n = 0, 1, 2$ and using the initial conditions,

$$A + B + C = 4 \ , \quad A + 2B - 5C = -8 \ , \quad A + 4B + 25C = 10 \ .$$

Solving these simultaneous equations (*exercise!*), we obtain $A = 9$, $B = -6$, $C = 1$ and hence

$$a_n = 9 - 6 \times 2^n + (-5)^n \ ,$$

the solution of the initial value problem.

In fact, **first order recurrences** can also be solved by using the characteristic equation! The recurrence

$$a_n = \alpha a_{n-1} \tag{$*$}$$

can be rewritten in the form

$$a_n - \alpha a_{n-1} = 0 \ ;$$

it has characteristic equation

$$r - \alpha = 0$$

with one (non–repeated) root $r = \alpha$, and so the general solution is

$$a_n = A\alpha^n \ .$$

However, we usually would not bother doing all this because solving the recurrence in its original form $(*)$ is so simple.

Challenge. See if you can find the general solution of the fourth order recurrence

$$a_n - 6a_{n-2} - 8a_{n-3} - 3a_{n-4} = 0 \ .$$

Recurrences are very important in analysing the **efficiency of algorithms**. An example is the problem of multiplying square matrices: it turns out that the "obvious" way of doing this is not always the best. For simplicity we consider $m \times m$ matrices where m is a power of 2, say $m = 2^n$. To multiply such matrices we can use a method developed by Volker Strassen in 1969. Given two matrices of size 2^n by 2^n,

- split the matrices into quarters, giving eight matrices of size 2^{n-1} by 2^{n-1};
- perform 7 matrix multiplications (using the same method) and 18 matrix additions with these smaller matrices to give another four 2^{n-1} by 2^{n-1} matrices;
- reassemble these to give the answer!

Let a_n be the number of numerical additions and multiplications required to multiply two matrices of size 2^n by 2^n using Strassen's method. Noting that to add two $m \times m$ matrices takes m^2 additions, we have

$$a_n = 7a_{n-1} + 18 \times 4^{n-1} \ ,$$

which can be rearranged to give

$$a_n - 7a_{n-1} = \tfrac{9}{2} \times 4^n \ .$$

This is an example of a first order linear **inhomogeneous** recurrence relation with constant coefficients. We shall see later how to show that this recurrence, together with the initial condition $a_0 = 1$, has the solution

$$a_n = 7^{n+1} - 6 \times 4^n \ .$$

Comments.

- The above procedure will involve some work in splitting and recombining the matrices, but we have ignored this because it will not require any actual arithmetic. However, a full analysis of the algorithm should take these operations into account too!
- **Exercise**. Show that the number of calculations required to multiply two matrices of size 2^n by 2^n by the "obvious" method is

$$b_n = 2 \times 8^n - 4^n \ .$$

  (You don't need to use recursion, just carefully count all the necessary operations.) Using `Maple`, or otherwise, show that Strassen's method takes fewer operations whenever $n \geq 10$.
- For specific information about Strassen's algorithm you could read the detailed explanation at `web.maths.unsw.edu.au/~angell/2410/section1.ps` or just search for `strassen matrix multiplication` on the web.
- There are more recent algorithms which are even better than Strassen's!
- The term "constant coefficients" refers to the coefficients of a_n and a_{n-1}. It doesn't matter that the right hand side is not constant.

Theorem. *Solving inhomogeneous recurrences.* Suppose that the homogeneous linear recurrence

$$a_n + sa_{n-1} + ta_{n-2} = 0$$

has general solution $a_n = h_n$, and that the inhomogeneous recurrence

$$a_n + sa_{n-1} + ta_{n-2} = f(n)$$

has a solution $a_n = p_n$. Then the inhomogeneous recurrence has general solution

$$a_n = h_n + p_n \ .$$

Proof. If $a_n = h_n$ is a solution of the homogeneous recurrence and $a_n = p_n$ is a solution of the inhomogeneous recurrence then $h_n + sh_{n-1} + th_{n-2} = 0$ and $p_n + sp_{n-1} + tp_{n-2} = f(n)$. Therefore, if $a_n = h_n + p_n$ we have

$$\begin{aligned} a_n + sa_{n-1} + ta_{n-2} &= (h_n + p_n) + s(h_{n-1} + p_{n-1}) + t(h_{n-2} + p_{n-2}) \\ &= (h_n + sh_{n-1} + th_{n-2}) + (p_n + sp_{n-1} + tp_{n-2}) \\ &= f(n) \ . \end{aligned}$$

