MAT1401: Discrete Methods
Vers. 2.55

Josef Lauri©
University of Malta

Contents

In Memoriam

Lil Gejtanu, li għallimni nħobb ngħodd.

1 What is the course about?

Since this course will be followed mostly by first year students, it is perhaps right to dwell a little on the meaning of the two words in the title, since this might also put other courses into better perspective.

1.1 "Discrete"

The word *discrete* can be largely taken to mean the opposite of *continuous*. For example, most of the mathematics you have been studying in your Advanced Level dealt with functions of a continuous variable, such as, $f(x) = \sin x$. A graph of such a function would look like a continuously drawn curve. In this course, however, the variables we shall generally be concerned with are usually from the set $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ or $\mathbb{N} = \{1, 2, \ldots\}$ or even more simply the finite set $N = \{1, 2, \ldots, n\}$. You are probably used to denoting such a variable (which takes on integer values only) by the letter n (this stands for the first letter in "Natural Numbers" which is the name of the set $\mathbb{N}$). Thus, for example, a plot of the above function $f(n)$ would consist of a series of isolated (discrete) points giving the values of $\sin 1, \sin 2, \ldots$ etc.

What bearing does this have on the type of mathematics we shall be doing in this course? For one thing, it will not be possible to do calculus on functions such as $f(n)$, since this would pre-suppose the ability to take limits arbitrarily close to a given point, whereas $f(n)$ is defined only in discrete jumps (the only limit we might be able to talk about here would be the limit as n tends to ∞). It also means that most of what we shall be doing will be conceptually simpler than for "continuous" mathematics: no limits, no irrational numbers, no tricky questions about infinity (especially if our variables will be limited to the set N).

But this is merely a sampling of what discrete mathematics is *not* about. What sort of things does this branch of mathematics study? One mathematical structure we could study is a graph, or network, which could model a set of towns joined by roadways. The variables involved would include the number of cities and the number of roadways and the interconnections between them, and questions investigated mathematically could be, for example, finding the minimum distance between two given cities. Just as in calculus, we have to find the minimum of some function (distance between the two cities), but new techniques will have to be sought since the functions we are dealing with are discrete. One could study data structures, the algorithms defined on them, and analyse their performance. For example, how many comparisons are required to sort n objects into increasing or decreasing order? One could study ways of sending over a transmission channel a code containing n bits in such a way that, if at most 3 errors, say, are committed by the channel, then the receiver would be able to detect, or even correct, the errors made.

These are only a few of the interesting topics studied in discrete mathematics, and in other courses you would have the opportunity to delve deeper into some of them. In this course we shall be considering something even more basic:

Enumeration. Basically, enumeration is just a fancy way of saying "counting". We shall be studying counting problems. In other words, most of our questions will be of the type "In how many ways can you ...". For example, "In how many ways can you put n objects into k boxes such that no box is empty?"

Such questions are basic to discrete mathematics because they crop up in various contexts: "How many comparisons does a given sorting algorithm make?", for example, is an important question in algorithmics. In probability, when the number of possibilities is finite, calculating the probability of an event is often a question of solving two counting problems—the number of possibilities which give rise to the event and the total number of possibilities—and taking the ratio. Even in statistical mechanics, some problems depend on the number of ways in which photons or particles can occupy a number of energy levels (a question similar to the last one in the preceding paragraph).

What techniques will be needed for such a course? The main pre-requisite is a good mathematical sense obtained during your "A" Level studies together with the love of solving mathematical problems. More specific pre-requisites will be the section usually entitled "Permutations and Combinations" from your "A" Level. Unfortunately, many skip this section because they find that, unlike many other "A" Level topics, the questions involved are not the rôte type questions. Some good news and some bad news for such students: The topics covered in "Permutations and Combinations" in the "A" Level syllabus will be revised in this course—that is the good news. The rest of the news is that this revision will be done rather quickly (two or three lectures at most) and that one thing which you have to learn in order to get through the course is to tackle those questions which are not of the rôte type. But one other bit of encouragement: Although many students find the going a bit tough early on, the great majority always manage to get through the unit, and with reasonably good marks.

1.2 "Methods"

You will soon find out that you can divide most of the mathematics courses in your B.Sc. degree into different families, for example, pure or applied. One other partitioning is between those courses which are more theoretical and others which are based more on developing methods or computational techniques. Actually this is more of a pedagogical distinction than a mathematical one, because good computational methods are grounded on good theory, and valid theory gives rise to effective methods of solving problems. However, it does make some difference for the student and in the way the course is taught. A "methods" course is, in many ways, more similar to what you have been accustomed in your "A" Level—the end of the solution to a problem is an "Answer"—and this is usually obtained by following techniques learnt during the course. A more theoretical course usually follows the format "Theorem-Proof-Theorem-Proof..." In such courses the solution to a problem is generally a whole proof. In a theoretical course little is taken for granted. For example, in the previous subsection I mentioned words like "continuous", "limit", "finite", "infinite" and "irrational" without giving any definitions. I simply relied on our intuition. I would not do this in a theoretical course. But this is a "methods" course and a first-year one, therefore a little less emphasis is placed on defining and proving everything rigorously and we rest a bit more on intuition. All this is done in order to emphasise the honing of our techniques and methods and to avoid missing the computational wood for the theoretical trees, not because we are against proofs in principle. Also, we do this in the full knowledge that what is being left to intuition will be amply covered in the more theoretical courses. Thus, a student who has followed a coherent selection of units from the degree course should see how the theory and methods complement each other. In fact, in the more advanced courses, this distinction between methods and theory will diminish, reflecting more the mathematical way in which these topics should be tackled.

But in this early "methods" course, the implications for the student are that there will be very little theory to learn and memorise, and more emphasis will be placed on learning how to solve problems. This has one important consequence regarding how you should study the course (although this advice, in my view, applies to all mathematics courses). Some students first try to understand the theory and leave the working of set problems to the end (usually close to the examination date). This is bad even for the theoretical courses, because the set problems are generally intended to help you understand the theory, which would be more difficult to do without attempting the problems. But it applies even more to a methods course, where the emphasis is on problem solving. Throughout these notes you will find "Problem Sheets". *THESE MUST BE ATTEMPTED AS YOU GO ALONG WITH THE COURSE.* Some of the problems will be marked by an asterisk. These will be the problems for which we shall have time to go over in more detail during tutorial sessions. But it would be a mistake to skip these problems simply because they will be done in class. Even if you attempt them without success, the time you would have spent going over them and trying to solve them will not be wasted. You would be gaining much more from the tutorial than someone who has not had a good crack at them. Most of the problems in the Problem Sheets are "drill" type problems, to help re-inforce what has been covered in class. Around half way through the course you should start attempting questions from past papers. Exam questions are not strictly of the short drill type which you will find in the problem sheets. They might require you to reproduce some "theory" from the lectures, and the problem would probably borrow ideas from different problems in the Problem Sheets. So you should also try out these past papers when indicated. Doing all this at a regular pace during the course should enable you to learn the elementary basics of discrete enumeration techniques sufficiently well to apply them when needed in other topics (probability, data structures, algorithms, some parts of algebra, etc) and to face the unit exam with confidence

1.3 The number of credits

This course is worth 4 ECTS credits. It will consist of 28 lectures, two lectures every week. After the first few weeks we shall be devoting one lecture a week to tutorials, that is, answering your difficulties in class and working out the marked questions in the problem sheets. Those who miss these lectures will be missing that part of the course which, hopefully, should make things clearer to the student.

The unit's exam paper will contain 4 questions and candidates are to attempt three questions in two and a half hours.

2 Set books

Hundreds of text-books on discrete mathematics have been published in recent years, so it is not easy to make suggestions. First of all: For those who only care about getting the credits for the unit and do not want (or think that they will not need) to hear again about discrete mathematics, these notes should be sufficient. However, most of you would want to read more about these topics, or at least read a more extended version than it is possible to present in class or in these short notes. At least one text book is essential for this.

There are essentially two types of student attending the course. I shall therefore suggest alternative texts. *(Be sure to get the paperback edition of any of these books when available.)*

2.1 B.Sc. & B.Ed. students

One category of student is the B.Sc. student taking Mathematics as a main subject (or Stats. & OR) and I include in this category the B.Ed. student who will be following at least half of the mathematics courses given in the B.Sc. For these I very strongly recommend the book:

***Discrete Mathematics (2nd Edition)* by N.L. Biggs (Oxford University Press)**

I am suggesting this book for this category of student because I shall be using parts of it in other courses that I shall eventually be teaching you, and it also contains sections which can prove useful background reading to courses given by other lecturers. Also, it is written with the “mathematics” student in mind. Therefore it is a good investment. Moreover, it is supported by an OUP website which contains solutions to many of the exercises.

To guide you find those parts which are more relevant to this course, here is a list of more specific reading from Biggs:

- Chapter 5
- Chapter 6
- Chapter 10 Sections 1,4,5
- Chapter 11 Sections 1–5
- Chapter 12 Sections 1,2,4
- Chapter 19 Sections 1,2
- Chapter 25 Sections 1–4
- Chapter 26 Sections 1–4

Of course, there is nothing to stop you reading other parts of Biggs, even during this unit!

I would also highly recommend the book

***Applied Combinatorics (2nd Edition)* by F.S. Roberts & B. Tesman. (Prentice-Hall)**

Although this excellent text does not fit in with the other units I teach as well as Biggs does, I recommend it because it brings the subject to life mainly through its several examples of applications of combinatorial theory. I would particulary recommend the book to future teachers who might be teaching higher forms and who would find very useful the real-life applications of mathematics which this book contains.

Here is a guide to those sections of this book which are more relevant to this course.

- Chapter 1
- Chapter 2 Sections 2.1–2.3, 2.5–2.7, 2.10, 2.14
- Chapter 5 Sections 5.1, 5.3
- Chapter 6 Sections 6.1–6.2
- Chapter 7 Sections 7.1.1, 7.1.2

For those who find that they really enjoy discrete mathematics and combinatorics (a word almost synonymous with discrete mathematics) here are two very good books which you can buy. (The first one is perhaps slightly more difficult than Biggs. The second is, on the other hand, slightly more easy going, but it does not overlap so much with what I do in this unit.)

Combinatorics: Topics, Techniques, Algorithms by Peter J. Cameron (Cambridge University Press)

Aspects of Combinatorics: A Wide-ranging Introduction by V. Bryant (Cambridge University Press)

Finally, for this category of students I would like to mention another excellent reading text:

Mathematics: A Discrete Introduction by E.R. Scheinerman (Brooks/Cole)

As its name suggests, the scope of this book is wider than the scope of these lectures. But if you like mathematics, and if you discover that you like discrete mathematics, you will find that this book not only helps you in other branches of mathematics but is also enjoyable to read.

But I also want to recommend here a book which, in my opinion, is highly suitable to mathematics students and, in particular, to future teachers of mathematics, that is, the B.Ed. students taking this course:

Discrete Mathematics—Elementary and Beyond by L. Lovász, J. Pelikán & K. Vesztergombi (Springer-Verlag)

There are several reasons for suggesting this book. These three authors are world class mathematicians, but they are also superb teachers. The book they have written is a joy to read–if you like mathematics, as I presume you do! The difference, however, between this book and the other two recommended texts, is that there is less of a one-one correspondence between it and my course; in other words, if you are just looking for a text which will help you with this course, you might be disappointed. So why am I recommending this text, particularly for future mathematics teachers? It is because I hope that as teachers you will occasionally be privileged to teach students who are able and want to learn more than is taught in the normal mathematics curriculum. Often, a teacher in this position needs problems and ideas which bring out more of the hidden beauty and usefulness of mathematics. Combinatorics, or discrete mathematics, is one very fertile field of mathematics for finding such ideas. This book is not a collection of challenge problems for the more able children, but it is an education for the reader on how elementary looking ideas in mathematics can be developed into diverse results and applications. It is a book which might help you somewhat marginally in improving your grade in this course, but which I am sure will go a very long way towards making you a better mathematics teacher.

2.2 Students from the Faculty of ICT

The other category of student following this course is made up of those who are following the FICT degree course. The two main texts I suggested above are quite OK even for these students, since in those the authors devote several chapters to applications of discrete mathematics to computer science (algorithms, data structures etc). But I shall be suggesting an alternative text which is slightly less "mathematical" (at least in the chapters we shall be needing) and gives more motivation for the IT student. Moreover, it was in the reading list of some of the IT course units some time ago (this could have changed and I have no control over it) so it could also be a good investment. The book is:

Discrete and Combinatorial Mathematics: An Applied Introduction (4th (or later) edition) **by R.P. Grimaldi (Addison Wesley)**

Again here is a guide to those sections from this book which are more relevant to this course (these refer to the 3rd edition):

- Chapter 1
- Chapter 5
- Chapter 8 Sections 1,2,3
- Chapter 9 Sections 1,2,3
- Chapter 10 Sections 1,2,3

medskip Another book which I strongly recommend to this class of student is the following:

Discrete Mathematics (5th Edition) by Dossey, Otto, Spence and Vanden Eyndon (Prentice-Hall)

These are the sections of this book which are most relevant to this course:

- Section 2.4
- Chapter 8, Sections 8-1–8.6
- Chapter 9, Sections 9.1, 9.2, 9.3, 9.5

One advantage of buying this text is that I shall be suggesting it again for my Networks course which the IT students will be taking next year together with the BSc and BEd students, so it is also a good investment.

Of course, no harm will be done if you buy all three alternative texts, or even all the books I have mentioned!

2.3 For all students

Recently I have come across the following book:

***Combinatorics)* by Balakrishnan (Schaum Series)**

You might be already familiar with the mathematics books in The Schaum Series. They are very useful study aids, based on worked problems. They are also quite inexpensive. In general, they vary in quality, but this is a very good representative for the series, especially for a course of this type, based very much on problem solving. I recommend this book very much. Especially from the point of view of the student who wants a direct no-fuss overview of most of what we shall do in the course, this is an extremely helpful text. Have a look at it in our bookshops and decide for yourselves.

Some of you might find the two main texts I recommended above a little difficult to start off with. In that case I recommend the following text.

***Introductory Combinatorics, 4th Ed* by R.A. Brualdi (Pearson-Prentice Hall)**

Although I am saying that this is somewhat more easy-going than Biggs, I do not mean that it is an “A”-Level standard book (see the next subsection). Brualdi is a first class combinatorialist and a superb teacher. The book basically contains all the material which the main texts I suggested do, but not in the same order and not so much in correspondence with this course—which is another good reason to have it as a secondary text!

2.4 Not cook-books

One final word about these texts. You will find that they are *not* “A” Level texts—the student is not spoon-fed, although the authors go through some pains to describe the subject slowly (they know that they are addressing first year undergraduates). They are not based on the “cook-book approach”: a method is given, followed by one or two examples, then the problems or exercises are just a drill exercise to reinforce the methods learned. This cannot be helped at this level, and I would not have tried to help it even if I could, because now you must begin to read text books which give a true representation of what mathematics is: you learn certain things by drill, you memorise techniques and theory, but mostly you have to think for yourself by applying what you have learnt in order to solve new problems.

3 Past examination papers

As I have already said, apart from working through the problem sheets in parallel with the lectures, around half way through the course you should start looking at and attempting questions from the past examination papers. There you will find out that many questions contain parts which ask for things which you can reproduce from studying the course and these usually hint at what methods will be required to solve the subsequent problem—this usually helps the weaker student who is not so confident at working out unseen problems to get a few marks and gain some confidence during the exam. Some exam problems are similar to those in the problem sheets, and some go slightly beyond, but can always be worked out by the student who has followed the course. Most of these past papers can be downloaded from my web site. Others would be available from the secretary of the Department of Mathematics. Again, to guide you through this (since there are several past papers now) here is a list of suggested problems, roughly divided into three categories corresponding to the topics covered in the course. Those problems marked by an asterisk are somewhat more typical or instructive (this is a subjective opinion). During the later tutorials we should have time to go over some of these questions.

Elementary Counting, Stirling Numbers, Inclusion-Exclusion, Generating Functions

- Jan 99 No. 1 *
- Jan 99 No. 4 *
- Jan 98 No. 1 *
- Jan 98 No. 2(a) *
- Jan 97 No.2 *
- Jan 97 No. 3 *
- Jan 96 No. 1 *
- Jan 96 No. 4(a), (b) *
- Jan 95 No. 1
- Jan 95 No. 2(b)
- Jan 94 No. 1

Recurrence relations

- Jan 99 No. 2 *
- Jan 98 No. 3
- Jan 98 No. 4 *
- Jan 97 No. 1 *
- Jan 96 No. 2(a)

- Jan 95 No. 3
- Jan 94 No. 2

Partitions of an integer

- Jan 99 No. 3 *
- Jan 98 No. 2(b)
- Jan 97 No. 4 *
- Jan 96 No. 2(b)
- Jan 96 No. 4(d)
- Jan 95 No. 2(a) *
- Jan 94 No. 3

The last problem sheet in these notes contains many of the asterisked questions, so this last sheet is definitely not one which you should start attempting at the end of the course!

4 Four motivating problems

In this section I shall pose four simple counting problems which will, in a sense, form a leitmotif throughout the course. After the first couple of lectures you will be able to solve the first problem. The solution to the fourth problem will emerge towards the end of the course (I shall point out, during the course, the moments when these problems become "solvable".)

The basic question is this:

In how many ways can you put six golf balls into four boxes such that no box remains empty?

The four variations of the problem arise if the balls or boxes are distinguishable or identical. By distinguishable balls or boxes I mean that some marks, say $1, 2, 3, \ldots$, are placed on each object. By identical we mean, of course, that there is no way to distinguish between the balls or the boxes. These distinctions give different solutions to the above question. The four variations which therefore arise (in the order in which they will be solvable along the course) are:

1. *Golf balls are identical, boxes are distinguishable*
2. *Golf balls are distinguishable, boxes are identical*
3. *Golf balls are distinguishable, boxes are distinguishable*
4. *Golf balls are identical, boxes are identical*

A few more words about the above four problems: I have chosen the small numbers 6 and 4 so that one can solve these questions by trial and error, that is, by hand, trying all possibilities, without the need for any of the theory or techniques developed in this course. And it would be a very useful (and not very time-consuming) exercise to try and do these problems this way now, because this would help you develop a better "combinatorial" sense. However, think what would happen if, instead of 6 and 4 I had posed the problem with, say, 100 golf balls and 70 boxes. Trial and error would not have been much of a help here, and this is where a more systematic way of solving these problems would be necessary. But we want to be even more ambitious. I would want you eventually to be able to solve the above four instances for this question:

In how many ways can you put n golf balls into k boxes such that no box remains empty?

What is now required is a formula in terms of n and k, or at least an efficient method (algorithm) in which to get the result when the numbers are the general n and k. Definitely, without some theory or systematic way to solve such problems it would be impossible to get at the result. This is one thing which you will learn to do along this course.

5 A little probability

In order to enliven the discussion, I shall occasionally present a problem or an example involving probability. These problems are only meant to highlight some of the counting techniques we will be learning. You will not need to know any probability really, beyond these intuitive notions.

Suppose you are "carrying out an experiment." This could have various meanings, for example: throwing a die and recording the number shown on the uppermost face; knocking on the door of a residence and recording the number of children of each sex living inside the residence; putting three letters in there envelopes in any possible order. We will always assume that the total number of possible results of the experiment is finite, and we denote the set of all possibilities by Ω. For example, in the die experiment,

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

and in the letters-in envelopes-experiment, if we call the letters X, Y and Z, imagine the envelopes to be the "first", the "second" and the "third" envelopes, and keep a record of the order in which the letters were put in the envelopes, then

$$\Omega = \{XYZ, XZY, YXZ, YZX, ZXY, ZYX\}.$$

An "event" in the experiment will be just a subset A of Ω. For example, in the letters-in-envelopes experiment, the event "The letter X is in the second envelope" is the set

$$A = \{YXZ, ZYX\}.$$

Now, *if all the possibilities in Ω are equally likely*, then we say that the probability of event A happening is the ratio of the sizes of A and Ω, that is,

$$Prob(A) = \frac{|A|}{|\Omega|}.$$

There is nothing more to know about probability for this course. The problems encountered will be to determine what Ω should be in order to model the problem under consideration and to find the sizes $|A|$ and $|\Omega|$. If you can understand the following then that is all the probability theory you need to know! (By a fair or unbiased die we mean one for which all the possible outcomes are equally likely.)

1. The probability of getting a 4 if a fair die is rolled is 1/6.
2. The probability of getting an even number if a fair die is rolled is 1/2.
3. The probability of getting an odd prime number if a fair die is rolled is 1/3.