...continued

Conversely, if both a_n and p_n are solutions of the inhomogeneous recurrence, write $h_n = a_n - p_n$. Then

$$\begin{aligned} h_n + sh_{n-1} + th_{n-2} &= (a_n - p_n) + s(a_{n-1} - p_{n-1}) + t(a_{n-2} - p_{n-2}) \\ &= (a_n + sa_{n-1} + ta_{n-2}) - (p_n + sp_{n-1} + tp_{n-2}) \\ &= f(n) - f(n) \\ &= 0 \end{aligned}$$

and so h_n is a solution of the homogeneous recurrence.

Comment. The point of this theorem is that we already know how to find h_n, the general solution of a homogeneous recurrence, by writing down and solving its characteristic equation. So in order to fully solve the inhomogeneous recurrence we only need to find a single solution p_n. This is usually referred to as a **particular solution** and we shall see how to find it by trial and error. Once again the whole procedure is very similar to that employed in solving linear constant–coefficient differential equations.

Finding a particular solution.

Example 1. Find the general solution of the recurrence

$$a_n - 7a_{n-1} + 12a_{n-2} = 30 \ ;$$

also find the solution satisfying the initial conditions $a_0 = 3$ and $a_1 = 1$.

Solution. To solve the homogeneous recurrence we write down the characteristic equation

$$r^2 - 7r + 12 = 0$$

and solve it to get $r = 3$ or $r = 4$. So the homogeneous recurrence has general solution

$$h_n = A3^n + B4^n \qquad (*)$$

where A and B are constants. We'll now **guess a particular solution**: usually a good way to start is to try something like the right hand side of the recurrence. So in this case we guess that the particular solution will be a constant,

$$p_n = c \ ,$$

and we substitute $a_n = p_n$ into the given recurrence:

$$c - 7c + 12c = 30 \ .$$

This gives $6c = 30$ and hence $c = 5$; thus

$$p_n = 5$$

is a particular solution, and the general solution of the inhomogeneous recurrence is

$$a_n = h_n + p_n = A3^n + B4^n + 5 \ .$$

...continued

Now take $n = 0$ and $n = 1$. Using the initial conditions

$$A + B + 5 = 3 \quad \text{and} \quad 3A + 4B + 5 = 1 \ ,$$

that is,

$$A + B = -2 \quad \text{and} \quad 3A + 4B = -4 \ ,$$

so $A = -4$ and $B = 2$. Therefore, the solution of the inhomogeneous initial value problem is

$$a_n = -4 \times 3^n + 2 \times 4^n + 5 \ .$$

Warning! – two common mistakes.

- Don't fall into the trap of thinking that the characteristic equation is $r^2 - 7r + 12 = 30$. The characteristic equation is used to solve the *homogeneous* recurrence and so its right hand side will always be zero!
- You can't find A and B by substituting values of n into equation $(*)$. This is because we are given values of a_0 and a_1, not h_0 and h_1: finding the values of A and B which satisfy the initial conditions must generally be the *last* step of the solution procedure.

Example 2. Find the general solution of the recurrence

$$a_n - 7a_{n-1} + 12a_{n-2} = 30n \ .$$

Solution. Since the left hand side of the recurrence is the same as in the previous example we have again

$$h_n = A3^n + B4^n \ .$$

As the right hand side is a constant times n we might try a particular solution

$$p_n = cn \ .$$

Substituting $a_n = p_n$ into the recurrence,

$$cn - 7c(n-1) + 12c(n-2) = 30n \ , \qquad (*)$$

that is,

$$6cn - 17c = 30n \ ;$$

comparing the coefficients of n gives $6c = 30$, so $c = 5$ and a particular solution is $p_n = 5n$.

Wrong! Why?

We must remember to equate *all* the coefficients in the above equation. From the "n" terms and the constant terms we have

$$6c = 30 \quad \text{and} \quad 17c = 0 \ ;$$

but these simultaneous equations have no solution and the attempt fails. This shows that our guess for p_n was not correct.

If we carefully examine equations $(*)$ on the previous page we can understand why our attempt failed: substituting p_n, p_{n-1} and p_{n-2} into the recurrence produced constants which did not match anything on the right hand side. What we shall do is to introduce a constant term into p_n in the first place, so that we can make all the constants cancel out. Thus

$$p_n = cn + d \;;$$

substituting gives

$$(cn + d) - 7(c(n-1) + d) + 12(c(n-2) + d) = 30n \;,$$

and upon equating coefficients (of *all* terms!) we obtain

$$c - 7c + 12c = 30 \quad \text{and} \quad d + 7c - 7d - 24c + 12d = 0 \;.$$

Simplifying these equations yields

$$6c = 30 \quad \text{and} \quad -17c + 6d = 0 \;;$$

hence $c = 5$ and $d = \frac{85}{6}$, a particular solution is

$$p_n = 5n + \tfrac{85}{6} \;,$$

and the general solution of the recurrence is

$$a_n = h_n + p_n = A3^n + B4^n + 5n + \tfrac{85}{6} \;.$$

Exercise. Find the general solution of the recurrence relation

$$a_n - a_{n-1} - 2a_{n-2} = 6n - 1 \ .$$

As we might suspect from these examples, if the right hand side of the recurrence is a polynomial, we should guess that p_n is a polynomial of the same degree, with unspecified coefficients. The next example shows something different.

Example 3. Find the solution of the recurrence

$$a_n - 7a_{n-1} + 12a_{n-2} = 3 \times 2^n \ ,$$

given that $a_0 = 5$ and $a_1 = 6$.

Solution. Once again we have $h_n = A3^n + B4^n$; we shall try a particular solution similar to the right hand side,

$$p_n = c2^n \ .$$

Substituting,

$$c2^n - 7c2^{n-1} + 12c2^{n-2} = 3 \times 2^n \ ;$$

cancelling 2^{n-2} from both sides we have

$$4c - 14c + 12c = 12$$

and hence $c = 6$. Thus the recurrence has general solution

$$a_n = A3^n + B4^n + 6 \times 2^n \ ;$$

applying the initial conditions yields

$$A + B + 6 = 5 \quad \text{and} \quad 3A + 4B + 12 = 6 \ ,$$

so $A = 2$, $B = -3$ and thus

$$a_n = 2 \times 3^n - 3 \times 4^n + 6 \times 2^n \ .$$

Example 4. Find a particular solution of the recurrence

$$a_n - 7a_{n-1} + 12a_{n-2} = 3n2^n \ .$$

Solution. We could try $p_n = cn2^n$; however, if we remember Example 2 we might realise that we shall also need a constant times 2^n. Therefore we guess

$$p_n = (cn + d)2^n \ .$$

We treat this guess in the same way as previous examples, being careful to get the algebra correct. Thus

$$(cn + d)2^n - 7(c(n - 1) + d)2^{n-1} + 12(c(n - 2) + d)2^{n-2} = 3n2^n \ ,$$

and the best way to start will be to cancel 2^{n-2}, giving

$$4(cn + d) - 14(cn - c + d) + 12(cn - 2c + d) = 12n \ .$$

Then equating coefficients produces the equations

$$4c - 14c + 12c = 12 \quad \text{and} \quad 4d + 14c - 14d - 24c + 12d = 0 \ ;$$

simplifying gives $2c = 12$ and $-10c + 2d = 0$; solving gives $c = 6$ and $d = 30$; and so a particular solution is

$$p_n = (6n + 30)2^n \ .$$

Exercise. Find the general solution of the recurrence relation

$$a_n - a_{n-1} - 2a_{n-2} = (-3)^n \ ,$$

and also of

$$a_n - a_{n-1} - 2a_{n-2} = 2^n \ .$$

The second example on the previous slide may be a bit of a surprise! We attempted to find a particular solution

$$p_n = c2^n \ ,$$

following our previous ideas, but ended up with an impossible equation $0 = 2^n$, indicating that this guess was wrong. How did this happen, and what can we do about it?

Remember that the homogeneous equation

$$a_n - a_{n-1} - 2a_{n-2} = 0$$

has a solution

$$a_n = A(-1)^n + B2^n$$

for any constants A and B. In particular, taking $A = 0$ and $B = c$ we see that our guess $p_n = c2^n$ is a solution of the homogeneous equation. This means, by definition, that when we substitute $a_n = p_n$, the left hand side will simplify to zero, and cannot equal the 2^n on the right hand side.