6 Choosing k objects out of n

In this section we shall look at the most elementary and basic of all counting problems (and we shall incidentally cover most of what you should have learnt in the "Permutations and Combinations" part of your "A" Level). The question is:

In how many ways can k objects be chosen from n available distinct objects?

Again there are four variations in which this problem can be posed. (Note: These four ways *do not* correspond to the four motivating problems from the previous section). We can either allow or not allow repeated choice of the same objects, and we can decide to count differently or not the same choice of objects but picked in different order (for example, we can decide that the choice of three letters *cab* or *abc* is different, or the same—if they are considered to be different we say that *order is significant*). These considerations, of course, lead to different results to the above question. We shall deal with these four variations separately, starting from the easiest.

6.1 Repetition allowed, order is significant

Think of the problem as choosing k letters in order from an alphabet containing n letters—therefore you are asked to form all possible k-letter words. How many such words are there in all? Well, the first letter can be chosen in n ways. Having chosen this first letter, *for each of these choices* the second letter can be chosen also in n ways, because we are allowed to repeat letters. Therefore the first two letters can be chosen in n^2 ways. Continuing this way, we see that the first three letters can be chosen in n^3 ways, and in general, the first k letters (which is our problem here) can be chosen in

$$n^k$$

ways, and this is the solution to our problem.

6.1.1 The multiplication and addition rules

Note that we have here used what is called the multiplication rule, which is sometimes described by saying that if we can do one thing in a ways and another in b ways, then *both things can be done simultaneously* in ab ways.

We have to be careful however: We can only apply this rule if we can say that, *for each of the a ways in which the first thing can be done*, the second one can be done in b ways. You must always be very careful that you only apply the multiplication rule when you can say loud and clear the condition which I have highlighted. Very often, mistakes are made because one applies the multiplication rule when the first thing can be done in a ways, but the second thing can be done in b ways for some of these a ways, in c ways for some of the others, etc. In this case, we obviously cannot multiply.

Another source of confusion here is that there is another rule, the addition rule, and sometimes a student asks when do we multiply and when do we add? The addition rule says that if one thing can be done in a ways and another thing can be done in b ways, then *one or another of these things can be done* in $a+b$ ways.

Example 6.1 1. *In how many ways can one choose two books of different languages from amongst 5 in English, 7 in Maltese and 10 in Italian.*

2. *In how many ways can the choice be made if the books need not be of different languages?*

Solution

1. $5 \times 7 + 5 \times 10 + 7 \times 10 = 155$.

2. $22 \times 21 = 462$.

$\square$

To help you draw better the distinction between the multiplication and addition rules, try Exercise 5 and Exercise 6 from Problem Sheet 1.

6.2 Repetition not allowed, order is significant

We can argue as in the previous case. The first object can be chosen in n ways. *For each of these ways* (therefore we can use the multiplication rule) the second object can now be chosen in $n-1$ ways, because repetition is now not allowed. Therefore the first two objects can be chosen together in $n \cdot (n-1)$ ways. Continuing this way, the first k objects (which is our problem here) can be chosen in

$$n \cdot (n-1) \cdot (n-2) \ldots (n-k+1)$$

ways, and this is the answer to our problem.

You might already be familiar with this number, and also know that this way of selecting k out of n objects is called the number of *permutations* of k objects out of n. You might also be familiar with notation such as ${}^nP_k$ for this number. The notation which we shall use will be

$$[n]_k$$

and we shall call this number the *falling factorial*. When $k = n$, then

$$[n]_n = n(n-1)(n-2) \ldots 2 \cdot 1$$

and we denote this number by $n!$, calling it "n factorial". Note that $[n]_k$ can be written in terms of factorials as,

$$[n]_k = \frac{n!}{(n-k)!}$$

and this is a formula with which you are most probably also already familiar from your "A" Level.

6.2.1 Stirling's approximation

Evaluating $n!$ is no easy task—as n becomes large $n!$ increases very rapidly. This last sentence could form a good part of a whole course, but we shall not spend much more time on it here. We shall only mention that there is a very good approximation to $n!$ called *Stirling's Approximation*, and it says,

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

Here, $f(n) \simeq g(n)$ means that $f(n)/g(n)$ tends to 1 as n tends to ∞, that is, $f(n)$ is closer in value to $g(n)$ the larger n is. When $n = 10$, the percentage error made by Stirling's approximation is high: 83%. But when $n = 100$, say, the percentage error is only 0.09%. This approximation for $n!$ makes it possible to obtain very good estimates of the values of $n!$ and of expressions in which it appears.

Sometimes, a rougher, but maybe easier to use, estimate for $n!$ is applied, namely that

$$\left(\frac{n}{e}\right)^n \leq n! \leq \frac{(n+1)^{n+1}}{e^n}.$$

Although the proofs of these approximations are not terribly difficult, we shall not give them here since they require calculus.

Example 6.2 *The approximation*

$$n! \simeq \left(\frac{n}{e}\right)^n$$

is a very useful estimate for $n!$. Using this estimate show that

$$\binom{n}{k} \simeq \frac{n^n}{(n-k)^(n-k)k^k}.$$

Using a mathematical computational package such as Mathematica*, investigate how good an approximation for $\binom{n}{k}$ this is for large values of n and k.*

6.3 Repetition not allowed, order not significant

Let x be the number of ways in which the k objects can now be chosen. Having listed these x ways of choosing the k objects, suppose we decide that order will become significant. Then, *each one of these x ways* (so we can use multiplication rule) will, by jumbling up in all possible ways the k objects selected, give us $k!$ different selections with order significant. That is, we would have $x \cdot k!$ ways of choosing the k objects, without repetition, but order significant. But we have just seen that this is equal to $[n]_k$. Therefore

$$\begin{aligned} x \cdot k! &= [n]_k \\ x &= \frac{n!}{(n-k!)k!}. \end{aligned}$$

Therefore our required answer is

$$\frac{n!}{(n-k)!k!}$$

and we shall write this, for short, as

$$\binom{n}{k}$$

calling it "n choose k". This is the famous *binomial number* or *binomial coefficient*, with which you are almost certainly familiar from your "A" Level. We shall have more to say about the binomial coefficient in a subsequent section.

Example 6.3 *In some problems it is clear whether, in the given counting problem, order is or is not significant, sometimes because the problem specifically says so. But sometimes you have to decide for yourself. This is true, for example, in probability problems. Consider the following elementary example but which, nevertheless, gave problems to mathematicians when the theory of probability was still in its infancy.*

Suppose that you know that a family has two children but you know nothing else about the family except that it is equally likely that a child is a boy or a girl. Let p denote the probability that the two children in the family are both boys.

1. *Why do you think that some argue that $p = 1/2$?*
2. *Why do you think that some argue that $p = 2/3$?*
3. *Which answer do you think is correct? Remember the italicised requirement above about the set Ω.*
4. *What simulation would you carry out to verify which is the correct result. Does this method of verification indicate an interpretation of what "probability" really means? Do you know any results in probability theory which justify this interpretation? Do you know if this interpretation is controversial even amongst mathematicians? (These last questions are for those students who know more probability theory than is explained in Section 5.)*

Example 6.4 *Do not be mislead by the wording of a problem. When we say that we are counting "unordered" selections or that order is not "significant" we mean that, in the situation we are interested in, any selection will be counted once, if if that selection can be ordered in several different ways.*

For example: How many three-letter words are there? By a "three-letter word" we mean here any string of three distinct letter written in alphabetical order.

In spite of superficial appearance, the answer is $\binom{3}{26}$*, that is, the number of* unordered *choice of three distinct letters. Just think a few seconds about it. Every choice counts only once not six times, because it is presented in a unique pre-determined order. This unique order means that* the different ways of ordering the three chosen letters are not counted as different choices, that is, order is not significant for the purpose of counting here.

6.4 Repetition allowed, order not significant

The previous three cases should all have been familiar to you from your previous studies. This fourth case is slightly more difficult to solve, and it is, in fact, probably the first item you are meeting in this course which is not a revision of what you have done in your "A" Level.

First of all note that we cannot use the multiplication rule here. Yes, the first choice can be made in n ways, and for each of these ways, the second choice can be made in another n ways (since repetition is allowed). But we are already overcounting, because we are counting choices such as ab and ba to be different, whereas order is not significant here. You might think, well let us use the same system we employed for calculating $\binom{n}{k}$ based on $[n]_k$, that is, divide by the number of repetitions. However, in this case, it is not true that *every choice is repeated the same number of times.* (This emphasised text has the same importance as the emphasis we place on the phrase *for every one of the previous choices* when applying the multiplication rule—we can be said, after all, to be applying the "division principle" here.) Some choices like ab are repeated twice while others, like aa, appear only once. One could carefully consider those which are repeated twice, divide their number by two, and add the rest, and this would work well here. But what happens when we come to a choice of three letters? Now, some choices, like abc, are repeated six times, others are repeated three times, others are not repeated at all. And the situation becomes more complex when four objects are chosen. And what about the case when k objects are chosen? It seems clear that proceeding this way will not give us a nice neat formula for the result we are seeking. And all this because, just as in those situations where we cannot use the multiplication rule, we cannot here say that *all* choices are repeated the same number of times.

We therefore have to solve this problem more carefully. The technique we shall use is to translate the problem into a second counting problem, show that the second counting problem has the same solution as the first, and then show that there is an easy way to solve the second problem. This is a very important technique in combinatorics. Seeing it applied for the first time, most students feel that the solution is like a magic rabbit drawn out of a hat, without any hint as to how one should be stumble on the idea, if not by accident. This student would worry, not without some justification, that presented with such a problem himself, he would be at a complete loss how to proceed. Well, the best advice is that you are seeing this type of mathematics for the first time, and that is the reason why the technique appears to be all trickery. But when you would have seen other problems solved in this way, the mystery begins to dissolve, and you would eventually be able to find the "tricks" yourself.

(In fact, the mystery never really dissolves. This is what a mathematician does for a living: Finding clever new ways of solving problems. The better the mathematician, the more difficult the problems she would be tackling, and the more clever the "tricks" have to be. But even a world class mathematician builds his cleverness upon an accumulation of techniques learnt gradually over the years. And in any case, for the student, especially a first year student taking his first mathematical steps, the advice of the previous paragraph is generally what is expected of him.)

But let us now continue with solving the problem at hand. Let us reconsider the problem as one of a distribution of k objects into n boxes. For example, if $k = 4$ and $n = 3$ and the n objects are the letters $\{a, b, c\}$, then the choice, say, $aabc$ (remember that repetition is allowed but order is not important), can be represented as a distribution of four identical objects into the three boxes marked a, b, c, respectively: Two objects go into box marked a, and one object goes into each of boxes marked b and c. Note that this representation ensures that the number of choices is the same as the number of distributions of the objects into the boxes. So we might as well count the latter in order to solve our problem.

Suppose we want to represent each distribution by some code. We can decide to represent each object by the symbol $\times$. We shall also decide which box will have its symbols listed first in our code, which listed second, and so on. We shall show the separation between boxes by the symbol $|$ (hence, only 2 separation symbols $|$ will be required in this case). Therefore the above choice would be represented by the code

$$\times \times | \times | \times .$$

If the choice had been, say, $aacc$, then the code representing it would have been

$$\times \times || \times \times .$$

Note that we need four symbols $\times$ (corresponding to $k = 4$) and two separation symbols $|$ (one less than the number $n = 3$ of boxes). Note also that again we have that the number of codes is the same as the number of distributions: So we might as well count the number of codes.

How many codes of this type are there? Well, each code has a total of six symbols, and a code is determined once we decide where to put the four symbols $\times$. And this can be done in

$$\binom{6}{4} = 13$$

ways. This is then the solution to our problem when $k = 4$ and $n = 3$.

We can now repeat the same argument for general k and n. It all boils down to determining the number of codes containing exactly k symbols $\times$ and $n - 1$ symbols $|$ (for a total of $n + k - 1$ symbols). This is equivalent to the number of ways in which we can choose the positions of the k symbols $\times$, and this can be done in

$$\binom{n+k-1}{k}$$

ways. This is the solution to our problem, that is, the number of ways in which we can choose k objects out of n allowing repetition but disregarding order.

The above result will be *very* important throughout the rest of the course. You will be meeting it under different guises. You will at first forget the result, or fail to see the connections, but I expect that by the end of the course this result will become second nature to you!

Example 6.5 *You can now do the first of the "four motivating problems" posed earlier on.*

6.4.1 Choosing objects versus distributing objects

The above technique of transforming a problem involving choosing objects into one involving distributing objects into boxes is quite general. In fact, the problem:

In how many ways can you choose k objects out of n?

can be translated into the equivalent problem:

In how many ways can k objects be distributed into n boxes?

Each one of the boxes is made to correspond to one of the n objects, and a choice of one of the n objects in the first problem is represented by putting one of the k objects into the corresponding box. In this way, k boxes out of the n are "chosen". The main question is how to translate the conditions of repetition/non-repetition and ordered/unordered.

Well, the first question is easy. If repetition is allowed in the choice problem, then putting more than one object into the same box is allowed in the distribution problem. For the second question, if order of choice is to be taken into consideration, then the k objects will be numbered $1, 2, \ldots, k$, respectively. This way, the order in which an object is chosen is represented by placing an appropriately numbered (corresponding to the order) object into the corresponding box. But if order is not significant, then the objects distributed into the boxes will be identical.

It is best to illustrate by some examples. Suppose that we are required to choose $k = 3$ objects out of $n = 5$. Let the five objects be $\{a, b, c, d, e\}$. Suppose first that order is significant and repetition is allowed. Consider the choice ada. This will be represented by putting objects marked 1 and 3 into box a and object marked 2 into box d. The "code" repersenting this distribution will be $13|||2|$.

Now consider the same choice ada where, this time, order is not significant. This would be represented by placing two objects into box a and another object into box d, where all the objects will be identical. The code for this distribution would be $\times\times|||\times|$.

Now consider the choice dac, where order is significant, but there is no repetition. This would be represented by placing object numbered 1 into box d, object 2 into box a and object 3 into box c. The code for this distribution would be $2||3|1|$.

Finally, consider the same choice dac but where now the order is not important. This would be represented by placing three identical objects one each into boxes a, c and d. The corresponding code would be $\times||\times|\times|$.

The study of distribution of objects in boxes under various conditions is of some importance in statistical mechanics (see Exercise 8).

7 The binomial coefficient and a first look at the interplay between algebra and combinatorics

A whole combinatorics course can be built around the binomial coefficient. We shall, of course, not be so ambitious here, but we shall open a small window on the many properties of the binomial coefficient and we shall use this discussion as a vehicle in order to introduce some very important techniques involving the interplay between algebra and combinatorics.

7.1 Two simple identities

The simplest identity involving the binomial coefficient is surely the following:

$$\binom{n}{k} = \binom{n}{n-k}.$$

We can prove this identity in two ways. One way is purely algebraic: Use the formula for the binomial coefficient to write $\binom{n}{k}$ as $n!/(n-k)!k!$, and similarly $\binom{n}{n-k}$ as $n!/k!(n-k)$, and clearly these quantities are equal.

We can also see that the above identity is true combinatorially, that is, by interpreting the binomial coefficient as the number of ways of choosing objects in a certain fashion, without resorting to an algebraic formula for the coefficient. OK, how does the argument go. Well, $\binom{n}{k}$ means the number of choosing without order or repetition k objects out of n—think of it as choosing the best k players out of a pool of n in order to play in an important match. But choosing the k players who will be playing is exactly equivalent to choosing the $n-k$ who will not be playing, that is, the number of ways of doing the former is equal to the number of ways of doing the latter, in other words, $\binom{n}{k} = \binom{n}{n-k}$.

Let us now look at another less trivial (but still simple) identity involving the binomial coefficient. This is:

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

The algebraic way to prove this runs as follows. Using the formula for the binomial coefficient, write the right hand side of the proposed identity as

$$\frac{(n-1)!}{k!(n-k-1)!} + \frac{(n-1)!}{(k-1)!(n-k)!}$$

then simplify it (take the least common multiple of the denominators, etc—all this is left as an easy exercise for the reader) to obtain $n!/k!(n-k)!$, which is the formula for $\binom{n}{k}$, which is the left hand side. Therefore the identity has been proved.

Now for the combinatorial proof. The left hand side counts the number of ways of choosing k objects out of n without order or repetition. Suppose the n objects are $\{1, 2, 3 \ldots, n\}$. Then let us divide all these possible choices of k objects into two types: Type I all those in which "1" is chosen, and Type II all those in which "1" is not chosen. Clearly, $\binom{n}{k}$ is equal to the sum of all choices of Type I and Type II (because no choice is left out this way, and no choice is counted more than once, since Type I and Type II are "disjoint").

So let us count the number of choices of Type I. We now have $n-1$ objects to choose from, and we need to choose $k-1$ objects, since "1" has already been chosen. So the number of ways to do this is

$$\binom{n-1}{k-1}.$$

Let us consider the Type II choices. Again we have $n-1$ objects to choose from, but we now still need to choose k objects, because all we know is that "1" is not to be chosen. The number of ways to do this is clearly

$$\binom{n-1}{k},$$

and therefore adding these two results gives the required identity.

You might well ask here: Why do we need two ways to prove the same result? There are many valid reasons for this. First of all, in mathematics, it is generally a font of new results and ideas when we can get to the same place by different routes. Very often, different expressions for the same thing are obtained, and by equating the different expressions new identities are discovered. Also, the different routes can give different insights into why the result is true. Note above that the algebraic method gave us an almost mechanical way of proving our results—algebra took over and we did not need to think very hard on why the results were true. This is often one advantage of the algebraic way when dealing with more difficult and complex identities. On the other hand, the combinatorial method forced us to interpret the meaning of the coefficients and hence of the identity we were proving. This often gives a deeper understanding of the results we are proving, although this is sometimes more difficult to achieve.

In our context, there is another very compelling reason for learning these two types of proofs. We shall soon encounter another counting number like the combinatorial coefficient, but for which *no closed formula is known.* In spite of this we shall be proving an identity very similar to the second one above. Clearly, in this case, since no formula is available, we do not have the luxury of both the algebraic and the combinatorial proof, but we must instead know how to make the latter work.

7.2 Pascal's Triangle and a first look at recurrence relations

Deriving combinatorial identities is not simply a fun exercise in itself. Very often these identities can be used for practical purposes, such as computing the value of some binomial coefficients. This is the case with the second identity above. We have already noted that computing factorials is no mean task for large numbers. However, we can obtain the value of the binomial coefficient for reasonable values of n and k (and for much higher, with a computing machine) by using this identity. Notice that this identity gives us the binomial coefficient in terms of itself, therefore we seem to be going round in circles. However, note that the parameters on the right hand side are smaller than those on the left hand side. That is, we can find what $\binom{n}{k}$ is if we know what it is for smaller values of n and k—a sort of cascade effect going backwards. Such an identity is called a *recurrence relation.* However, in order to stop this cascade effect from going backwards without ever stopping, we need to know the value of $\binom{n}{k}$ from some starting values. A recurrence relation always comes equipped with such starting values, and they are usually called *initial conditions* or *boundary conditions.* These initial conditions have to be obtained not from the recurrence relation (Of course! Why?), but from the nature of the problem itself. In our case this is easy. For any value of n, there is only one way to choose all n objects (remember, no repetition, no order), and only one way to pick no object at all. Therefore, our initial conditions are

$$\binom{n}{n} = 1 \qquad \text{and} \qquad \binom{n}{0} = 1.$$

Starting from $\binom{0}{0} = 1$ or even $\binom{1}{0} = \binom{1}{1} = 1$ we can then calculate $\binom{2}{k}$ for all appropriate values of k, then $\binom{3}{k}$, $\binom{4}{k}$ and so on, using the above identity, and stopping when we hit boundary conditions. You are already familiar with how this is done in this case, and the values obtained in this order form what is called *Pascal's Triangle*. Here are the first few lines from Pascal's Triangle. The kth number (starting from $k = 0$) in the nth row (also starting from $n = 0$, which is the top row, consisting of just "1") is the value of $\binom{n}{k}$.

$$\begin{array}{ccccccccccc}
&&&&&1&&&&& \\
&&&&1&&1&&&& \\
&&&1&&2&&1&&& \\
&&1&&3&&3&&1&& \\
&1&&4&&6&&4&&1& \\
1&&5&&10&&10&&5&&1
\end{array}$$

The above discussion on recurrence relations and initial conditions should have reminded you of two topics you have already encountered (one, if you are not studying any computing), namely, proofs by induction and recursive programming. These three topics (recurrence relations, proof by induction and recursive programming) are different facets of the same mathematical idea. In fact, at this stage, computing students should find it instructive to try out the next example (unless you have already been assigned this or similar exercises).

Example 7.1 *Write a recursive program which will input n and k and compute $\binom{n}{k}$ using the above recurrence relation. Write also a non-recursive program to do this. Is the recursive program easier to write? On the other hand, which is the most efficient program in terms of calculations involved?*

7.3 The Binomial Theorem—Our first generating function

You should already know that the binomial coefficients are the main components of the, so-called, *Binomial Theorem*, which says:

Let x be any real number and n a positive integer. Then

$$\begin{aligned}(1+x)^n &= 1 + x + \binom{n}{2}x^2 + \binom{n}{3}x^3 + \ldots + x^n \\ & \sum_{k=0}^{n} \binom{n}{k} x^k.\end{aligned}$$

Again, there are two ways to prove this. The algebraic way is by induction. Assume the result is true for n. Then, for $n+1$, since $(1+x)^{n+1} = (1+x)(1+x)^n$, we need to multiply the right-hand-side by $(1 + x)$, collect like terms (using the recurrence relation which we have just discussed for the binomial coefficients) and get the formula for $n + 1$.

We shall not dwell any longer on this because we want to discuss in more detail the other method of proof, the combinatorial method. (The interested reader is invited to fill in the details of the above proof by induction, or else to look it up in some good "A" Level textbook.) So, imagine writing out $(1+x)^n$ as a product of n brackets $(1+x)$. Each coefficient in the expansion of this product is obtained by "visiting" each of the brackets once and picking either 1 or x and multiplying the choices together. Therefore the term x^k appears every time we choose x exactly k times and 1 the rest of the times. The number of ways this can happen is equal to the number of terms x^k which, when collected together, give x^k multiplied by its coefficient. Therefore this coefficient is the number of ways of picking k objects (x's) out of n possibilities (given by the n brackets). Repetition is not allowed (we can only choose one object from each bracket) and order is not important, since multiplication is commutative. Therefore the required number of ways (that is, this coefficient) is equal to $\binom{n}{k}$, and the result is proved.

We can also think of the binomial expansion in this way: The expression $(1+x)^n$ is storing, in a compact algebraic expression, all the binomial coefficients $\binom{n}{k}$. It is as if carrying with us $(1+x)^n$ means that we have in our pocket all these binomial coefficients. All we need to do to extract one or any of these coefficients is to expand $(1+x)^n$ and, if we need $\binom{n}{k}$, then we look up the coefficient of x^k. This process of expanding an algebraic expression $p(x)$ and taking the coefficient of the term x^k is so often used that we have a special notation for it. Thus, instead of saying "the coefficient of x^k in the expansion of $p(x)$" we write

$$[x^k]p(x).$$

Therefore,

$$\binom{n}{k} = [x^k](1+x)^n,$$

and this is just another way of writing the Binomial Theorem. But here we are emphasising the role of $(1+x)^n$ as a "carrier" of the binomial coefficients, rather than seeing the theorem as a way of expanding $(1+x)^n$ which, incidentally, gives $\binom{n}{k}$ as coefficients.

When an expression is viewed like $p(x)$ above as a carrier of its coefficients, we say that $p(x)$ is the *generating function* of its coefficients. Of course, given a generating function we might still need to do a lot of work in order to extract its coefficients, and we shall see situations where this cannot be done completely. However, we shall see that even in such cases, having the generating function without knowing exactly what the coefficients are can also be very useful in order to get results.

In this section we have started from known coefficients (the binomials) and arrived at their generating function. In solving counting problems we often need to do the reverse, that is, construct a generating function whose coefficients are the answer to our counting problem. The question remains: Are we able to expand the generating function in order to get at the coefficients?

We shall see an example of this in the next section.

7.4 Using generating functions to solve counting problems

Suppose we have the following counting problem. We are to choose five letters from the letters a, b, c, d with order not important, but we are given that the letter a can be chosen up to twice (including the possibility of not being chosen), the letters b and d can each be chosen at most once, while the letter c must be chosen at least once but not more than three times. In how many ways can a choice of five letters be made subject to these restrictions?

Let us try and solve this by taking a leaf out of the combinatorial proof of the Binomial Theorem. Let us write down a product of four brackets, one each for the letters a, b, c, d, respectively, and let us do this in such a way that, when multiplying out these brackets, the coefficient of x^k will be the number of ways of choosing k letters under these conditions (we are interested in $k = 5$, but that is just an example). The product would be

$$(1 + x + x^2)(1 + x)(x + x^2 + x^3)(1 + x).$$

The first bracket contains these powers of x: 0, 1 and 2, since a can be chosen zero times, once or twice. Similarly for the other brackets. In particular, notice that the third bracket does not contain the constant term 1. This is because c cannot be chosen zero times, that is, it must be chosen at least once. It should not be difficult to convince yourself that the coefficient of x^k in the above expansion is equal to the number of ways in which we can choose k letters out of a, b, c, d under the given restrictions. In other words, the above expression is the generating function for our counting problem, and we can write, using our notation defined previously, that the number of ways of choosing k letters in this problem is equal to

$$[x^k]x(1 + x)^2(1 + x + x^2)^2,$$

after simplifying a little the expression.

Example 7.2 *Find the coefficient of x^5 in the expansion of the above expression.*

Solution This is now a simple "A" Level problem, and the result is 8. □

Sometimes we require what is sometimes called a *full inventory* of the choices possible, not just their number, that is, a complete listing of the possible choices. Algebra can also help us to achieve this. First, let us decide that a choice of, say, two a's, one b and two c's will be written as a product a^2bc^2. Then, to get the full inventory we replace the above expression by

$$(1 + ax + a^2x^2)(1 + bx)(cx + c^2x^2 + c^3x^3)(1 + dx),$$

and the coefficient of x^k (which will now not be a number but a sum of products of the letters a, b, c, d) will actually list the possible choices of k letters from a, b, c, d under the given conditions.

Example 7.3 *Draw up a list of the possible choices of five letters in the above example.*

Solution This is now given by the coefficient

$$[x^5](1 + ax + a^2x^2)(1 + bx)(cx + c^2x^2 + c^3x^3)(1 + dx),$$

and although it is still an easy problem in multiplication of algebraic variables, more work needs to be done here (of course, we are asking for more information). This is left as an exercise to the reader. Notice that, in this case, finding this coefficient is really no more or less different or difficult than actually listing the possible choices. $\square$

7.5 Returning the favour: using the result of a counting problem to obtain an algebraic result—the Binomial Theorem with negative integral exponent

The problem we would now like to solve is a purely algebraic one, namely, to find an expression for the expansion of

$$(1 - x)^{-n}$$

where n is a positive integer. That is, we shall be proving the Binomial Theorem for negative integral exponent. However, we shall be using mainly counting techniques to do this (we shall actually use a result we have proved above), although this is, of course, not the only way to prove this theorem.

First consider the expression $(1 - x)^{-1}$. If we assume that $-1 < x < 1$ then this is the infinite sum of the geometric progression

$$1 + x^2 + x^3 + x^4 + \ldots$$

We shall henceforth assume that all conditions required for convergence of any infinite series we might meet do hold (this is, after all, a "methods" course, although all these convergence arguments can be made completely rigorous and water-tight—but this treatment is enough for a first course in generating functions).

Well, back to our expression. If we write out $(1 - x)^{-n}$ as a product of n brackets, each a G.P., as

$$(1 + x^2 + x^3 + x^4 + \ldots)(1 + x^2 + x^3 + x^4 + \ldots) \ldots (1 + x^2 + x^3 + x^4 + \ldots)$$

then our problem becomes that of finding the coefficient of x^k in this expansion. It looks rather hopeless, doesn't it, multiplying n infinite expressions! But look at it this way. We are required to collect k x's (x^k) out of n possibilities (the brackets). But each possibility allows us to: not choose x (choose the term 1), or choose one x, or choose two x's (x^2), etc. So the problem can be re-worded as follows: In how many ways can you choose k objects out of n, order not significant, and repetition allowed without limit (except that you cannot, of course, make more than k choices in all)? But we know the answer to this, it is $\binom{n+k-1}{k}$. And this is the required coefficient of x^k, and so we have obtained the Binomial Theorem we were aiming for. We shall write down this theorem in different equivalent ways, not only so that you can make the extra effort to remember this very important result, but also so that we can revise the different notations we have developed up to now.

The Binomial Theorem for negative integral exponent
Let x be a real number, $-1 < x < 1$, and let n be a positive integer. Then

$$\begin{aligned}(1-x)^{-n} &= 1 + \binom{n}{1}x + \binom{n+1}{2}x^2 + \binom{n+2}{3}x^3 + \dots \\ &= \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^k.\end{aligned}$$

Alternatively, we can say that the coefficient of x^k in the expansion in powers of x of $(1-x)^{-n}$ is

$$\binom{n+k-1}{k}$$

or, using the shorthand notation we have defined for such situations,

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

7.6 Probability generating functions

This section requires more knowledge of probability than what is explained in Section 5 and may be omitted.

The use of generating functions in probability is very important, and there they are called probability generating functions, because the coefficients of the x^k, that is, the numbers which they generate, are probabilities. Let us consider a simple example which illustrates a non-trivial use of generating functions.

Let A be a random variable which can take on the values $\{0, 1, \dots, n\}$ and let the probability $P(A = k)$ that A is k be denoted by p_k and let it be equal to

$$\binom{n}{k} p^k q^{n-k}$$

where $q = p - 1$. Let us form the generating function

$$g(x) = \sum_0^n p_k x^k = \sum_0^n \binom{n}{k} p^k q^{n-k} x^k.$$

It is clear that

$$g(x) = (q + px)^n$$

by the Binomial Theorem. In fact, the random variable A is said to have the binomial distribution. Now, to what use can we put the function $g(x)$? One thing which we would like to calculate for a given random variable is its mean. For A this is equal to

$$\sum_0^n kp^k.$$

We can evaluate this sum by some simple manipulations with $g(x)$. Note that, by differentiating $g(x)$ term-by-term,

$$\frac{d}{dx} g(x) = \sum_0^n k \binom{n}{k} p^k q^{n-k} x^{k-1} = \sum_0^n k p_k x^{k-1},$$

And substituting $x = 1$ in this sum gives exactly the mean of A. Therefore let us differentiate $g(x) = (q + px)^n$ with respect to x and then substitute $x = 1$. Since $g'(x) = n(q + px)^{n-1}$ this gives

$$g'(1) = np(q + p.1)^{n-1} = np,$$

since $p+q = 1$. Therefore the mean of A is np, and we got this result quite easily, without having to evaluate any more summations, but just by manipulating $g(x)$.

8 Counting functions

In this section we shall look at some of the previous counting problems in a slightly different guise. This will have several aims. Firstly, as we have already commented, looking at the same problem from a different angle helps us to gain more insight into the problem. Secondly, these next few sections will enable us to look at possible the single most important concept in mathematics: functions. Of course, in other courses you will be doing much more on functions—here we shall only be interested in counting particular types of functions. But reviewing the basic definitions will help to consolidate what you are learning in other units. Finally, this point of view will lead us to a totally new counting problem, a new combinatorial coefficient, and the opportunity to apply some of the previous principles and techniques in a different setting.

First of all, what is a *function*? We shall not give the most rigorous definition here, but simply say that a function involves two sets A and B (often, these sets are subsets of the real numbers, like intervals; and we are not excluding the possibility that $A = B$; A is called the *domain* of the function and B is called its *co-domain*) and it is a rule which assigns to *every* element in A a *unique* element in B. This situation is often described by the notation

$$f : A \to B$$

and if an element $a \in A$ is mapped by f into the element $b \in B$ we denote this either by

$$f : a \mapsto b$$

or

$$f(a) = b.$$

The set of all elements of B on which the elements from A are mapped is called the *range* of the function. Note that the range is a subset of B—it can consist of only one element, or, at the other extreme, it might be equal to all of B.

Notice the two emphasised words in the definition of a function. No element of A can fail to be mapped onto something in B, (although not all elements in B need have some element mapped onto them) and no element in A can be mapped into more than one element from B. Thus, for example, the "function" $f(x) = \pm\sqrt{x}$ is *not* a function. The two functions $f(x) = |\sqrt{x}|$ and $g(x) = -|\sqrt{x}|$ however are legitimate functions. The best way to remember all this is perhaps by keeping in mind the two "forbidden" pictures for a function, that is, what a function cannot be (see Figure 1).

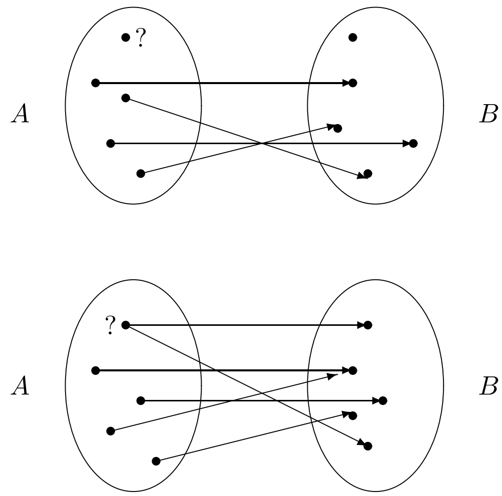

Figure 1: The forbidden pictures for functions

The most basic counting question which now arises is the following. Suppose A and B are finite sets with $|A| = k$ and $|B| = n$. How many functions are there from A to B?

Well, suppose $A = \{a_1, a_2, \ldots, a_k\}$. Any function f is completely determined once we know $f(a_i)$ for all $1 \leq i \leq k$. In how many ways can $f(a_1)$ be chosen? Clearly in n ways, because a_1 can be mapped into any element of B. Now, *for each of these choices* the value of $f(a_2)$ can also be chosen in n ways (remember, in the forbidden pictures of a function we are not excluding the possibility that different elements of A are mapped into the same element of B and neither that some elements of B are "left out"). Therefore the first two elements of A can, between them, be mapped into B in n^2 ways. Continuing this way we conclude that all the elements of A can be mapped into elements of B in n^k ways, and each of these ways gives a different function. And there are no more! Therefore the number of functions from a k-set to an n-set is

$$n^k.$$

Of course, we have already seen this expression in the very first counting problem we have studied in this course, and this is because a function from A to B is simply another way of choosing k elements (the $f(a_i)$) from the n elements of B. Repetition is allowed by the definition of a function, and order is important, because, if we, say, interchange the values of $f(a_1)$ and $f(a_2)$, then we get a different function.

8.1 Counting injections

The definition of a function is very lenient. Sometimes we require more restrictions. For example, we might require that no different two elements of A are mapped into the same element of B. When a function has this property we say that it is *one-to-one* or *injective*. The forbidden picture for injective functions (called *injections*) is shown in Figure 2.

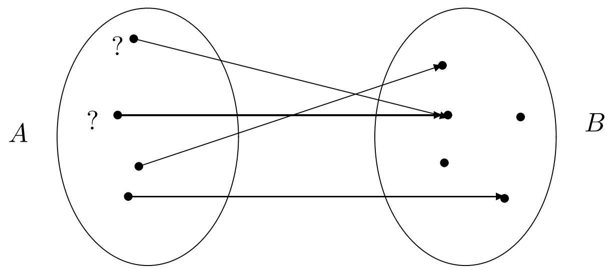

Figure 2: The forbidden picture for injective functions

Sometimes we require that no element of B is left unmapped, that is, for any element $b \in B$ there must be at least one element $a \in A$ such that $f(a) = b$. Such functions are said to be *onto* or *surjective*. The forbidden picture for surjective functions (called *surjections*) is shown in Figure 3.

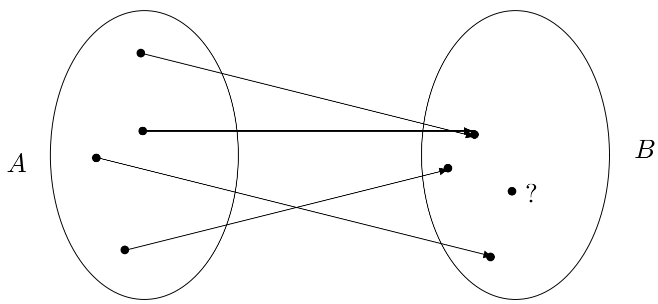

Figure 3: The forbidden picture for surjective functions

A function which is both injective and surjective is called a *bijective function* or a *bijection* or a *one-one-correspondence*.

Example 8.1 *When there is a bijection between two* <u>*finite*</u> *sets, then the two sets have the same number of elements. (If this common number is n, then the number of bijections between the sets is $n!$—why?) Thus bijections are one of the two most important counting tools in combinatorics (the other is generating functions!).*

Can you find examples of two infinite sets A and B such that one is apparently "larger" than the other but it is still possible to construct a bijection from A to B?

Solution Consider the set A of positive integers and the set B of even positive integers (that is, B is a subset of A). But $f : A \to B$ defined by $f(a) = 2a$ is a bijection. This seemingly strange phenomenon does not occur with finite sets.

That the number of bijections between two n-sets is $n!$ can be seen in various ways, one being that $[n]_n = n!$. □

In this section we shall be counting injections. Counting surjections is more difficult, and we shall tackle that in the next section.

Thus, let A and B be finite sets with $|A| = k$ and $|B| = n$. How many injective functions are there from A to B?

Again, suppose $A = \{a_1, a_2, \dots, a_k\}$, and again we note that any function f is completely determined once we know $f(a_i)$ for all $1 \leq i \leq k$. In how many ways can $f(a_1)$ be chosen? Clearly in n ways, because a_1 can be mapped into any element of B. Now, *for each of these choices* the value of $f(a_2)$ can be chosen in $n-1$ ways (remember, in the forbidden pictures of an injective function we are excluding the possibility that different elements of A are mapped into the same element of B). Therefore the first two elements of A can, between them, be mapped into B in $n(n-1)$ ways. Continuing this way we conclude that all the elements of A can be mapped into elements of B in $n(n-1)\dots(n-k+1)$ ways, and each of these ways gives a different function. And there are no more! Therefore the number of functions from a k-set to an n-set is

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} = [n]_k,$$

the falling factorial.

We have already seen this expression in the second counting problem we have studied in this course, and this is because an injective function from A to B is simply another way of choosing k elements (the $f(a_i)$) from the n elements of B. Repetition is not allowed by the definition of an injective function, and order is, as explained above, important.

8.2 Counting surjections

Recall that a surjection from A to B is one for which the range is all of B—that is, every element of B has at least one element from A mapped onto it. Counting surjections is a trickier problem. Suppose that $|A| = n$ and $|B| = k$ (we have reversed the roles of n and k, not that it makes any difference to the problem, but so that our final result will agree with the notation used to solve a related problem). Let $B = \{b_1, b_2, \dots, b_k\}$. For any surjection f from A to B, each element b_i will have at least one element (possibly more) from A which is mapped onto it by f. So, let $f^{-1}(b_i)$ denote the set of all those elements of A which are mapped onto b_i. The subsets $f^{-1}(b_1), f^{-1}(b_2), \dots, f^{-1}(b_k)$ of A have two very important properties (which stem from the two "forbidden" diagrams for a function):

1. *No two subsets overlap, in other words they are disjoint.* This is because, by the definition of a function, no element of A can be mapped by f onto two different elements of B.

2. *The union of all these subsets covers all elements of A.* This is because, again by the definition of a function, no element of A can fail to be mapped onto some element of B.

These two properties of a system of subsets of a set A (whether or not the subsets arise from a surjection) are so important in mathematics that we give such systems a special name. Thus, if a system of subsets of a set A satisfies the above two properties, then we call the system of subsets a *partition of the set* A. The subsets in the partition are often called *parts*. From a counting point of view we can see why partitions are particularly useful to work with: if we know the number of elements in each subset of the partition, then, adding all these numbers, gives us the total number of elements in A. This is obviously not true if the subsets do not form a partition, and in a later section we shall see the difficulties in dealing with such a situation.

But let us go back to our problem, that of counting surjections. We have just seen that every surjection from the n-set A to the k-set B determines a partition of A into k parts. On the other hand, a partition of A into k parts can determine several surjections from A to B, because now the situation is this. We have the k parts of the partition, and these are to be mapped one-to-one onto the k elements of B. This mapping can be done in $k!$ ways. Therefore, if we let $S(n,k)$ denote the number of partitions on an n-set into k parts (we do not yet have any idea what value $S(n,k)$ might have) then, since *every* partition leads to $k!$ surjections, the number of surjections is therefore

$$k! \cdot S(n,k).$$

Of course, to make this result in any way meaningful we have to find $S(n,k)$. This is what we shall attempt to do in the next section.