The situation where an attempted particular solution is also a homogeneous solution is similar to the "repeated root" case where we have two identical solutions and need to modify one of them. So, a rule of thumb is: if our guess for p_n has any term in common with h_n, multiply p_n by n and try again. For the second problem on the previous page, . . .

. . . *continued*

we shall try

$$p_n = cn2^n \ .$$

Substituting into the recurrence,

$$cn2^n - c(n-1)2^{n-1} - 2c(n-2)2^{n-2} = 2^n \ ,$$

and cancelling 2^{n-2} from both sides leaves

$$4cn - 2c(n-1) - 2c(n-2) = 4 \ .$$

This apparently must be wrong because we have terms in n on the left hand side only; however, upon simplification all these terms cancel and we are left with

$$6c = 4 \ .$$

Therefore $c = \frac{2}{3}$, a particular solution is

$$p_n = \tfrac{2}{3}n2^n \ ,$$

and the general solution of the recurrence is

$$a_n = A(-1)^n + B2^n + \tfrac{2}{3}n2^n = A(-1)^n + (B + \tfrac{2}{3}n)2^n \ .$$

Summary. The following ideas may be helpful in finding a particular solution of an inhomogeneous recurrence relation.

1. Guess that p_n will be something like the right hand side of the recurrence.
2. When we substitute p_{n-1} into the left hand side, different kinds of terms may appear which were not in p_n itself. Add extra terms to p_n to allow for these new terms.
3. Compare p_n with the homogeneous solution h_n. (So, it would have been best to calculate h_n *before* starting to think about p_n.) If p_n has any terms in common with h_n, multiply p_n by n. Repeat if necessary.

Example 5. Guess a formula for a particular solution of the recurrence

$$a_n + 2a_{n-1} - 35a_{n-2} = 4 \times 5^n \ .$$

Solution. The characteristic polynomial $r^2 + 2r - 35$ has roots 5 and -7; therefore

$$h_n = A5^n + B(-7)^n \ .$$

So first we would guess something like the right hand side, $p_n = c5^n$; but then we would notice that this is one of the terms in h_n and would change our guess to

$$p_n = cn5^n \ .$$

Example 6. Guess a formula for a particular solution of

$$a_n + 2a_{n-1} - 35a_{n-2} = 72n5^n \ .$$

Solution. As above, $h_n = A5^n + B(-7)^n$. We might first try $p_n = cn5^n$; then we would realise that p_{n-1} will also have terms of the form constant times 5^n, so we adjust the guess to

$$p_n = (cn + d)5^n = cn5^n + d5^n \ .$$

Now this includes a term (the second) from h_n, so we multiply *all terms of* p_n by n to give

$$p_n = n(cn + d)5^n = cn^2 5^n + dn5^n \ .$$

As this now has no terms in common with h_n, we expect it to work.

Example 7. Guess a particular solution of the recurrence $a_n + 7a_{n-1} - 8a_{n-2} = 6n$.

Solution. As on page 75, $h_n = A + B(-8)^n$. Our first guess for a particular solution would be $p_n = cn$, modified to

$$p_n = cn + d\ ;$$

since this shares a term with h_n we would actually try

$$p_n = n(cn + d) = cn^2 + dn\ .$$

Example 8. Guess a formula for p_n if the recurrence is $a_n - 6a_{n-1} + 9a_{n-2} = n3^n$.

Solution. This is a "repeated root" case with

$$h_n = A3^n + Bn3^n\ .$$

We guess first $p_n = cn3^n$, then $p_n = cn3^n + d3^n$, and because this has a term $d3^n$ which is also in h_n, we change it to

$$p_n = cn^2 3^n + dn3^n\ .$$

But this still has a term in common with h_n! So we need to multiply by n again and consider

$$p_n = cn^3 3^n + dn^2 3^n\ .$$

This (at last) has no terms shared with h_n, so it should work.

Exercise. Check that the guesses in examples 5, 6 and 7 do actually work by finding the values of the constants. If you're feeling brave, try example 8 as well.

Exercise. Confirm the formula given at the bottom of page 80 for the number of arithmetic operations required by Strassen's matrix multiplication algorithm.