8.2.1 Counting partitions of a set: Stirling Numbers of the Second kind and "Stirling's" Triangle

The numbers $S(n,k)$, called *Stirling numbers of the second kind* are very similar to the binomial coefficients (they "count" something, each depends on two parameters, and we shall see more significant similarities soon). However, let us say at the outset that, unlike for the binomial coefficient, we shall not be able to obtain a nice closed formula for $S(n,k)$.

Values for $S(n,k)$ can be found "by hand" for small values of the parameters n,k. In particular, it is easy to see that

$$S(n,1) = 1$$

and

$$S(n,n) = 1.$$

These will be the initial or boundary conditions of the basic recurrence relation for Stirling numbers which is

$$S(n,k) = S(n-1,k-1) + kS(n-1,k) \quad (2 \le k \le n-1).$$

We immediately have these comments to make about this recurrence relation.

1. We have stated the relation without any proof. This will be remedied soon below.
2. There is a very close similarity with the recurrence relation for the binomial coefficient which gave us Pascal's Traingle. The only difference is the appearance of a k multiplied to one of the terms on the right hand side.

3. Just as with the analogous relation for the binomial coefficients, we can use this relation to compute recursively values of $S(n, k)$, and this can be done by constructing a "triangle" of values similar to Pascal's Triangle. Here are the first few values of this triangle, which we shall call "Stirling's Triangle"—the task of checking that these values can be obtained recursively from the above recurrence relation is left to the reader (note the lack of symmetry which we had in Pascal's Triangle).

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 3 & & 1 & & & \\
 & & 1 & & 7 & & 6 & & 1 & & \\
 & 1 & & 15 & & 25 & & 10 & & 1 & \\
1 & & 31 & & 90 & & 65 & & 15 & & 1
\end{array}
$$

So now, what we have to do is to proof the recurrence relation for Stirling numbers.

Recall that for the binomial coefficients we had the luxury of two proofs: an algebraic and a combinatorial proof. However, since we do not have (at least as yet) a formula for the Stirling numbers, the algebraic proof is not available. We must therefore proof the relation combinatorially.

Let us proceed this way. Let the n-set A which is to be partitioned into k parts be $\{1, 2, \ldots, n\}$. Let us divide all partitions of A into k parts into two types, Type I and Type II. Type I will contain all those partitions in which $\{1\}$ is one of the k parts, whereas Type II will contain all those partitions in which the element "1" appears in a part together with some other element or elements of A. Clearly, if we add the number of partitions of Type I and of Type II we would obtain the total number of partitions of A into k parts, that is, $S(n, k)$.

Now, how many partitions are there of Type I? One part, that is $\{1\}$ has already been determined. So the other parts can be chosen in $S(n-1, k-1)$ ways, because we now need only to subdivide the resulting set into $k-1$ subsets, and the set has an element (the element 1) missing, that is, it has $n-1$ elements.

Now, how many partitions are there of Type II? Again, the base set has an element missing, therefore we are partitioning an $(n-1)$-set, and we need to obtain k parts, because now the element 1 does not form a partition on its own. This can be done in $S(n-1, k)$ ways. But we still need to put back the element 1. (In the Type I case, there was only one way to do this, by placing it alone as a subset $\{1\}$.) This element can be put in any one of the k parts, each such choice giving a different partition. That is, *each of the above $S(n-1, k)$ partitions* then leads to k possible partitions of A by putting 1 back. This gives $k \cdot S(n-1, k)$ partitions.

Adding the number of partitions of Types I and II and equating to $S(n, k)$ gives the required recurrence relation.

9 Problem Sheet 1: Elementary Counting

1. *

   (a) In how many ways can you deal a bridge hand (13 cards) from a full pack (52 cards)?

(b) In how many ways can you choose a Chairman, Secretary and a Treasurer from a 10-person committee?

(c) How many binary numbers (consisting of 0's and 1's) are there with eight digits?

(d) How many different *patterns* can be obtained by throwing three indistinguishable dice? (By "patterns" we mean that, for example, $4, 4, 6$ and $4, 6, 4$ are considered to give the same pattern.)

2. $*$ A family has four children. It is assumed that a chid is equally likely to be a boy or a girl. What is more likely, that the family has two boys and two girls or that it has three children of the same sex? (You can see that, even with very elementary counting ideas, counter-intuitive results in probability are accessible.)

3. *

(a) In how many ways can k indistinguishable golf balls be coloured with any one of n given colours?

(b) In how many ways can k identical objects be put into n boxes, where each box can accommodate any number of objects, including none?

(c) What is the number of solutions of the equation

$$t_1 + t_2 + \ldots + t_n = k$$

in non-negative integers $t_1, t_2, \ldots, t_n$?

(d) Consider the effect of multiplying out the following n bracketed expressions:

$$(1 + x + x^2 + \ldots)(1 + x + x^2 + \ldots)\ldots(1 + x + x^2 + \ldots).$$

What is the coefficient of x^k in the resulting expansion?

(e) How many different terms are there in the expansion of

$$(w + x + y + z)^5?$$

4. *

(a) In how many ways can a total of 16 be obtained by rolling four dice once? (Here, configurations such as $4, 4, 3, 5$ and $4, 3, 4, 5$ are to be considered different.)

(b) Calculate the coefficient of t^{12} in the expansion of

$$\left(\frac{1 - t^6}{1 - t}\right)^4.$$

(c) Express the number $N(n, p)$ of ways of obtaining a total of n with p dice as a coefficient in a suitable product of binomial expansions.

5. * How many times is the word "Hello" written by the following program fragment?

```
For i:= 1 to 10 do
   Writeln ("Hello");
For j:= 1 to 10 do
   Writeln ("Hello");
For k:= 1 to 10 do
   Writeln ("Hello");
```

6. * How many times is the word "Hello" written by the following program fragment?

```
For i:= 1 to 10 do
   For j:= 1 to 10 do
      For k:= 1 to 10 do
         Writeln ("Hello");
```

7. * How many times is the word "Hello" written by the following program fragment?

```
For i:= 1 to 10 do
   For j:= 1 to i do
      For k:= 1 to j do
         Writeln ("Hello");
```

   What would have been your answer if there were $r(\geq 1)$ *For* loops instead of three in this last program segment?

8. In statistical mechanics this type of problem occurs: A system consists of N particles. Each of the particles can be in one of two states, A or B.

   (a) * In how many possible states can the system of particles be?

   (b) * Suppose that a particle in state A has zero energy while a particle in state B has energy 1. In how many of these possible states does the system have total energy E?

   Now suppose that each particle can be in any one of three possible states, A, B or C, and suppose that the energies associated with these states are 0, 1 and 2, respectively. Answer (a) and (b) above in this case.

9. * Prove that:

   (a) $\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n$.

   (b) $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$.

   (c) $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^n\binom{n}{n} = 0$.

   (d) $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n-1}{k} + \ldots + \binom{k+1}{k} + \binom{k}{k}$.

   (e) $\binom{n}{0} + \frac{1}{2}\binom{n}{1} + \frac{1}{4}\binom{n}{2} + \ldots + \frac{1}{2^n}\binom{n}{n} = \left(\frac{3}{2}\right)^n$.

   (f) $\binom{n}{1} + 2\binom{n}{2} + \ldots + n\binom{n}{n} = n2^{n-1}$.

   (g) $\binom{n}{0} + \frac{1}{2}\binom{n}{1} + \ldots + \frac{n}{n+1}\binom{n}{n} = \frac{1}{n+1}(2^{n+1} - 1)$.

10. (a) In how many ways can a pack of 52 cards be dealt out to four players so that each gets at least one card, but the four do not necessarily get the same number of cards?

    (b) In how many ways can a pack of 52 cards be dealt out to four players so that each gets a full hand? (See Section 9.2 below.)

11. $*$ (This problem tests your understanding of various ideas we have covered: the addition and the multiplication rules, the ordered and unordered selections, repetition and no repetition, using generating functions to solve counting problems, and the binomial and multinomial (see Section 9.2 below) coefficients.)

    Determine the number of different 4-letter words which can be constructed from the letters in MISSISSIPPI. Here a word means a selection of not necessarily distinct letters such that different orders of the letters selected are counted as different words. Solve the problem as follows. Find, using inventory generating functions, how many *unordered* words there are of each type, where, by a type, we mean, for example, words with two S's one P and one I, or, two I's and two P's, etc. Then, for each type, find how many orderings there are using the multinomial coefficient (see Section 9.2 below).

(The general technique for solving this type of problem is the use of *exponential generating functions*. Some of the suggested texts cover exponential generating functions, and you are encouraged to have a look if you are interested. The generating functions we discuss in these notes are called *ordinary generating functions*.)

12. (See Section 9.1 below.)

    (a) Show that in a party of six people there are either three mutual acquaintances or three persons none of whom knows any of the two others.

    (b) Show that in a party of people there are at least two persons with the same number of acquaintances within the party. (In these last two questions assume that if person a knows person b then b also knows a.)

    (c) Show that in a group of 37 people there are always at least 4 who were born in the same month of the year.

    (d) Suppose five points are chosen inside an equilateral triangle of side-length 1. Show that there is at least one pair of points whose distance apart is at most 1/2.

    (e) $*$ (This is called the "Birthday Paradox." Again we see a counter-intuitive result in probability accessible with very simple counting techniques.)

    Show that in a group of at least 23 persons it is more likely than not (that is, there is more than 50% probability) that at least two persons in the group have the same birthday. (Assume 365 days in a year and that all days are equally likely to be birthdays.)

13. (a) For a given positive integer n, a *partition* of n is an equation

    $$n_1 + n_2 + \ldots + n_k = n$$

    in positive integers n_i. For example, 5+1 and 3+1+1+1 are two partitions of 6, but $1+5$ is considered to be the same partition as $5+1$. The number of partitions of n is denoted by $p(n)$; k is said to be the number of *parts* of the partition. The number of partitions of n with k parts is denoted by $p_k(n)$.

    Find, by listing all the partitions, $p(6)$.

    (b) What is the coefficient of x^6 in the expansion of

    $$(1 + x + x^2 + \ldots)(1 + x^2 + x^4 + \ldots)\ldots(1 + x^6 + x^{12} + \ldots)?$$

    Can you write down $p(n)$ as a coefficient of x^n in some suitable expansion?

    (c) In how many ways can Lm1 be exchanged for 25 cent, 10 cent and 5 cent coins? What is

    $$[x^{100}](1 - x^5)^{-1}(1 - x^{10})^{-1}(1 - x^{25})^{-1}?$$

14. (a) Write down the first four terms and the general term in the expansion of the power series $(1 - x)^{-3}$.

    (b) Find the general term in the power series $\frac{1+3x}{(1-x)^2}$.

(c) Find

$$[x^n]\frac{1+2x+2x^2}{1-3x+3x^2-x^3}.$$

Selected answers

1. (a) $\binom{52}{13}$; (b) $[10]_3$; (c) 2^8; (d) $\binom{6+3-1}{3}$.
2. (a)–(d) $\binom{n+k-1}{k}$; (e) $\binom{4+5-1}{5}$.
3. (a), (b) $\binom{15}{3}-4\binom{9}{3}+6$; (c) $[t^{n-p}](1-t^6)(1-t)^{-p}$.
4. 30.
5. 10^3.
6. $\binom{10+3-1}{3}$; with r (≥ 1) loops: $\binom{10+r-1}{r}$.
7. (a) 2^N; (b) $\binom{N}{E}$; (c) $[x^E](1+x+x^2)^N$.
9. (a) $4!S(52,4)$; (b) $\frac{52!}{13!13!13!13!}$.
11. (a) 11; (b) 11; (c) 29.
12. (a) General term: $\binom{r+2}{r}x^r$; (b) $(4n+1)x^n$; (c) $\frac{1}{2}(5n^2+3n+2)$.

9.1 The Pigeonhole Principle

Parts of Problem 12 involves the use of the so-called Pigeonhole Principle:

> *If more than n objects are placed into n pigeonholes, then at least one pigeonhole must contain more than one object*

A slight extension of this principle says:

> *If more than rn objects are placed into n pigeonholes, then at least one pigeonhole must contain more than r objects*

(The Pigeonhole Principle is effectively saying that $[n]_k = 0$ if $k > n$.)

We shall not dwell too much in this course on this principle or its applications. Suffice it to say that it crops up in various branches of mathematics, particularly combinatorics, and that although it is perhaps one of the most trivial mathematical principles in its bare form, it is often cleverly employed to obtain results which seem to be very far removed from what the principle trivially asserts. Parts of Problem 12 illustrate simple (although clever) ways in which the Pigeonhole Principle can be used, or is somehow disguised in some seemingly unrelated problem. Those of you who will eventually decide to do a Combinatorics Elective in the final year of the B.Sc. will encounter a topic, "Ramsey Theory", which, in many ways, can be considered to be a very non-trivial extension of the Pigeonhole Principle.

9.2 The multinomial coefficient

Problem 10b involves what is called the "multinomial coefficient". We shall limit ourselves to a brief discussion here. The suggested texts all give a more extended treatment.

Suppose that we are required to arrange n objects into k boxes so that exactly n_i objects go into box i, $i = 1, \ldots, k$, order is not important, and $n_1 + n_2 + \ldots + n_k = n$ (that is, each object is placed in some box). Then the number of ways of doing this is

$$\frac{n!}{n_1!n_2!\ldots n_k!}.$$

This number is denoted by

$$\binom{n}{n_1, n_2, \ldots, n_k}$$

and it is called the *multinomial number* or *multinomial coefficient*. A few points will be, very briefly, summarised here about this number.

1. The binomial number can be viewed as a special of the multinomial number, since the former can be written as

$$\binom{n}{k} = \binom{n}{n-k, k}.$$

Can you give a combinatorial interpretation of this?

2. Just as the binomial coefficient is the main component of the Binomial Theorem, then so is the multinomial ceofficient the main component of Multinomial Theorem, which is an extension of the Binomial Theorem: Let $x_1, \ldots, x_k$ be real numbers and n a positive integer. Then

$$(x_1 + x_2 + \ldots + x_k)^n = \sum_{r_1+\ldots+r_k=n} \binom{n}{r_1, r_2, \ldots, r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}.$$

3. The number $\binom{n}{n_1, n_2, \ldots, n_k}$ is equal to the number of partitions of an n-set into k parts such that one part contains n_1 elements, another contains n_2 elements, etc.

The binomial coefficient can also be defined as follows. Let there be a total of n objects such that n_1 are identical to each other, n_2 are identical to each other, and so on up to n_k. Therefore $n = n_1 + n_2 + \cdots + n_k$ In how many ways can these objects be permuted (order significant, no repetition)?

If no two elements were the same, the result would be $n!$. But among these $n!$ permutations every arrangement of the n_1 elements which are the same is counted $n_1!$ times instead of once. Therefore since each arrangement of these elements is counted $n_1!$ times, instead of once, the total number of permutations taking into consideration that they are identical is $n!/n_1!$. Arguing similarly for the other sets of identical objects gives a total of

$$\frac{n!}{n_1! n_2! \ldots n_k!},$$

and this is the multinomial coefficient.

10 The Inclusion-Exclusion Principle

Suppose we have two finite sets A, B and we know their respective sizes $|A|, |B|$. What is $|A \cup B|$? The first temptation would be to answer $|A| + |B|$. But only a moment's thought will reveal that this is wrong in general, because if there are elements in both A and B (that is, in the intersection $A \cap B$), then these elements will be counted twice. And we want all elements in $A \cup B$ to be counted once and only once. Well, the solution is to subtract a count for each one of these elements in the intersection, and we would be ok. This then gives,

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

Let us try this for three sets (it would be helpful if you were to draw a Venn Diagram). If we write

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

then we could be overcounting several elements which are in the intersections of these sets. So, our "second approximation" to the result would be to take off the sizes of the intersections in order to compensate for the overcount:

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

All elements are counted exactly once (which is what we are aiming for) except those elements which are in all three sets. Each of these elements is first counted three times ($|A| + |B| + |C|$), then removed three times, so in total it is not counted at all. To compensate for this undercounting we need to add another term, and the formula is then correct:

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

I am sure that you are now beginning to see the pattern (and to understand why this is called "inclusion-exclusion"). We are therefore ready to state and proof the result for n sets.

The Inclusion-Exclusion, or the Sieve, Formula
Let $A_1, A_2, \ldots, A_n$ be finite sets, and let α_i for $1 \leq i \leq n$ denote the sum of the sizes of all intersections of these sets taken i at a time. Then

$$|A_i \cup A_2 \cup \ldots \cup A_n| = \alpha_1 - \alpha_2 + \ldots + (-1)^{n-1}\alpha_n.$$

Proof We shall take an arbitrary element x in $A_i \cup A_2 \cup \ldots \cup A_n$ and show that it is counted exactly once by the RHS.

Suppose that x is contained in exactly t of the sets, $1 \leq t \leq n$. Then how many times is x counted by α_1? Clearly the answer is t. How many times is it counted by α_2? Well, every time we take an intersection of two sets from the t in which x lies, and this is $\binom{t}{2}$. Similarly in α_i x is counted $\binom{t}{i}$ for $i \leq t$, and in α_i, for any $i > t$, it is not counted at all since, in this case, the intersection would contain at least one set in which x does not lie.

Therefore the RHS of the result we have to prove counts x the following number of times,

$$\begin{aligned}
& t - \binom{t}{2} + \binom{t}{3} + \ldots + (-1)^{t-1}\binom{t}{t} \\
= \quad & \binom{t}{1} - \binom{t}{2} + \binom{t}{3} + \ldots + (-1)^{t-1}\binom{t}{t} \\
= \quad & 1 - 1 + \binom{t}{1} - \binom{t}{2} + \binom{t}{3} + \ldots + (-1)^{t-1}\binom{t}{t} \\
= \quad & 1 - \left(\binom{t}{0} - \binom{t}{1} + \binom{t}{2} - \binom{t}{3} + \ldots + (-1)^{t}\binom{t}{t}\right) \\
= \quad & 1 - (1-1)^t \\
= \quad & 1,
\end{aligned}$$

as required. □

In the next three sections we shall see some applications of the inclusion-exclusion principle in cases where its use is not immediately obvious.

10.1 An application in Number Theory

We shall now make a short excursion into Number Theory. This branch of mathematics deals with properties of the integers, their factorisations, prime numbers and their distributions, solutions of equations in which the variables can only take integral value and several other problems of this type. Of course, ours will be a very minor excursion, and what we shall present will only be scratching the surface of this vast field. However, it will enable us to use inclusion-exclusion in what will at first seem to be quite an unexpected way.

The problem is to find, for any positive integer n, the number of integers $\{1, 2, \ldots, n\}$ which are relatively prime to n. This number will be denoted by $\phi(n)$, and it is often called *Euler's ϕ-function.* Recall that two integers are said to be *relatively prime* if they have no common factor apart from 1; for example, 15 and 22. A pair of relatively prime integers need not be prime themselves, as this example shows.

Let us first start with a simple concrete example, finding $\phi(60)$. Note that to find the *factors* of 60 is quite easy. Just write the prime factorisation of $60 = 2^2 \times 3 \times 5$, and then consider all those numbers which have at least one of these prime numbers as factors. These factors of 60 are quickly listed knowing these prime factors and are shown in Figure 4. This figure, in fact, gives the lattices of factors of 60, in which a number m is joined by a straight line to a number n below it if n is a factor of m.

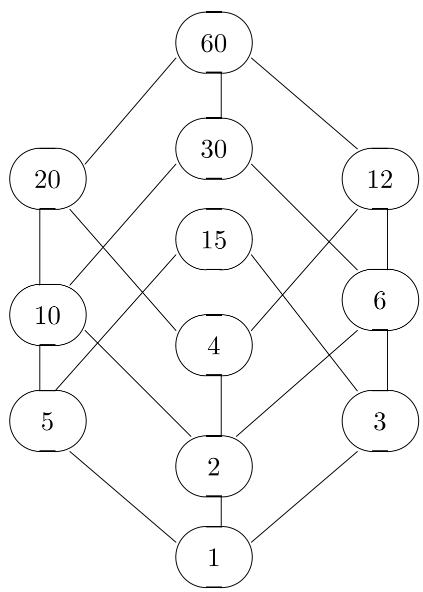

Figure 4: The lattice of factors of 60

To find $\phi(60)$ is, however, more tricky. Let start by defining those sets of numbers which we do not want to count, and then we shall subtract them from all possibilities (You will soon learn that this is a common trick when working with the inclusion exclusion principle.)

Thus,

let A be the set of numbers in $\{1, 2, \ldots, 60\}$which have 2 as a factor
let B be the set of numbers in $\{1, 2, \ldots, 60\}$which have 3 as a factor
and let C be the set of numbers in $\{1, 2, \ldots, 60\}$which have 5 as a factor.

Our required $\phi(60)$ is simply

$$60 - |A \cup B \cup C|$$

and this is where the inclusion-exclusion formula comes in. Let us first compute the cardinalities of the respective sets (You will see how much easier it is to count those numbers which we do not want rather than those which we do.)

Clearly, $|A| = \frac{60}{2} = 30$, $|B| = \frac{60}{3} = 20$ and $|C| = \frac{60}{5} = 12$. Also, $|A \cap B| = \frac{60}{2 \cdot 3} = 10$, $|A \cap C| = \frac{60}{2 \cdot 5} = 6$, and $|B \cap C| = \frac{60}{3 \cdot 5} = 4$. Finally, $|A \cap B \cap C| = \frac{60}{2 \cdot 3 \cdot 5} = 2$.

We now have all the values we need to expand $|A \cup B \cup C|$ using the inclusion-exclusion formula, giving,

$$\phi(60) = 60 - (30 + 20 + 12) + (10 + 6 + 4) - (2) = 11$$

We now have seen enough to enable us to attack the general problem, that is, finding $\phi(n)$. So, suppose writing n as a product of prime factors gives

$$n = p_1^{e_1} p_2^{e_2} \ldots, p_t^{e_t}$$

. Let $N = \{1, 2, \ldots, n\}$, and let A_i (for $i = 1, 2, \ldots, t$) be the set of all those numbers in N which have p_i as a factor. Clearly

$$\phi(n) = n - |A_1 \cup A_2 \cup \ldots \cup A_t|.$$

Now, as we had when n was 60, each $|A_i|$ equals $\frac{n}{p_i}$. Similarly $|A_i \cap A_j| = \frac{n}{p_i p_j}$, and so on for the other intersections.

Therefore, by the inclusion exclusion formula,

$$\phi(n) = n - \alpha_1 + \alpha_2 - \ldots + (-1)^t \alpha_t$$

where the α's have their usual meaning as above when we proved the inclusion-exclusion formula. From the example with the number 60, it is not difficult to see that each α_i equals n divided by the sum of all possible products of i distinct prime factors of n, and again it is not difficult to see that this gives

$$\phi(n) = n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2}) \ldots (1 - \frac{1}{p_t})$$

which is our required formula.

10.2 Derangements

We know that there are $n!$ permutations of the numbers in the set

$$N = \{1, 2, \ldots, n\}.$$

In some of these permutations the number 1 could find itself in the first position, or the number 2 in the second position or, in general, the number i might be placed in the ith position. In this section we are going to count all those permutations in which this never happens, that is, all those permutations of N in which no $i \in N$ appears in its natural (that is, the ith) position. Such a permutation is called a derangement, and the number of derangements of the set N is denoted by d_n.

This problem is often presented in more popular language. The most usual setting is the following. Suppose a secretary has n letters which are to be put in n addressed envelopes, and suppose that he or she does this in such a way that no letter is put in the correct envelope. In how many ways can this be done. Clearly, the answer is d_n, as defined above, and which we shall now proceed to find a formula for.

So, as before, let us count those permutations which are not "allowed". Thus, let A_i, for $1 \leq i \leq n$, be the set of all those permutations in which the number i occurs in its natural position. Therefore,

$$d_n = n! - |A_1 \cup A_2 \cup \ldots \cup A_n|$$

and, by the inclusion-exclusion formula, this equals

$$n! - \alpha_1 + \alpha_2 - \ldots + (-1)^n \alpha_n$$

where the α's have their usual meaning.

Consider first $|A_1|$. Since the number 1 is fixed (in the first position) this set contains all permutations of the other $n-1$ elements. Therefore $|A_1| = (n-1)!$. And similarly, for all the other sets, $|A_i| = (n-1)!$. Therefore,

$$\alpha_1 = n \cdot (n-1)!$$

since there are n sets with this size.

Similarly, consider $|A_1 \cap A_2|$. Here two numbers (1 and 2) are fixed, and therefore the size of this intersection is equal to $(n-2)!$. Therefore,

$$\alpha_2 = \binom{n}{2} \cdot (n-2)!$$

since there are $\binom{n}{2}$ intersections of the sets $A_1, \ldots, A_n$ taken two at a time, and they all have size $(n-2)!$.

It should now be easy to write down the value of α_i. There are $\binom{n}{i}$ of choosing the i sets to be intersected, and each intersection has size $(n-i)!$. Therefore

$$\alpha_i = \binom{n}{i} \cdot (n-i)! = \frac{n!}{i!}$$

Therefore

$$\begin{aligned} d_n &= n! \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + (-1)^n \frac{1}{n!} \right) \\ &\simeq \frac{n!}{e}. \end{aligned}$$

Note that the last line is only an approximation, since we can evaluate (to e^{-1}) the summation $\sum \frac{1}{n!}$ only when this sum is to infinity. Yet this is a good approximation since the series converges very rapidly. In fact, the first six terms give the value of $1/e$ as 0.36806 while the first four decimal places of $1/e$ are 0.36788. This means that the probability that a random permutation (where all permutations are equally likely) is a derangement is equal approximately to $d_n/n! \simeq 1/e$ and this practically gives the same result whether $n = 10$ or $n = 10,000$.

The next few examples require knowledge of some elementary probability.

Example 10.1 *Show that the probability that at least a number is in its natural position in a random permutation (where all permutations are equally likely) is approximately 0.63, and find the order of accuracy of this value.*

Solution

$$\begin{aligned}\text{Prob(at least one match out of } n) &= 1 - \text{Prob(no match)} \\ &= 1 - d_n/n! \\ &\simeq 1 - e^{-1} \\ &\simeq 0.63.\end{aligned}$$

But
$$|e^{-1} - d_n| \leq \frac{1}{(n+1)!}$$
and, for $n \geq 4$, $1/(n+1)! \leq 1/120$. Therefore Prob(at least one match) is 0.63 with an error of less than 1% for $n \geq 4$. □

Example 10.2 *Show that the expected number of matches in a random permutation is 1.*

Solution 1 Prob(i is in its correct position) equals $(n-1)!/n! = 1/n$. Therefore Exp(no of matches of symbol i) equals $1 \times 1/n + 0 \times (1 - 1/n) = 1/n$. Therefore, by linearity of expectation, Exp(no. of matches)=Exp(no. of matches of 1) +...+ Exp(no. of matches of n) $= 1/n + 1/n + \ldots 1/n = 1$.

Solution 2 Prob(exactly r symbols match) equals
$$\frac{\binom{n}{r} \cdot d_{n-r}}{n!}$$
which equals
$$\frac{1 - 1 + \frac{1}{2!} - \frac{1}{3!} + (-1)^{n-r}\frac{1}{(n-r)!}}{r!}$$
which equals $e^{-1}/r!$ for large $n - r$. Therefore expected number of matches equals
$$e^{-1}\sum_{r=0}^{\infty}\frac{r}{r!} = e^{-1} \cdot e = 1.$$
□

10.3 Counting surjections again and a summation formula for the Stirling numbers

We have already seen that the number of surjections from an n-set A to a k-set B is equal to $k!S(n,k)$. This result puts the onus of counting surjections on the ability to find a formula for $S(n,k)$. We shall now reverse our point of view. Let us, for the time being, denote by $T(n,k)$ the number of surjections from A to B. Therefore
$$S(n,k) = \frac{1}{k!}T(n,k).$$

Now, let us concentrate on $T(n,k)$. Let $B = \{b_1, b_2, \ldots, b_k\}$. Let A_1 be all those functions from A to B for which no element of A is mapped onto b_1. Clearly, we have to remove these functions from our count of surjections. Similarly, for $1 \leq i \leq k$, let A_i be the set of all functions from A to B for which no element of A is mapped onto b_i. Clearly, $T(n,k)$ is equal to all functions from A to B (which is equal to k^n) less $|A_1 \cup A_2, \ldots, A_k|$, that is, less those functions which omit some element or elements of B in their range. It is here where we need the inclusion exclusion principle (together with our earlier formula for the total number of functions from one set to another).

Thus, consider A_1. This contains *all* functions which omit b_1 from its range. Therefore we have all functions from an n-set A to a $(k-1)$-set $B - \{b_1\}$. We know that there are $(k-1)^n$ such functions, and this therefore is the value of $|A_1|$. But the same holds for $A_2, \ldots, A_k$, that is, each $|A|$ is equal to $(k-1)^n$.

Now consider $A_1 \cap A_2$. Here we have *all* functions which omit both b_1 and b_2 from their range (they could omit others, for all we care—inclusion-exclusion will take care of that for us). Therefore we have all functions from an n-set A to a $(k-2)$-set $B - \{b_1, b_2\}$. We know that there are $(k-2)^n$ such functions, and this therefore is the value of $|A_1 \cap A_2|$. But the same holds for all intersections taken two at a time, that is, each such intersection has size equal to $(k-2)^n$.

We can now begin to see the pattern. Each intersection $A_i \cup A_j \cup A_l$ taken three at a time has size $(k-3)^n$, and so on for the other intersections.

Therefore the α's in the inclusion-exclusion formula are given by

$$\alpha_i = \binom{k}{i}(k-i)^n,$$

the binomial coefficient being there because there are that many intersections of i sets from k, while the $(k-i)^n$ counts the size of each such intersection.

Therefore, $T(n,k)$ is equal to

$$\begin{aligned} & n^k - |A_1 \cup A_2, \ldots, A_k| \\ = \; & n^k - \alpha_1 + \alpha_2 - \ldots + (-1)^n \alpha_n \\ = \; & n^k - \sum_{i=1}^{k} \binom{k}{i}(k-i)^n \\ = \; & \sum_{i=0}^{k} (-1)^i \binom{k}{i}(k-i)^n \end{aligned}$$

Therefore the Stirling numbers $S(n,k)$ are given by

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

Of course, although this gives us an alternative to what we called "Stirling's Triangle" for calculating Stirling numbers, it is still far from being a closed formula for $S(n,k)$. However, such an expression for the Stirling numbers can help us obtain an approximate closed form representation for $S(n,k)$, and we will briefly consider this now.

The remainder of this section can be omitted.

Let us first write

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

We shall concentrate on obtaining an approximate value for the summation. First, since

$$(k-i)^i < [k]_i < k^i$$

then

$$k^i(1-i/k)^{i+n} < i!B_i < k^i(1-i/k)^n.$$

Now, for $0 < t < 1$, $-\log(1-t) = t + t^2/2 + t^3/3 + \ldots$, therefore therefore

$$t < -\log(1-t) < t/(1-t)$$

therefore

$$[ne^{-(i+n)/(k-i)}]^i < i!B_i < [ne^{-n/k}]^i.$$

Let

$$\lambda = ne^{-n/k}.$$

(This procedure should remind students who have done a course in probability of the Poisson approximation to the binomial distribution). Let $n, k \to \infty$ such that λ remains bounded. For fixed i, the ratio of the left-hand side and right-hand side of the last inequality tends to one as n and k tend to ∞. Therefore

$$0 \le \frac{\lambda^i}{i!} - B_i \to 0.$$

But

$$e^{-\lambda} - \frac{k!S(n,k)}{k^n} = \sum_{i=0}^{\infty}(-1)^i(\frac{\lambda^i}{i!} - A_i)$$

therefore

$$e^{-\lambda} - \frac{k!S(n,k)}{k^n} \to 0$$

as $n, k \to \infty$ in the manner described above. Therefore

$$S(n,k) \simeq \frac{k^n e^{-lambda}}{k!}$$

and, using the approximation $n! = (n/e)^n$ this gives

$$S(n,k) \simeq k^{n-k}\exp[k(1-e^{-n/k})].$$

Use a package like *Mathematica* to investigate how good this approximation is for large n and k.

11 Problem Sheet 2: Inclusion-Exclusion

1. In a class of 67, 47 can read French, 35 can read German and 23 can read both languages. How many can read neither? If, furthermore, 20 can read Russian, of whom 12 read French also, 11 German also and 5 read all languages, how many cannot read any of the three languages.
2. Is there something wrong with this data? In a class of 30 students, 18 study mathematics, 20 study IT and 7 study both mathematics and IT.
3. * In how many ways can the letters A,E,M,O,U,Y be arranged in an ordered sequence such that the words ME and YOU do not occur?
4. Show that if the prime factorisation of n is
$$p_1^{e_1} p_2^{e_2} \dots p_r^{e_r}$$
then the number of divisors of n is
$$(1 + e_1)(1 + e_2) \dots (1 + e_r).$$
5. Show that if m, n are relatively prime then $\phi(mn) = \phi(m)\phi(n)$.
6. * How many integers from 1 to 1000 are divisible by none of 3,7,11?
7. How many ways are there of placing n non-taking rooks on an $n \times n$ chessboard? How many if none lie on the main diagonal? How many if exactly one lies on the main diagonal? How many if exactly k lie on the main diagonal?
8. $*$ Some "Scratch and Win" lotteries are organised as follows. The company running the election prints tickets with two numbers. The first number is a number within the range $\{1, 2, \dots, 100, 000\}$. This number is shown. The second number is chosen randomly within the same range and also printed on the ticket, but this second number is hidden by some silvering which can be scratched off. You win if the number revealed is the same as the first number shown.

   What is the probability of finding a winning ticket? What would this probability be if the two numbers are chosen within the range $\{1, 2, \dots, 10, 000, 000\}$?
9. $*$ How many permutations are there of the digits $1, 2, \dots, 8$ in which none of the patterns 12, 34, 56, 78 appears?

   In how many ways can the letters
$$\alpha, \alpha, \beta, \beta, \beta, \gamma, \delta, \delta, \delta, \delta$$
   be permuted so that all the letters of the same kind are not in a single block?
10. * In how any ways can the numbers $1, 2, 3, \dots, 9$ be permuted such that,
    (a) 1,2,7,9 are not in their natural positions;
    (b) 1,2,7,9 are in their natural positions but none of the others are;
    (c) exactly four integers are in their natural positions;
    (d) at least four are in their natural positions?

Selected answers

1. $8, 6$
2. 582
5. 520
6. $n!$, d_n (the number of derangements of n objects), $n \cdot d_{n-1}$, $\binom{n}{k} d_{n-k}$.
7. 24024, $\frac{10!}{2!3!4!1!} - (\frac{9!}{3!4!} + \frac{8!}{2!4!} + \frac{7!}{2!3!}) + (\frac{7!}{4!} + \frac{5!}{2!} + \frac{6!}{3!}) - 4!$.
8. (a) $9! - \binom{4}{1}8! + \binom{4}{2}7! - \binom{4}{3}6! + \binom{4}{4}5!$; (b) d_5; (c) $\binom{9}{4}d_5$; (d) $\binom{9}{4}d_5 + \binom{9}{5}d_4 + \binom{9}{6}d_3 + \binom{9}{7}d_2 + 1$.

12 Recurrence relations (linear, mainly first & second order)

Consider the following problem: n lines are to be drawn in the plane in what is sometimes called the "general position", that is, no two lines are parallel and no three pass through a common point. Let a_n be the number of sectors in which the plane is divided by these n lines. Find a_n.

Well, we could certainly write down the values of the first few terms of the sequence $\langle a_n \rangle$: $a_0 = 1, a_1 = 2, a_2 = 4$ and $a_3 = 7$. These values do not however point to any obvious pattern or, better still, to a formula for a_n. So we shall proceed this way.

Suppose that n lines have already been drawn, therefore the plane is divided into a_n pieces, and suppose that the $(n + 1)$-st line is to be drawn. This new line L will meet each one of the other lines once. Therefore there will be n new points of intersection $P_1, P_2, \ldots, P_n$. Between each pair of points P_i, P_j a section of the plane (defined by the first n lines) will be divided into two sections. These pairs of points will therefore give $n - 1$ new sections of the plane. But also, the part of the line from P_1 to infinity will divide an existing section into two, as will that part of the line from P_n to infinity, giving a total of $n - 1 + 2 = n + 1$ new sections of the plane. That is, when the $(n + 1)$-st line is drawn, $n + 1$ new sections of the plane will be created, and this can be written as the recurrence relation

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

Now, it takes only a little common sense to find a formula for a_n from this. Thus,

$$\begin{aligned}
a_1 &= a_0 + 1 = 2 \\
a_2 &= a_1 + 2 = 2 + 2 \\
a_3 &= a_2 + 3 = 2 + 2 + 3.
\end{aligned}$$

And one can easily see that, in general,

$$a_n = 1 + 1 + 2 + 3 + \ldots + n = 1 + n(n + 1)/2.$$

What we have seen in this simple example contains most of the ingredients which we shall elaborate upon. We had a first order recurrence relation (a_n depends only on the previous term of the sequence, apart from some functions of n) and we had to have at least one initial condition to get from the recurrence relation to a formula for n. All of this should remind you of two other situations: Proof by Induction and Recursive Programming. As we have said earlier, Recurrence Relations, Proof by Induction and Recursive Programming are all different faces of the same mathematical coin.

The formula we are looking for is what we mean when we say that we are "solving" the recurrence relation. Very often, finding the formula finally boils down to a summation which we either are able to evaluate (as we did in the simple case above) or else it is too difficult to evaluate and we must consider the summation formula as the best we can do towards finding an expression for a_n (just as what we had to do for $S(n, k)$, for example). Here is another example of a simple first order recurrence relation which arises in computer science.

Example 12.1 *In bubble-sort, a linear array of n numbers is given and they are sorted this way: a scan (involving $n-1$ comparisons) is made through the whole array to find the largest number, and this is placed at the end of the array. The procedure is repeated for the remaining $n-1$ numbers in the array, and so on, until the array is sorted in non-decreasing order. How many comparisons are required to sort this way an array of n numbers?*

Solution Let a_n be the required number. After the first $n-1$ comparisons to locate the largest number, we are back to the same problem, but this time on an array of $n-1$ numbers, and this takes a_{n-1} comparisons. Therefore

$$a_n = (n-1) + a_{n-1}$$

with $a_1 = 0$. It is easy to see that

$$\begin{array}{rcl} a_2 & = & a_1 + (2-1) = 1 \\ a_3 & = & a_2 + (3-1) = 1+2 \\ a_4 & = & a_3 + (4-1) = 1+2+3 \\ & & \text{etc} \end{array}$$

and the general solution is $a_n = 1+2+3+\ldots+(n-1) = n(n-1)/2$. □

12.1 Linear first order

Let us now consider first order linear recurrence relations in more systematic detail. As we have just explained, by first order we mean that in the recurrence relation the terms a_n and a_{n-1} appear, that is, each term of the sequence depends on the previous one. To solve completely such a recurrence relation we require also the value of a first term of the sequence, say a_0 or a_1. With just this initial condition and the relation, we can effectively find numerical values of the terms of the sequence (just as we do with Pascal's Triangle or "Stirling's" triangle). But we would like to get more. We are after a formula for the term a_n.

By linear we could say that what we mean is that the recurrence relation does not contain terms like a_n^2 or $a_n a_{n-1}$. A better definition of linear is that if b_n and c_n are both solutions of the recurrence relation, then so is $Ab_n + Bc_n$, where A, B are constants. We shall see later on the power of linearity when we come to second order relations.

The most general form of a linear first order recurrence relation is

$$a_n = f(n)a_{n-1} + g(n)$$

where f and g are functions of n, together with a value for a_0, say. For example, we might have

$$a_n = (n^2+1)a_{n-1} + \log(n) \quad (n \geq 1); \quad a_0 = 7.$$

But first let us start with some simpler examples.