Solution. We have to solve the initial value problem

$$a_n - 7a_{n-1} = \tfrac{9}{2} \times 4^n \quad \text{with} \quad a_0 = 1 \ .$$

The homogeneous recurrence is $a_n - 7a_{n-1} = 0$, that is,

$$a_n = 7a_{n-1} \ ,$$

and this has general solution $h_n = A7^n$. For a particular solution we try substituting $p_n = c4^n$. This gives

$$c4^n - 7c4^{n-1} = \tfrac{9}{2} \times 4^n \ ;$$

by cancelling 4^{n-1} we have $4c - 7c = 18$, so $c = -6$ and our general solution is

$$a_n = A7^n - 6 \times 4^n \ .$$

The initial condition gives $A - 6 = 1$, so $A = 7$ and the solution is

$$a_n \ = \ 7 \times 7^n - 6 \times 4^n \ = \ 7^{n+1} - 6 \times 4^n \ ,$$

as stated on page 80.

An economic application. An *annuity* is a financial arrangement whereby an amount of money is received at regular intervals. For example, I might wish to invest a sum of money in such a way that I can withdraw \$1000 at the end of each year, with all the initial investment and interest having been exhausted at the end of n years. Assuming that I can obtain 5% interest per annum every year, how much do I need to invest initially?

Solution. Let v_n dollars be the initial amount required (this is referred to as the *present value* of the annuity). If I invest this amount, then one year later I will have earned $0{\cdot}05v_n$ interest for a total of $1{\cdot}05v_n$ dollars; I will then withdraw \$1000, leaving $1{\cdot}05v_n - 1000$ dollars; and this has to be enough to pay for the remaining $n-1$ years of the annuity. Therefore,

$$1{\cdot}05v_n - 1000 = v_{n-1} \ .$$

If I don't want to receive any payments at all then I don't need to invest any money! Therefore we have the initial condition $v_0 = 0$.

Exercise. Show that this initial value problem has solution

$$v_n = 20000\left(1 - \frac{1}{1{\cdot}05^n}\right) \ .$$

Example. We can calculate $v_{30} = 15372$, to the nearest dollar. So, if I want this scheme to operate for 30 years I must initially invest \$15372; over the duration of the annuity I will receive approximately \$14628 in interest to make up 30 payments of \$1000.

Exercises.

- What happens to v_n as $n \to \infty$? What does this mean in economic terms? Does it make sense?
- Generalise this problem by taking the regular payment to be p dollars and the interest rate i. You should obtain the "present value of an annuity" formula

$$v_n = \frac{p}{i}\left(1 - \frac{1}{(1+i)^n}\right) ,$$

  which is well known to economists.
- This formula is obviously not valid when $i = 0$.
  - What is a formula for v_n in the case $i = 0$? (*Hint.* This is easy! – just go back to the original problem and think what it is actually asking.)
  - How does this answer fit in with what we have learnt about solving recurrence relations?

An application to graph theory. Given two different vertices x, y in the complete graph K_{100}, how many walks of length n are there from x to y?

Solution. Let v be the second last vertex in a walk from x to y. Since K_{100} has no loops, v cannot equal y. However, once the second last vertex is known, there is only one way to take the last step to y. So, to count the walks of length n from x to y we need to count the total number of walks of length $n-1$ from x to all vertices $v \neq y$.

If v is different from x we can do this recursively. If $v = x$ we cannot; but then the third last vertex u cannot be x, and we count walks from x to u to x to y.

Let a_n be the required number of walks. First count walks $x \to \cdots \to v \to y$ with $v \neq x, y$.

1. Choose v .... 98 ways
2. Choose a walk of length $n-1$ from x to v .... a_{n-1} ways

Next count walks $x \to \cdots \to u \to x \to y$ with $u \neq x$.

1. Choose u .... 99 ways
2. Choose a walk of length $n-2$ from x to u .... a_{n-2} ways

Putting all this together, we obtain the recurrence

$$a_n = 98a_{n-1} + 99a_{n-2} \ .$$

As this is of second order we require two initial conditions:

$$a_1 = 1 \quad \text{and} \quad a_2 = 98 \ .$$

Exercises.

- Solve this recurrence to show that

$$a_n = \frac{99^n - (-1)^n}{100} \ .$$

- Given two different vertices x and y in K_m, where m is a fixed number, find by similar methods the number of walks of length n from x to y.
- Given a vertex x in K_m, how many walks of length n are there starting at x and finishing at x?