12.1.1 Four easy pieces

Example 12.2 *Solve $a_n = ka_{n-1}$, where k is a constant and $a_0 = A$.*

Solution It is easy to see that $a_1 = ka_0 = kA$, $a_2 = ka_1 = k^2A$ and, in general, the solution is

$$a_n = k^n A.$$

□

Example 12.3 *Solve $a_n = f(n)a_{n-1}$, where $f(n)$ is a function of n and $a_0 = A$.*

Solution Again it is not difficult to see that $a_1 = f(1)a_0 = f(1)A$, $a_2 = f(2)a_1 = f(1)f(2)A$ and, in general, the solution is

$$a_n = f(1)f(2)\dots f(n)A.$$

For example, if $f(n) = n$, then the solution would have been $a_n = n!A$. □

Example 12.4 *Solve $a_n = a_{n-1} + k$, where k is a constant and $a_0 = A$.*

Solution We proceed as above: $a_1 = a_0 + k = A + k$, $a_2 = a_1 + k = A + 2k$ and the general pattern is again easy to see, giving the solution

$$a_n = A + nk.$$

□

Example 12.5 *Solve $a_n = a_{n-1} + g(n)$, where $g(n)$ is a function of n and $a_0 = A$.*

Solution Again we look for the general pattern in the usual way: $a_1 = a_0 + g(1) = A + g(1)$, $a_2 = a_1 + g(2) = A + g(1) + g(2)$ it is easy to see that the solution is

$$a_n = A + g(1) + g(2) + \dots + g(n).$$

For example, if $g(n) = n$, then the solution would have been $a_n = A + n(n+1)/2$.
□

12.1.2 A method for the general case and a few examples

The general first order linear recurrence relation is, as we have said,

$$a_n = f(n)a_{n-1} + g(n) \quad (n \geq 1) \quad a_0 = A.$$

In the previous four examples we considered cases where only f or g was involved. But now we need to consider what to do when both are present. Taking a cue from the second example in the previous section, we define a new sequence $\langle b_n \rangle$ as follows. Let

$$b_n = \frac{a_n}{f(1)f(2)\dots f(n)} \quad (n \geq 1)$$

(we are tacitly assuming that none of the $f(i)$ equals zero) with $b_0 = a_0 = A$. Then substitute into the recurrence relation:

$$f(1)f(2)\dots f(n)b_n = f(n)f(1)f(2)\dots f(n-1)b_{n-1} + g(n).$$

This is the whole point of the substitution: after rearranging the product on the RHS we get $f(1)f(2)\dots f(n)$ which we can factor off, leaving us with a recurrence relation for b_n similar to the last example in the previous section. Thus,

$$\begin{aligned} b_n &= b_{n-1} + \frac{g(n)}{\prod_1^n f(i)} \\ &= b_{n-1} + h(n) \end{aligned}$$

say. Then, as in the last example of the previous section, the solution for b_n is

$$b_n = A + \sum_1^n h(i)$$

and therefore

$$a_n = \prod_1^n f(i)b_n = \prod_1^n f(i)\left(A + \sum_1^n h(i)\right).$$

Of course, our ability to obtain a closed solution for a_n would then depend on whether or not we can evaluate the product of $f(i)$'s and the summation of $h(i)$'s, which does look rather horrid. But the important thing is for you to remember the *method* not the last formula above, and you will see that, often enough, the product and the summation boil down to something which we can easily handle.

Example 12.6 *Solve*

$$a_n = na_{n-1} + 4; \quad a_0 = 7.$$

Solution Let

$$b_n = \frac{a_n}{1 \cdot 2 \cdot \ldots \cdot n} = a_n/n!$$

and $b_0 = 7$. Therefore

$$\begin{aligned} n!b_n &= n!b_{n-1} + 4 \\ b_n &= b_{n-1} + \frac{4}{n!} \\ b_n &= 7 + \sum_{i=1}^n \frac{4}{n!} \\ a_n &= n!\left(7 + \sum_i^n \frac{4}{n!}\right) \end{aligned}$$

and this is the best we can do with this series (we can only sum it if it were an infinite series). □

Example 12.7 *Solve*

$$a_n = 3a_{n-1} + n; \quad a_0 = 1.$$

Solution Let

$$b_n = \frac{a_n}{3 \cdot 3 \ldots \cdot 3} = a_n/3^n$$

and $b_0 = a_0 = 1$. Then substituting in the recurrence relation gives

$$\begin{aligned}
3^n b_n &= 3^n b_{n-1} + n \\
b_n &= b_{n-1} + \frac{n}{3^n} \\
b_n &= b_0 + \sum_{i=1}^{n} \frac{i}{3^i} \\
a_n &= 3^n \left(1 + \sum_{i=1}^{n} \frac{i}{3^i}\right)
\end{aligned}$$

Now how are we going to evaluate this summation, which is a mixture of an arithmetic and a geometric sequence? We proceed this way, starting from the well-known formula for a geometric progression. (This is a technique well worth remembering.)

$$\begin{aligned}
\sum_{i=1}^{n} r^i &= \frac{r(1-r^n)}{1-r} \\
\sum_{i=1}^{n} i r^{i-1} &= \frac{d}{dr}\left(\frac{r(1-r^n)}{1-r}\right) \\
\sum_{i=1}^{n} i r^i &= r\frac{d}{dr}\left(\frac{r(1-r^n)}{1-r}\right)
\end{aligned}$$

which, after differentiation, multiplication by r and substituting $r = 1/3$, which is the common ratio in our case, gives

$$\sum_{i=1}^{n} \frac{i}{3^i} = 1 - (2n+3)(1/3)^n$$

and after substituting into the above expression for a_n this gives

$$a_n = 3^n \left(2 - (2n+3)\frac{1}{3}^{n+1}\right).$$

□

12.2 Second order

We shall now consider second order linear recurrence relations. But to make this easier we shall only allow constant coefficients, and we shall only deal with a few special cases of the right hand side. Thus, we shall be considering the recurrence

$$a_n + ba_{n-1} + ca_{n-2} = f(n), \quad (n \geq 2); a_0 = c_1, a_1 = c_2$$

where a, b, c are constants, c_1, c_2 are the given initial values of a_0, a_1, respectively, and where the function $f(n)$ of n will take on a few particular forms which we shall study below.

One important case of this recurrence relation is when $f(n) \equiv 0$. This is called the homogeneous case, and it is with this that we shall start our investigations.

12.2.1 Homogeneous case: two mysterious substitutions

We shall start with a concrete example. We shall effect two substitutions which will transform the second order recurrence into two first order ones. At this stage, these substitutions might seem quite mysterious—they seem to have been plucked out of thin air. In this section concentrate only on checking that these substitutions really do what is required of them—that is, just check that the working is ok. In the next section you will be told how to find these two substitutions, and it will turn out to be very easy indeed!

Example 12.8 *Solve*

$$a_n - 5a_{n-1} + 6a_{n-2} = 0; \quad a_0 = 5, a_1 = 22$$

Solution Let

$$c_n = a_{n-1} - 3a_{n-2}, (n \geq 2).$$

Therefore, from a_0 and a_1, $c_2 = 7$. This is the first of our "mysterious" substitutions. Here is the second one. Consider

$$c_{n+1} - 2c_n.$$

In terms of a_n, this is equal to,

$$a_n - 3a_{n-1} - 2a_{n-1} + 6a_{n-2}$$

which equals

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

which equals 0. Therefore

$$\begin{aligned} c_{n+1} - 2c_n &= 0 \\ c_{n+1} &= 2c_n. \end{aligned}$$

Therefore $c_3 = 2c_2 = 2 \cdot 7$, $c_4 = 2^2 \cdot 7$, $c_5 = 2^3 \cdot 7$ and, in general, $c_n = 7 \cdot 2^{n-2} = \frac{7}{4} \cdot 2^n$.

Having solved our first of our pair of first order recurrence relation we can now substitute in $c_n = a_{n-1} - 3_{n-2}$ to solve for a_n. Thus,

$$a_{n-1} = 3a_{n-2} + \frac{7}{4} \cdot 2^n$$

or

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

As per our usual "recipe" for solving such recurrences, let

$$b_n = a_n/3^n.$$

Therefore

$$\begin{aligned}
3^n b_n &= 3^n b_{n-1} + \frac{7}{2}2^n \\
b_n &= b_{n-1} + \frac{5}{2} \cdot \frac{2^n}{3^n} \\
b_n &= b_0 + 7/2 \sum_1^n (2/3)^n \\
&= 5 + (7/2)\frac{(2/3)(1-(2/3)^n)}{1-(2/3)} \\
&= 5 + 7(1-(2/3)^n).
\end{aligned}$$

Therefore

$$a_n = 3^n b_n = 12 \cdot 3^n - 7 \cdot 2^n.$$

□

12.2.2 All is made clear: the auxiliary equation!

But how did we manage to guess the two substitutions for c_n which made the last one equal to 0 starting us off with a first order recurrence relation? The little thing which helped us is a quadratic equation called the *auxiliary equation*. Thus, corresponding to the recurrence relation

$$a_n - 5a_{n-1} + 6a_{n-2} = 0$$

the auxiliary equation is

$$k^2 - 5k + 6 = 0.$$

This factorises into

$$(k-2)(k-3)$$

that is, has roots 2 and 3. When we see an auxiliary equation of this type we know that the substitution

$$c_n = a_{n-1} - 3a_{n-2}$$

will lead to

$$c_{n+1} - 2c_n = 0.$$

(The other way round is just as valid: $c_n = a_{n-1} - 2a_{n-2}$ would lead to $c_{n+1} - 3c_n = 0$—try it out!)

If you study carefully the previous section, you will see the important part which the roots 3 and 2 played, and that the solutions to such a pair of first order recurrence relations will ultimately be of the form

$$A2^n + B3^n$$

so much so that, after finding the roots of the auxiliary equation we can immediately write out this answer without actually working out the two first order recurrences. Let us repeat the above example working this way.

Example 12.9 *Solve*

$$a_n - 5a_{n-1} + 6a_{n-2} = 0; \quad a_0 = 5, a_1 = 22$$

Solution The auxiliary equation is

$$k^2 - 5k + 6 = 0$$

which has roots 3 and 2. Therefore the solution is

$$a_n = A2^n + B3^n$$

where A, B are arbitrary constants which need to be determined. We determine them this way. Since $a_0 = 5$ and $a_1 = 22$, this gives two simultaneous equations

$$\begin{aligned} 5 &= A + B \\ 22 &= 2A + 3B \end{aligned}$$

which easily gives $A = -7$ and $B = 12$, as above. □

Therefore the whole thing boils down to solving a simple quadratic equation. There is occasion in which the above answer would vary, and that occurs if the auxiliary equation has equal roots.

Example 12.10 *Solve*

$$a_n - 4a_{n-1} + 4a_{n-2} = 0; \quad a_0 = a_1 = 1$$

Solution The auxiliary equation is

$$k^2 - 4k - 4 = 0$$

which has repeated roots 2,2. The solution cannot be $A2^n + B2^n$ since this would have only one arbitrary constant which can be determined just knowing a_0, and we know that a second order recurrence relation needs two initial values to start the solution off. We must again repeat, at least for the first time until we get the pattern right, the long method involving two first order recurrence relations. Thus, let

$$c_n = a_{n-1} - 2a_{n-2}, (n \geq 2)$$

and consider

$$c_{n+1} - 2c_n$$

which, when written out in terms of a_n, 0. Solving these resulting first order recurrences is left to the reader, but the general form of the solution will be

$$a_n = (A + Bn)2^n.$$

getting this via the two first order equations will also give you A and B. In our case, we can find them as usual:

$$\begin{aligned} a_0 = 1 \Rightarrow 1 &= A \\ a_1 = 1 \Rightarrow 1 &= 2A + 2B \end{aligned}$$

which immediately gives $A = 1$ and $B = -1/2$. □

So we can write down the general rule for solving second order homogeneous recurrence relations with constant coefficients: *Find the roots α, β of the auxiliary equation. If $\alpha \neq \beta$ then the solution is*

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

whereas if $\alpha = \beta$, then the solution is

$$a_n = (A + nB)\alpha^n.$$

This rule, appropriately extended, also works for higher order.

Example 12.11 *Solve*

$$4a_n - 20a_{n-1} + 17a_{n-2} - 4a_{n-3} = 0.$$

Solution The auxiliary equation is

$$4k^3 - 20k^2 + 17k - 4 = 0$$

which has roots $k = 1/2, 1/2, 4$. Therefore the general solution is

$$a_n = (A + nB)\left(\frac{1}{2}\right)^n + C4^n$$

where the arbitrary constants A, B, C can be found from three initial values a_0, a_1, a_2. □

12.2.3 The auxiliary equation via matrices

This section can be omitted without loss of understanding of the rest of the course. It requires knowledge of some linear algebra. We shall gloss over these linear algebra details somewhat rapidly.

Consider the 2nd order recurrence relation

$$a_n + ba_{n-1} + ca_{n-2} = 0.$$

This can be represented by the matrix equation

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}.$$

In other words, to find a formula for a_n in terms of the initial conditions a_0, a_1 we need a formula for

$$\begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}^n.$$

Now, there are standard techniques in linear algebra for finding nth powers of matrices. Let us first find the eigenvalues of the matrix. These are the solutions of the equation,

$$\det \begin{bmatrix} -b-\lambda & -c \\ 1 & -\lambda \end{bmatrix} = 0,$$

that is,

$$\lambda^2 + b\lambda + c = 0,$$

which is just the auxiliary equation!

As usual, suppose first that the auxiliary equation has distinct roots α, β. Then it follows, again from linear algebra, that the matrix has a complete set of eigenvectors, and that if P is the matrix whose columns are linearly independent vectors, then the matrix is similar to a diagonal matrix, that is,

$$\begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix} = P \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} P^{-1}.$$

Therefore

$$\begin{aligned} \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}^n &= P \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}^n P^{-1} \\ &= P \begin{bmatrix} \alpha^n & 0 \\ 0 & \beta^n \end{bmatrix} P^{-1} \end{aligned}$$

and hence

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = P \begin{bmatrix} \alpha^n & 0 \\ 0 & \beta^n \end{bmatrix} P^{-1} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$$

which gives

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

where A and B are appropriate constants. Rather than finding the matrix P to obtain these constants, we can now use the initial values of a_1 and a_0, as before.

Now suppose that the roots α, β of the auxiliary equation are equal—denote them by α. The matrix is now not similar to a diagonal form, but it almost is. In fact, one can find a matrix P such that

$$\begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix} = P \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix} P^{-1}.$$

Therefore

$$\begin{aligned} \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}^n &= P \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}^n P^{-1} \\ &= P \begin{bmatrix} \alpha^n & 0 \\ 0 & n\alpha^{n-1}\alpha^n \end{bmatrix} P^{-1} \end{aligned}$$

and hence

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = P \begin{bmatrix} \alpha^n & n\alpha^{n-1} \\ 0 & \alpha^n \end{bmatrix} P^{-1} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$$

which gives

$$a_n = (A + nB)\alpha^n$$

where A and B are appropriate constants. Again, rather than finding the matrix P to obtain these constants, we use the initial values of a_1 and a_0, as before.

12.2.4 Nonhomogeneous case: the power of linearity

Well, now that we know how to solve the homogeneous case, the next step is certainly to learn how to solve the case where the RHS of the recurrence relation is not 0. Specifically we shall be concerned with the recurrence relation

$$a_n + ba_{n-1} + ca_{n-2} = f(n)$$

where $f(n)$ will be allowed to take on some special forms. These details will be considered in the next section. Here we shall make some observations which apply for any $f(n)$.

First of all, suppose we have solved, via the auxiliary equation, this recurrence relation with $f(n) = 0$. Suppose that our solution still has its arbitrary constants to be determined (for example, $A2^n + B3^n$ or $(A+nB)7^n$). Therefore what we have is really not just one solution but a whole family of solutions: thus *any* solution of

$$a_n + ba_{n-1} + ca_{n-2} = 0$$

is contained in this family by appropriately choosing A and B. Let us call this solution S_n, where we have used a capital S in order to remind us that this is not just one solution but a family which contains all solutions.

Now, suppose that *the* solution of

$$a_n + ba_{n-1} + ca_{n-2} = f(n)$$

satisfying the initial conditions is found (we shall soon see how to do that); let us call it $s(n)$. Suppose also that we can somehow "guess" a solution of

$$a_n + ba_{n-1} + ca_{n-2} = f(n)$$

which does not necessarily satisfy the initial conditions (we shall also see how to do this in the next section). Let us denote this "guessed" solution by p_n—all this means is that if p_n is substituted in the LHS of

$$a_n + ba_{n-1} + ca_{n-2} = f(n)$$

then the result will be $f(n)$. This "guessed" solution is called the *particular solution.*

Now consider the sequence $s_n - p_n$. By linearity, if we substitute this sequence in the LHS of the recurrence relation, the result would be the same as if we had substituted s_n and then p_n and then subtracted, that is

$$s_n + bs_{n-1} + cs_{n-2} - p_n + bp_{n-1} + cp_{n-2}.$$

But this would equal zero, since both s_n and p_n give the same result, $f(n)$, when plugged into the LHS. Therefore $s_n - p_n$ is a solution of the homogeneous recurrence, that is, it is of the form S_n, for appropriate arbitrary constants. That is,

$$\begin{aligned} s_n - p_n &= S_n \\ s_n &= S_n + p_n \end{aligned}$$

We therefore have a very simple way of solving these nonhomogeneous recurrence relation.

1. Solve with the RHS equal to zero, leaving the arbitrary constants undetermined.
2. "Guess" or somehow find a solution valid for the given RHS (the particular solution).
3. Add these two solutions together.
4. Finally determine the arbitrary constants.

In the next section we shall see how to determine the particular solution for a few types of RHS.

12.2.5 Nonhomogeneous case: particular solutions for some special right-hand-sides

We shall proceed by examples.

Example 12.12 *Solve the recurrence relation*

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

given that $a_0 = a_1 = 0$.

Solution The auxiliary equation has roots 2 and 3, therefore the homogeneous solution is

$$a_n = A2^n + B3^n.$$

But now we need to find a particular solution to add to the above. Let us try $a_n = n$, for want of a better guess at this stage. Substituting in the LHS gives $2n - 7$, and this is clearly not correct (we want the result to be n, like the RHS). Ok, so let us try $a_n = n/2$—at least this should get the coefficient of n right. In fact, when we substitute into the LHS we obtain $n - 7/2$. We must therefore get rid of the constant term. Let us try $a_n = n/2 + 7/4$, (since, on substituting in the LHS, the constant term is multiplied by 2). Substituting in the LHS now does give n (check this!), which is precisely what we want. So $a_n = n/2 + 7/4$ is a particular solution, and we got there after three trials.

However, with hindsight, it is now easy to realise that the solution must necessarily have been of the form $Hn + K$, where H and K are constants. So, instead of the above three trials, let us find the particular solution at one go by substituting $a_n = Hn + K$ into the LHS and finding H, K by equating to the RHS. Substitution gives

$$Hn + K - 5(H(n-1) + K) + 6(H(n-2) + K)$$

which, on simplifying gives

$$2Hn + 2K - 7H.$$

Equating with the RHS and comparing coefficients gives that $2H = 1$ and $2K - 7H = 0$, giving, as above, $H = 1/2$ and $K = 7/4$.

Therefore the complete solution is now,

$$a_n = A2^n + B3^n + \frac{n}{2} + \frac{7}{2}.$$

Only at this point do we substitute $a_0 = a_1 = 0$ in order to find the two arbitrary constants A and B. This is here left as a simple exercise for the reader. □

Example 12.13 *Solve the recurrence relation*

$$a_n - 5a_{n-1} + 6a_{n-2} = 4^n$$

given that $a_0 = a_1 = 0$.

Solution The homogeneous solution is exactly like the previous example. So all we have to do is find a particular solution and add it to the homogeneous solution, as we did above.

In the same vein as our first attempts in the previous example, let us try $a_n = 4^n$. Substituting in the LHS gives

$$\begin{aligned} & 4^n - 5 \cdot 4^{n-1} + 6 \cdot 4^{n-2} \\ = \; & 4^n(1 - 5/4 + 6/16) \\ = \; & \frac{1}{8}4^n. \end{aligned}$$

But our RHS is 4^n, therefore let us try $a_n = 8 \cdot 4^n$ (hoping that this will cancel out the $\frac{1}{8}$). Substituting in the LHS gives us the right answer, 4^n, so a good particular solution. Again, using hindsight, we now see that we could have guessed that the result would be of the type constant times 4^n. So, instead of making a number of trials, we could simply substitute $a_n = H4^n$ in the LHS, and after simplification compare coefficients with the RHS to obtain H. Thus,

$$\begin{aligned} H4^n - 5H4^{n-1} + 6H4^{n-2} &= \\ H4^n(1 - 5/4 + 6/16) &= \\ H4^n(1/8) &= \text{RHS} \\ &= 4^n \end{aligned}$$

giving, as above, that $H = 8$.

Now, adding to the homogeneous solution we obtain

$$a_n = A2^n + B3^n + 8 \cdot 4^n$$

which is the complete solution. All we have to do now is use $a_0 = a_1 = 0$ to obtain the values of the arbitrary constants a and B—this is left as an exercise for the reader. □

Example 12.14 *Solve the recurrence relation*

$$a_n - 5a_{n-1} + 6a_{n-2} = 3^n$$

given that $a_0 = a_1 = 0$.

Solution The homogeneous solution is

$$A2^n + B3^n,$$

as in the previous two examples. Now, continuing in the same vein as in the previous example, one would try substituting $a_n = H3^n$ into the LHS in order to find the particular solution. But if we do this now, the LHS will become 0, and therefore no value of H can make the LHS equal to the RHS. But if we look carefully at the homogeneous solution we note that this was bound to happen, because constant times 3^n is already a part of the solution there, and this means that substituting constant times 3^n into the LHS must give us zero.

Therefore we are alerted to this problem as soon as we see the roots of the auxiliary equation and we note that the "3" in 3^n of the RHS is already a root. So what should we do? The substitution here should be $a_n = Hn3^n$. Let us try putting this into the LHS.

$$\begin{aligned} Hn3^n - 5H(n-1)3^{n-1} + 6H(n-2)3^{n-2} &= \\ Hn3^n(1 - 5/3 + 6/9) + H3^n(1 + 5/3 - 12/9) &= \\ 0 + H3^n(4/3) &= RHS \\ &= 3^n. \end{aligned}$$

(Note how the 0 which was causing us trouble has now been diverted to the term containing n, which is precisely what we want, since the RHS does not contain terms like $n3^n$.)

Therefore $H = 3/4$, giving as particular solution $\frac{3}{4}n3^n$. This can now be added to the homogeneous solution and the arbitrary constants determined as usual. □

Example 12.15 *Solve the recurrence relation*

$$a_n - 6a_{n-1} + 9a_{n-2} = 3^n$$

given that $a_0 = a_1 = 0$.

Solution The auxiliary equation here has a double root at 3. Therefore the homogeneous solution is

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

We now need to find a particular solution. As before, we see that the attempt $a_n = H3^n$ does not work. But even $a_n = Hn3^n$ collapses the LHS to zero—and we should not be surprised really, because 3 being a double root, $n3^n$ is already in the homogeneous solution. Therefore we attempt $a_n = Hn^2 3^n$. If we substitute in the LHS we see that the terms in n^2 and those in n become zero, leaving us with a non-zero term containing $H3^n$ which can be compared to the RHS to give H. The details are left to the reader. The general rule is therefore that if α is a root of the auxiliary equation with multiplicity m, to find a particular solution we should try $a_n = Hn^m\alpha^n$. □

In the next problem sheet you will be asked to find the particular solution for combinations of these different RHS's.

12.2.6 The use of generating functions

This section may be omitted.

At this point the reader would be justified in thinking that he or she is being asked to take too many things for granted. Yes we have justified the use of the auxiliary equation for order 2, but what about higher orders? We are told that we can use the auxiliary equation without any real justification. And what about our way of getting a particular solution? For some situations (as in the last example) we did not give much justification (although one can always verify that the particular solution is correct by substituting in the LHS). Moreover, we have considered only a very restricted type of RHS. What about other types? And what if the recurrence relation does not have constant coefficients?

Well, one way in which all this can be put on a surer mathematical footing is by means of generating functions. In this course we shall not be dealing too much with this method of solving recurrence relations, but, for completeness' sake, and also as another illustration of the power and versatility of generating functions, we shall redo one of the above examples by this method.

Example 12.16 *Solve the recurrence relation*

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

given that $a_0 = a_1 = 0$.

Solution We are looking for a sequence $a_0, a_1, a_2, \ldots$ which satisfies the above recurrence relation, including the initial conditions. Let us form the generating function for this sequence:

$$g(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_j x^j + \ldots$$

Using the recurrence relation we shall obtain an algebraic expression for $g(x)$, and then we shall expand $g(x)$ in order to get the coefficient of x^n, which is a_n, precisely what we want.

So, let us multiply the recurrence relation by x^n and let us sum the LHS and RHS from $n = 2$ (why?) to infinity.

$$\begin{aligned}
\sum_{n=2} a_n x^n - 5\sum_{n=2} a_{n-1}x^n + 6\sum_{n=2} a_{n-2}x^n &= \sum_{n=2} nx^n \\
g(x) - a_0 - a_1 - 5x\sum_{n=2} a_{n-1}x^{n-1} + 6x^2\sum_{n=0} a_n x^n &= x\sum_{n=1}(n+1)x^n \\
g(x) - 5x(g(x) - a_0) + 6x^2 g(x) &= x((1-x)^{-2} - 1)) \\
g(x) &= \frac{2x^2 - x^3}{(1-x)^2(1-3x)(1-2x)}.
\end{aligned}$$

Expanding the last expression in partial fractions gives

$$g(x) = \frac{4/5}{1-x} + \frac{1/2}{(1-x)^2} + \frac{5/4}{1-3x} + \frac{-3}{1-2x}.$$

Therefore the coefficient of x^n in $g(x)$ is

$$4/5 + (n+1)/2 + 3^n(5/4) - 2^n \cdot 3$$

which is therefore a_n. $\square$

Of course, this method is much longer than the ones we have been studying so far, but this is only because our intention was to give an illustration of the use of generating functions in a simple situation whose solution can be verified by simpler techniques. Keep in mind, however, that there are many situations in which our "short" methods do not apply (non-linear recurrences, non-constant coefficients, more complex right-hand-sides, etc) and it is the powerful and versatile method of generating functions which has to be used. The next problem sheet presents one such example.

13 Problem Sheet 3: Recurrence Relations

1. * Solve the following recurrence relations:
   (a) $a_{n+2} - 3a_{n+1} - 4a_n = 0$ $(n \geq 0)$; $a_0 = 1, a_1 = 3$.
   (b) $a_{n+3} - 6a_{n+2} + 11a_{n+1} - 6a_n = 0$ $(n \geq 0)$; $a_0 = 2, a_1 = 0, a_2 = -2$.
   (c) $a_{n+3} - 3a_{n+1} + 2a_n = 0$ $(n \geq 0)$; $a_0 = 1, a_1 = 0, a_2 = 0$.
2. *
   (a) A man climbs stairs in the following fashion. Sometimes he takes two stairs in one stride, sometimes only one. Find a formula for a_n, the number of different ways he can climb n stairs.
   (b) Find b_n, the number of n-digit binary words containing no two consecutive zeroes.
3. * Let q_n be the number of words of length n in the alphabet $\{a, b, c, d\}$ which contain an odd number of b's. Show that
   $$q_{n+1} = 4^n + 2q_n, (n \geq 1)$$
   and hence find q_n. [*Hint:* Divide the set of such words of length $n+1$ into those which begin with b and those which do not.]
4. * Find a particular solution for each of the following recurrence relations:
   (a) $a_n + 5a_{n-1} + 6a_{n-2} = 2n^2 + n - 5$.
   (b) $a_n - 5a_{n-1} + 6a_{n-2} = 3$.
   (c) $a_n + a_{n-1} = 4n3^n$.
   (d) $a_n - 2a_{n-1} = 5 \cdot 2^n$.
   (e) $a_n - 6a_{n-1} + 9a_{n-2} = (n-1)3^n$.
   (f) $a_n = a_{n-1} + 5$.
   (g) $a_n - 2a_{n-1} + a_{n-2} = 5$.
   (h) $a_n - 7a_{n-1} + 12a_{n-2} = 5$.
   (i) $a_n - 5a_{n-1} + 6a_{n-2} = 3^n + n$.
   (j) $a_n - 2a_{n-1} + a_{n-2} = n$.
5. Let the sequence $\langle a_n \rangle$ be defined by
   $$a_{n+2} + a_{n+1} + na_n = 0, \quad (n \geq 0)$$
   with $a_0 = 1; a_1 = 0$.

Let g be the generating function of the sequence. Show that

$$g'(x) = -\frac{g(x)}{x^3 + x^2}.$$

[*Hint:* Multiply the recurrence relation by x^{n-1} and take summations.]

Can you find an expression for $g(x)$ and, from it, the term a_n?

6. *The remaining set of questions up to Question 10 are all illustrations of the use of recurrence relations to problems in computer science. You might have already met them or some variants, particularly in courses on data structures and algorithms.*

   A *full binary tree* T is defined to be a rooted tree such that every vertex is either a leaf or else has both a left and a right child. Show that, if T has l leaves and i internal vertices, then $l = i + 1$.

   Now assume that all the leaves are at level k (the level of the root is taken to be zero). How many leaves does T have? Deduce that the total number of vertices is $2^{k+1} - 1$.

   Now let u_k be the sum of the levels of all the vertices of T. By considering $T - r$, that is, T with the root vertex r deleted, show that

   $$u_k = 2u_{k-1} + 2^{k+1} - 2.$$

   Hence deduce that $u_k = (k-1)2^{k+1} + 2$.

   What does this result say about the average number of comparisons required to find an element in a sorted array using binary search?

7. Let c_n be the average number of comparisons required to *quicksort* n items in an array (assuming all possible initial orderings are equally likely). It can be shown that

   $$c_n = (n-1) + \frac{2}{n}\sum_{i=0}^{n-1} c_i, \;\; (c_0 = 0).$$

   CS students should know this or at least should know where to look up the proof of the result.

   Show that

   $$c_n = (n+1)H(n) + 4 - 4(n+1)$$

   where $H(n) = 1 + \frac{1}{2} + \frac{1}{3} + \ldots \frac{1}{n}$ (called the *harmonic function*), and hence that

   $$c_n \simeq 1.39n \log_2 n.$$

8. Consider the following algorithm for sorting n numbers, for $n \geq 2$.

   (a) Use $2n - 3$ comparisons to determine the largest and second largest of the n numbers;

   (b) recursively sort the remaining $n - 2$ numbers.

   Let a_n denote the number of comparisons used for sorting n numbers in this fashion. Find a_n.

9. Let S be the set of $n = 2^k$ distinct integers ($k \geq 1$). Show that it is possible to find the maximum *and* the minimum of S using $3n/2 - 2$ comparisons.

10. Consider a "divide-and-conquer" recursive algorithm which takes $T(n)$ computational steps when running on a problem whose input size is n. The algorithm proceeds by dividing the problem into two subproblems each of size $n/2$ (assume n is a power of 2) and so on recursively.

    Suppose that the number of steps to carry out this division is $c_1 n$ for some constant c_1 and that $T(1) = c_2$ for some constant c_2. Show that

    $$T(n) \leq (c_1 + c_2)n \log_2 n + c_2.$$

11. *The last two questions are taken from probability theory. These and the previous problems should convince the student that recurrence relations crop up in many areas of mathematics*

    Two gamblers A and B play a game against each other. They repeatedly flip a coin which comes up heads with probability p and tails with probability $q = 1-p$. Gambler A starts with Lm n and ganbler B starts with LM $(N-n)$. If tails comes up, then A wins Lm 1 from B, whereas if heads comes up, then B wins Lm 1 from A. The game proceeds until one gambler has Lm 0. Let a_n be the probability that A wins starting with Lm n; clearly $a_0 = 0$ and $a_N = 1$. Show that

    $$a_n = pa_{n+1} + qa_{n-1},$$

    and solve the recurrence relation to find a_n.

12. A coin is tossed repeatedly. Each time there is a probability p of a head turning up. Let $_n$ be the probability that an even number of heads has occurred after n tosses. Show that

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

    Hence show that $a_n = \frac{1}{2}[1 + (1-2p)^n]$.

Selected answers

1. (a) $\frac{4^{n+1}+(-1)^n}{5}$; (b) $5 - 2^{n+2} + 3^n$; (c) $\frac{8-6n(-2)^n}{9}$.
2. Recurrence relations for the two quantities:
(a) $a_n = a_{n-1} + a_{n-2}, n \geq 2;\ a_0 = a_1 = 1$.
(b) $b_n = b_{n-1} + b_{n-2}, n \geq 3;\ b_1 = 2, b_2 = 3$.
3. $q_n = \frac{1}{2}(4^n - 2^n)$.
4. (a) Roots $=-3, -2$; Trial soln: $a_n = Hn^2 + Jn + K$.
(b) 3,2; $a_n = H$.
(c) -1; $a_n = (Hn + J)3^n$.
(d) 2; $a_n = Hn2^n$.
(e) 3,3; $a_n = (Hn + J)n^2 3^n$.
(f) 1; $a_n = Hn(1)^n = Hn$.
(g) 1,1; $a_n = Hn^2(1)^n = Hn^2$.
(h) 3,4; $a_n = H$.
(i) 3,2; $a_n = Hn3^n + Jn + K$.
(j) 1,1; $a_n = (Hn + J)n^2 1^n = (Hn + J)n^2$.
9. Biggs, worked example p.253.
10. Aho, Hopcroft & Ullman worked example 9.1 p.296.

14 Partitions of a positive integer: a brief introductory excursion

In this final section we shall again be looking at a problem from number theory. We shall only be scratching the surface of this topic, so what is presented here is only a small part of what would be covered in this field of study if this were a course in number theory. However, we want to end these lecture notes with this topic because, even with this cursory treatment, it is possible to see some non-trivial applications of things like generating functions, and this should not only help to consolidate what you have learned in the previous sections but it should help convince you that discrete mathematics is not an isolated subject but it is part of mainstream mathematics with applications to various other fields. Moreover, you will be able to see almost side by side the two principal methods of counting in combinatorics: counting by means of *bijections* and counting using *generating functions*.

Let us first define what a partition of an integer means, and also introduce some notation. In the first problem sheet you were asked to find in how many ways can the number 6 be written as a sum of positive integers, order not significant and repetition allowed. The number 6 is small enough that we can solve this without knowing any number theory. Thus,

$$\begin{aligned}
6 &= 6\\
6 &= 1+5\\
6 &= 2+4\\
6 &= 1+1+4\\
6 &= 3+3\\
6 &= 1+2+3\\
6 &= 1+1+1+3\\
6 &= 2+2+2\\
6 &= 1+1+2+2\\
6 &= 1+1+1+1+2\\
6 &= 1+1+1+1+1+1.
\end{aligned}$$

So there are 11 partitions of 6. We write this as $p(6) = 11$ and, in general, the number of partitions of the positive integer n is denoted by $p(n)$. The terms in the summation are called the *parts* of the partition. Sometimes we require partitions satisfying certain conditions. Thus let $\mathcal{P}$ be a property such as 'the number of parts cannot exceed five' or 'each part is even'. We denote the number of partitions of n with such a restriction by the notation $p(n|\mathcal{P})$.

Finding the number of partitions of 6 was easy. But this ease is very deceptive for numbers larger than 6. To give you an idea, note that $p(20) = 627, p(100) = 190,569,292$ and $p(200) = 3,972,999,029,388$. Clearly a more mathematical way of tackling this problem is required. This is what we shall attempt to do in the next two sections.

Example 14.1 *You can now look at the last of the four motivating problems we stated at the beginning of the course. What general problem is it a special case of?*

14.1 Ferrers Diagrams

The technique which we shall now present is so simple that it seems quite surprising that any nontrivial mathematics can be done with it. But this often happens in many areas of mathematics, not the least in combinatorics.

To every partition we shall associate what is called a Ferrers diagram of the partition. which is a diagram made up of dots or multiplication signs. The best way to explain this is by some examples. The partition $10 = 2 + 3 + 5$ has the Ferrers diagram

$$\begin{array}{ccccc}
\times & \times & \times & \times & \times\\
\times & \times & \times & & \\
\times & \times & & &
\end{array}$$

while the partition $7 = 4 + 1 + 1 + 1$ has

$$\begin{array}{cccc} \times & \times & \times & \times \\ \times & & & \\ \times & & & \\ \times & & & \end{array}$$

as its Ferrers diagram.

Note that the rows of a Ferrers diagram are written in nonincreasing length.

For any diagram D there is a corresponding diagram which is called the *conjugate* of D, and is denoted by $\mathrm{conj}(D)$, and whose ith column corresponds to the ith row of D (this is similar to taking the transpose of a matrix). For example, $\mathrm{conj}(D)$ for each of the above two diagrams is, respectively,

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \\ \times & & \\ \times & & \end{array}$$

and

$$\begin{array}{cccc} \times & \times & \times & \times \\ \times & & & \\ \times & & & \\ \times & & & \end{array}.$$

But can we do any mathematics with such a simple idea as the Ferrers Diagram? Let us try.

Theorem 14.1 *Let $\mathcal{P}_1$ be the property that the largest part of a partition is m and let $\mathcal{P}_2$ be the property that the partition has exactly m parts. Then,*

$$p(n|\mathcal{P}_1) = p(n|\mathcal{P}_2).$$

Proof Let A be the set of Ferrers diagrams corresponding to $p(n|\mathcal{P}_1)$ and let B be the set of Ferrers diagrams corresponding to $p(n|\mathcal{P}_2)$. We need to show that $|A| = |B|$, and, in order to show that two sets have the same size, one common technique is to find a bijection between them. Thus, define the function *conj* from A to B which maps any diagram D in A into its conjugate, $\mathrm{conj}(D)$. This is clearly a bijection. (Details are left to the reader.) Therefore $|A| = |B|$, as required. $\square$

Note a few things about that proof. First of all, as soon as we spotted what function to define it became easy to see that it is a bijection. This is true in many of the elementary cases you might be meeting, but not true in general. Sometimes very ingenious arguments are required to show that the function is a bijection. Secondly, what we have seen is what is called, in combinatorics, a proof by bijection, and it is a very important counting technique. One of its advantages is that not only does it show you that two sets are equal but it also shows you which elements of the two sets correspond to each other—it tells you *why* the two sets are equal. In the next section we shall see another method of counting which does not have this advantage but it has others instead.

Finally, note that we have no way of knowing what $p(n|\mathcal{P}_1)$ or $p(n|\mathcal{P}_2)$ is, yet we know that they are equal. Most of our results in this section will be of this type: It is too difficult to find $p(n)$ or variations of it, but we still can state nontrivial relationships between different types of partitions.

Theorem 14.2 *Let $\mathcal{P}_1$ be the property that the number of parts in the partition is at most r, and let $\mathcal{P}_2$ be the property that the partition has exactly r parts. Then*

$$p(n|\mathcal{P}_1) = p(n + r|\mathcal{P}_2).$$

[*For example*, the number of partitions of 12 in which there are at most 5 parts is equal to the number of partitions of 17 with exactly 5 parts.]

Proof Let A be the set of Ferrers diagrams corresponding to $p(n|\mathcal{P}_1$ and let B be the set of Ferrers diagrams corresponding to $p(n|\mathcal{P}_2$. We need to show that $|A| = |B|$, and again, in order to do this we shall find a bijection between the two sets. Thus, define the function $f : B \to A$ as follows: $f(D)$ is obtained from D by removing its first column. It is easy to see that a diagram in B is really sent to a diagram in A (remember what shape Ferrers diagrams must have), therefore f is a function from B to A. It is just as easy to show that f is a bijection. $\square$

14.2 Use of generating functions

When finding partitions of n, what we are actually doing is deciding how many 1's, if any, the partition will have, then how many 2's, how many 3's, etc. It is as if we have boxes with a limitless supply of 1's, 2's, 3's, etc to choose from. The fact that the supplies are limitless does not mean that we can take an infinity of 1's, say, because our sum cannot exceed the value of n. But it helps to imagine limitless supplies because this way we need not change the boxes if we are asked to find the partition of another number different from n—the same arrangement will suffice.

Let us think of this algebraically. Suppose we represent the first box by the barcket

$$(1 + x + x^2 + x^3 + \ldots)$$

the second by the bracket

$$(1 + x^2 + x^4 + x^6 + \ldots)$$

and so on. Then, the value of $p(n)$ equals the coefficient of x^n in the product

$$(1 + x + x^2 + x^3 + \ldots)(1 + x^2 + x^4 + x^6 + \ldots)(1 + x^3 + x^6 + x^9 \ldots) \ldots \ .$$

Note again that just because we have a potentially infinite number of brackets it does not mean that we can use all of them, since the sum of the powers of the terms we choose must equal to n.

In order to make this a bit clearer, think of the partition of 22 given by

$$1 + 1 + 1 + 3 + 4 + 4 + 8.$$

In our algebraic formulation this means that we have chosen to multiply x^3 from the first bracket (the three 1's) by x^3 from the third bracket (the single 3) by x^8 from the fourth bracket (the two 4's) and by x^8 from the eight bracket (the single 8). Therefore an x raised to a given power will come from different brackets depending on how many 1's or 2's or 3's, etc, we have chosen from the respective 'boxes'.

What we have just seen is that the generating function of $p(n)$ is given by

$$(1+x+x^2+x^3+\ldots)(1+x^2+x^4+x^6+\ldots)(1+x^3+x^6+x^9\ldots)\ldots$$

whose brackets contain infinite geometric progressions, and which therefore can be written as

$$\frac{1}{(1-x)(1-x^2)(1-x^3)(1-x^4)\ldots}.$$

In other words

$$p(n)=[x^n]\frac{1}{(1-x)(1-x^2)(1-x^3)(1-x^4)\ldots}.$$

What about partitions with restrictions? Can we find generating functions for them? Here are a few.

- Let $\mathcal{P}$ be the property that the value of the parts can only be equal to $m_1, m_2, \ldots, m_k$. Then

$$p(n)=[x^n]\frac{1}{(1-x^{m_1})(1-x^{m_2})\ldots(1-x^{m_k})}.$$

- Let $\mathcal{P}$ be the property that no part can occur more than k times. Then

$$\begin{aligned} p(n) &= [x^n](1+x+x^2+\ldots+x^k)(1+x^2+x^4+\ldots+x^{2k})(1+x^3+x^6+\ldots+x^{3k})\ldots \\ &= [x^n]\prod_{i=1}(1+x^i+x^{2i}+\ldots+x^{ki}) \\ &= [x^n]\prod_{i=1}\frac{1-x^{(k+1)i}}{1-x^i}. \end{aligned}$$

- Let $\mathcal{P}$ be the property that the value of each parts is odd. Then

$$p(n)=[x^n]\frac{1}{(1-x)(1-x^3)(1-x^5)\ldots}.$$

- Let $\mathcal{P}$ be the property that each part must have value at most m. Then

$$p(n)=[x^n]\frac{1}{(1-x^1)(1-x^2)\ldots(1-x^m)}.$$

- Let $\mathcal{P}_1$ be the property that there are precisely m parts. By the previous section, $p(n|\mathcal{P}_1)$ equals $p(n|\mathcal{P}_2)$ where $\mathcal{P}_2$ is the property that the largest part equals m. Therefore the generating function here is

$$(1+x+x^2+\ldots)(1+x^2+x^4+\ldots)\ldots(x^m+x^{2m}+\ldots)$$

which is equal to

$$\frac{x^m}{(1-x)(1-x^2)\ldots(1-x^m)}.$$

We shall now illustrate, by a very simple theorem due to Euler, the type of results obtainable using generating functions. Some more examples are give in the Miscellaneous Problem Sheet.

Theorem 14.3 *Let $\mathcal{P}_1$ be the property that the parts of the partition are distinct, and let $\mathcal{P}_2$ be the property that each part is odd. Then*

$$p(n|\mathcal{P}_1) = p(n|\mathcal{P}_2)$$

Proof Let D(x) and O(x) be the respective generating functions. Therefore

$$D(x) = (1+x)(+x^2)(1+x^3)\dots$$

and

$$O(x) = \frac{1}{(1-x)(1-x^3)(1-x^5)}\dots.$$

We need to show that $D(x) = O(x)$.

Recall that, for any number y, $1+y = (1-y^2)/(1-y)$. Applying this to all the terms of $D(x)$ (with y successively equal to $x, x^2, x^3, \dots$) gives

$$\begin{aligned} D(x) &= \frac{(1-x^2)(1-x^4)(1-x^6)\dots}{(1-x)(1-x^2)(1-x^3)\dots} \\ &= \frac{1}{(1-x)(1-x^3)(1-x^5)\dots} \\ &= O(x), \end{aligned}$$

which is what we had to prove. □

Note that in the generating function method, unlike the method of bijections, we do not obtain an explicit one-to-one correspondence between the two sets of partitions, although we still show that they have an equal number of elements. In a sense, we therefore obtain less of an understanding why the result is true. On the other hand, this method requires much less intuition than the bijection method, and it is as if the algebraic machinery is doing all the thinking for us. It is almost as if we can automate the method to generate the proof. This, and its very general applicability, is one of the advantages of the generating function method.

14.3 A recurrence relation for p(n)—combining the technique of Ferrers diagrams with generating function

Although there is no known formula for $p(n)$, there is a recurrence relation which computes $p(n)$ efficiently. Here it is, for completeness' sake.

$$p(n) = \sum_{m=1} (-1)^m \left[p\left(n - \frac{1}{2}m(3m-1)\right) + p\left(n - \frac{1}{2}m(3m+1)\right)\right]$$

with the convention that $p(k)$ is zero when k is negative.

Note that this is not a first, or second order relation or anything like that, because the number of previous terms required to find $p(n)$ increases with n. The derivation of this formula combines the use of Ferrers diagrams and generating functions. It is not so difficult that whoever has reached this stage cannot understand it; it is simply that we have no time for it in this course. An excellent exposition of its derivation and use can be found in Biggs, pp. 433–438.

15 Problem Sheet 4: Partitions of an Integer

1. $*$ This problem shows that the difficulty with partitions of an integer is that the order of the parts is not significant when counting $p(n)$.

   The number of *compositions of a positive integer* n, denoted by $\mathrm{comp}(n)$, is the number of ways of writing n as a sum of non-zero positive integers such that different orderings of the parts are counted as different compositions. Show that

   $$\mathrm{comp}(n) = 2^{n-1}.$$

   (Look back at Problem 3c

   When you see this result, what sort of proof does it suggest?

2. * Prove that the number of partitions of n which have at most m parts is equal to the number of partitions of $n + \frac{1}{2}m(m+1)$ in which there are m parts, all of which are different. [*Hint.* Use Ferrers diagrams; add a "triangle" with $\frac{1}{2}m(m+1)$ marks.]

3. * A partition of n is said to be *self-conjugate* if $D = D^T$ (where D is the Ferrers diagram of the partition and D^T is its conjugate). Prove that the number of self-conjugate partitions of n is equal to the number of partitions of n whose parts are distinct and odd.

4. * Write down the generating function for the sequence whose nth term is equal to the number of partitions of n such that no even number occurs more than once as a part.

   Hence show that this number is equal to the number of partitions of n in which each part occurs at most three times. [*Hint.* Use $(1-y^4) = (1-y)(1+y+y^2+y^3)$.]

5. The following rule is the basis of a method for listing all partitions of n in *lexicographic order*. The first partition is $[n]$. Suppose the current partition has parts $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r$. Then the next partition is found as follows:

   (a) If $\lambda_r \neq 1$, then the parts of the next partition are $\lambda_1, \lambda_2, \ldots, \lambda_r, \lambda_r - 1, 1$.

   (b) If $\lambda_r = \lambda_{r-1} = \ldots = \lambda_{r-s+1} = 1$ but $\lambda_{r-s} = x \neq 1$, then the parts of the next partition are obtained by replacing $\lambda_{r-s}, \ldots, \lambda_y$ by $x-1, x-1, \ldots, x-1, y$, where $1 \leq y \leq x-1$ and y is chosen so that the result is a partition of n.

   Use this algorithm to list lexicographically the partitions of 8. Write a computer programme based on this algorithm.

16 Solutions to the four motivating problems

We shall pose the four questions in a slightly more general setting:

In how many ways can you put r golf balls into q boxes such that no box remains empty?

Looking at the four variants of this problem will help us review a few of the main topics discussed in this course:

16.1 Golf balls are identical, boxes are distinguishable

This amounts to choosing r of the q boxes without order (since the balls are identical) and with repetition (equivalent to putting more than one ball in a box) allowed. We need, however, to deal with the proviso that no box is empty.

This can be taken care of by first putting one ball in each box, and this can be done in only one way, since the balls are identical. Then, the remaining $r-q$ balls are distributed, this being equivalent to choosing $r-q$ of the q boxes without order and with possible repetition. Therefore the number of ways of doing this is

$$\binom{p+(r-q)-1}{r-q} = \binom{r-1}{r-q}.$$

16.2 Golf balls are distinguishable, boxes are identical

What is important here is deciding which balls go together in the same box—which particular box is not important. This is equivalent to partitioning an r-set (the set of balls) into exactly q parts ("exactly" because no box is empty). This can be done in

$$S(r,q)$$

ways.

16.3 Golf balls are distinguishable, boxes are distinguishable

Now it does matter which box contains which balls. The problem is therefore not one of partitions, although it is closely related. Here we should see every arrangement as a surjection from an r-set to a q-set ("surjection" because no box is empty), and there are exactly

$$q!S(r,q)$$

of these.

16.4 Golf balls are identical, boxes are identical

This is actually the most difficult case, and it was only towards the end of the course that we could see the problem in its true perspective. Here, it is not important to know which balls go together and in which box—only the number of balls in a box and only the number of boxes. This is equivalent to asking for the number of partitions of the integer r into exactly (because no box is empty) q parts. The number of such arrangements is therefore

$$p(r|\text{ exactly } q \text{ parts }).$$

17 Problem sheet 5: Miscellaneous

Most of these problems are taken from past examination papers.

1. Find the number of ways of colouring k golf balls with four colours such that there are an odd number of balls coloured with the same colour.
2. Find the number of ways of placing m similar balls into n different boxes so that no box is empty.
3. Find the number of ways of distributing $2t+1$ similar objects amongst three distinct boxes so that no box will contain more than t objects.

4. A set of 14 identical golf balls is partitioned into five parts: two parts have 4 balls each while each of the other parts have two balls. The balls are to be coloured using two colours red and blue but balls in the same part of the partition are to be coloured the same. Obtain a generating function giving the inventory of the number of ways in which the colourings can be carried out. That is, the coefficient of of $b^i r^i$ in the expansion of the generating function should give the number of colourings with i blue balls and j red ones.

   What is the total number of possible colourings?

   What is the total number of possible colourings if six balls are to be coloured red and eight balls are to be coloured blue?

5. How many permutations are there of the digits $1, 2, \ldots, 9$ in which none of the patterns $23, 56, 89$ appears?

   How many permutations are there of the digits $1, 2, \ldots, 9$ in which exactly five digits are in their natural position?

   [You may leave factorials in your answers but any results on derangements must be derived and written out in full.]

6. (a) Let $S(n, k)$ denote the number of partitions of an n-set into k parts (subsets). Write down a recurrence relation giving $S(n, k)$ in terms of $S(n-1, k-1)$ and $S(n-1, k)$. Find the values of $S(6, k), 1 \leq k \leq 6$.

   (b) In how many ways can six golf balls labelled 1 to 6 be put into four identical boxes so that at most one box is empty.

   (c) Using induction on n or otherwise show that, for $n \geq 2$,

$$S(n, 2) = 2^{n-1} - 1.$$

7. (a) Two integers are said to be *relatively prime* if they have no common factors except for the factor 1. How many positive integers less than 1000 are relatively prime to both 1000 and 90?

   (b) Show that there are

$$\sum_{i=0}^{9} (-1)^i \binom{9}{i} (9-i)^n$$

   n-digit numbers using the digits $1, 2, \ldots, 9$ and such that each of the nine digits is used at least once.

   [**Hint:** Let A_k, $1 \leq k \leq 9$, be the set of all those n-digit numbers in which the digit k is not used.]

8. Use generating functions to find

   (i) The number of ways of placing $2r$ identical balls into eight distinct boxes such that each box gets an even number of balls.

   (ii) The number of ways in which a sum of 25 can be obtained when 10 distinct dice are rolled.

9. (a) Write down the coefficient of x^k in the expansion of $(1-x)^{-n}$ in ascending powers of x.

   (b) Find the coefficients of x^{20} in each of the following products:

   (i) $(x^2 + x^3 + x^4 + \ldots)^4$

   (ii) $(x^2 + x^3 + \cdots + x^{12})^4$

   (iii) $(x^2 + x^3 + x^4 + \ldots)^3 (x + x^2 + \cdots + x^6)$

   (c) How many ways are there to distribute $2k$ identical balls into 4 distinguishable boxes so that no box contains more than k balls?

10. (a) How many integers are there in the range 1 to 999 (inclusive) which are not divisible by any of 2, 5 or 23?

(b) Let $S(n,k)$ denote the number of ways of partitioning an n-set into k parts. Write down a recurrence relation for $S(n,k)$ and use this relation to find all values of $S(n,k)$ for $1 \le k \le n \le 5$.

(c) How many ways are there to distribute 5 distinguishable balls amongst 6 distinguishable boxes such that exactly three of the boxes are nonempty?

11. Each of twenty rooms contains five persons. Thirty persons are to be selected from these hundred with the condition that at least one person from each room is to be selected. Show that the number of ways in which this can be done is

$$\sum_{i=0}^{14} \binom{100-5i}{30}\binom{20}{i}(-1)^i.$$

[*Hint:* Let the rooms be numbered from 1 to 20 and let A_i be the set containing all the ways in which the selection can be made *without* choosing anyone from room i.]

12. Ten letters are to be chosen (order not important) using the letters a, b, c. In how many ways can this be done if:

    (i) Each of the three letters can be chosen an unlimited number of times?

    (ii) The letter a must be chosen at least once (but otherwise an unlimited number of times), the letter b cannot be chosen more than five times, and the letter c can be chosen an unlimited number of times?

13. (a) Solve the following recurrence relation

$$a_n = n^2 a_{n-1} + [(n+1)!]^2 \quad (n \ge 1)$$

given that $a_0 = 1$. [You may use the fact that $\sum_1^n i^2 = n(n+1)(2n+1)/6$.]

(b) Solve the recurrence relation

$$a_n - 5a_{n-1} + 6a_{n-2} = 3^n \quad (n \ge 2)$$

given that $a_0 = a_1 = 0$.

14. (a) Let $p(n)$ denote the number of partitions of the positive integer n and let $p(n|\mathcal{P})$ denote the number of partitions of n having property $\mathcal{P}$. Write down the generating functions of each of the following

    (i) $p(n)$;

    (ii) $p(n|$ all parts are distinct);

    (iii) $p(n|$ all parts are odd);

    (iv) $p(n|$ no part appears more than twice);

    (v) $p(n|$ no part is a multiple of 3).

Show that

$$p(n|\text{ all parts are distinct}) = p(n|\text{ all parts are odd})$$

and

$$p(n|\text{ no part appears more than twice}) = p(n|\text{ no part is a multiple of 3}).$$

[You may need to use $1+y = (1-y^2)/(1-y)$ and $1+y+y^2 = (1-y^3)/(1-y)$.]

(b) Using Ferrers diagrams, show that

(i) The number of partitions of $3n$ into n parts is equal to the number of partitions of $2n$ into at most n parts.

(ii) The number of partitions of n into exactly m parts is equal to the number of partitions of n such that the largest part equals m.

15. The sequence $\langle a_n \rangle$ satisfies the recurrence relation

$$a_{n+2} - 2\alpha a_{n+1} + \alpha^2 a_n = \beta^n \quad (n \geq 0)$$

(where α, β are non-zero constants) subject to the initial conditions $a_0 = a_1 = 0$.

Solve this recurrence relation (obtaining a_n in terms of α, β and n) in the two cases:

(i) $\beta \neq \alpha$.

(ii) $\beta = \alpha$.

16. (a) A loan of Lm3000 is taken from a bank. After a year, and at the end of every subsequent year, a repayment of LmP is effected. Moreover, at the end of every year the bank charges interest at the rate of 1 per cent of the amount owed during that year.

Let A_n denote the amount owed to the bank at the end of the nth year (therefore $A_0 = 3000$). Obtain and solve a recurrence relation for A_n.

How much should the repayment amount P be equal to if the loan (including all interests) is to be repaid by the end of the third year?

(b) Solve the recurrence relation

$$a_{n+2} - 5a_{n+1} + 6a_n = n.5^n (n \geq 0)$$

given that $a_0 = a_1 = 0$.

17. *In this question, $p(n)$ denotes the number of partitions of the positive integer n and $p(n|\mathcal{P})$ denotes the number of partitions of n satisfying a given property $\mathcal{P}$.*

(a) Write down a generating function for $p(n)$.

(b) Show that

$$p(n| \text{ number of parts } = m) = p(n| \text{ size of largest part } = m)$$

and

$$p(n| \text{ number of parts } \leq m) = p(n+m| \text{ number of parts } = m)$$

(c) Show that

$$p(n| \text{ all parts distinct}) = p(n| \text{ each part is odd})$$

and

$$\begin{aligned} & p(n|\text{no number occurs more than three times as a part}) \\ = \; & p(n| \text{ no multiple of 4 occurs as a part}) \\ = \; & p(n| \text{ no even number occurs more than once as a part}) \end{aligned}$$

[**Hint for last part of this question:**

$$(1 + y + y^2 + y^3) = \frac{1 - y^4}{1 - y} = \frac{(1 - y^2)(1 + y^2)}{1 - y} \;]$$

18. Let $p(n)$ denote the number of partitions of the positive integer n and let $p(n|\mathcal{P})$ denote the number of partitions satisfying a given property $\mathcal{P}$. Write down the generating function for $p(n)$.

Let $k \geq 1$ be fixed, let $\mathcal{P}_1$ be the property "No part in the partition appears more than k times" and let $\mathcal{P}_2$ be the property "No part in the partition is divisible by $k+1$". Prove that $p(n|\mathcal{P}_1) = p(n|\mathcal{P}_2)$.

19. (a) Solve the following recurrence relation

$$a_n = na_{n-1} + (n+1)! \quad (n \geq 1)$$

given that $a_0 = 1$.

(b) Let a_n denote the number of regions created by n mutually overlapping circles drawn on a sheet of paper such that no three circles have a common point of intersection.

Obtain and solve a recurrence relation for a_n.

[**Hint:** Two overlapping circles intersect at exactly two points. Consider by how much the number of regions increases when the nth circle is drawn.]

20. Solve the first order recurrence relation

$$a_n = na_{n-1} + (-1)^n \quad (n \geq 2)$$

given that $a_1 = 0$. What is the value of

$$\lim_{n\to\infty} \frac{a_n}{n!}?$$

21. (a) Find the number of five-letter words that use letters from the set $\{\alpha, \beta, \gamma\}$ and such that none of these three letters is missing from any word.

(b) How many permutations are there of the digits $1, 2, \ldots, 100$ in which no odd number appears in its natural position?

[In (c) you may give your answer as a summation of terms involving factorials.]

22. (a) Let $S(n, k)$ denote the number of partitions of an n-set into k parts (subsets). Write down a recurrence relation giving $S(n, k)$ in terms of $S(n-1, k-1)$ and $S(n-1, k)$. Find the values of $S(6, k), 1 \leq k \leq 6$.

(b) In how many ways can six golf balls labelled 1 to 6 be put into four identical boxes so that at most one box is empty.

(c) Using induction on n or otherwise show that, for $n \geq 2$,

$$S(n, 2) = 2^{n-1} - 1.$$

23. Suppose there are m boxes labelled $1, 2, \ldots, m$ respectively and n balls $(m \geq n)$ labelled $1, 2, \ldots .n$ respectively. The balls are put into the boxes such that no two balls are in the same box and no ball labelled i is put into box i, for all $1 \leq i \leq n$. Show that the number of ways in which this can be done is

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

24. Let $N = \{1, 2, \ldots, 2n\}$. Show that the number of permutations f on N with the property $f(2i) \neq 2i, 1 \leq i \leq n$ is given by

$$(2n)! \sum_{i=0}^{n} \frac{\binom{n}{i}}{[2n]_i} (-1)^i.$$

25. (a) Write down the coefficient of x^k in the expansion of $(1-x)^{-4}$ in ascending powers of x.

(b) Let $S(n,k)$ denote the number of ways of partitioning an n-set into k parts. Write down, without proof, a recurrence relation involving $S(n,k)$, $S(n-1,k-1)$ and $S(n-1,k)$. Use this recurrence relation and the values of $S(n,n)$ and $S(n,1)$ to calculate the value of $S(6,4)$.

(c) In how many ways can six balls be distributed amongst four boxes so that no box is empty, if:

(i) The balls are identical and the boxes are distinguishable?

(ii) The balls are distinguishable and the boxes are distinguishable?

(iii) The balls are distinguishable and the boxes are identical?

(d) How many ways are there to distribute 5 distinguishable balls amongst 6 distinguishable boxes such that exactly three of the boxes are nonempty?

26. Show that the number of partitions of n is equal to the number of partitions of $2n$ into exactly n parts and that this number is equal to the number of partitions of n with the largest part equal to n.

27. (a) Let $g(x) = \sum_{r=0}^{\infty} a_r x^r$ and $f(x) = \sum_{r=0}^{k} x^r$. Write down the coefficient of x^k in the product $f(x)g(x)$ in terms of the coefficients of $g(x)$.

(b) Write down the generating functions for:

(i) $p(n)$, the number of partitions of the integer n;

(ii) $p(0) + p(1) + \ldots + p(n)$, the number of partitions that add up to at most n;

(iii) The number of partitions that add up to an even number at most $2n$.