Week 5 — Induction and Recursion

Richard Earl*
Mathematical Institute, Oxford, OX1 2LB,

November 2003

Abstract

Induction. Strong induction. Binomial theorem. Difference equations. Fibonacci numbers.

Notation 1 *Throughout this article we shall use the notation* $\mathbb{N}$ *to denote the set of natural numbers* $\{0, 1, 2, 3, \ldots\}$. *So when we write that a statement is true for all* $n \in \mathbb{N}$ *then we are simply saying that the statement is true in each of the cases* $n = 0, 1, 2, 3, \ldots$

The symbol $n!$ *read 'n factorial' denotes the product* $1 \times 2 \times 3 \times \cdots \times n$. *As a convention* $0! = 1$.

1 Introduction

Mathematical statements can come in the form of a single proposition such as

$$3 < \pi \quad \text{or as} \quad 0 < x < y \implies x^2 < y^2,$$

but often they come as a family of statements such as

$$\begin{array}{ll} A & e^x > 0 \quad \text{for all real numbers } x; \\ B & 0 + 1 + 2 + \cdots + n = \dfrac{1}{2} n (n+1) \quad \text{for } n \in \mathbb{N}; \\ C & \displaystyle\int_0^\pi \sin^{2n} \theta \, \mathrm{d}\theta = \frac{(2n)!}{(n!)^2} \frac{\pi}{2^{2n}} \quad \text{for } n \in \mathbb{N}; \\ D & 2n + 4 \text{ can be written as the sum of two primes for all } n \in \mathbb{N}. \end{array}$$

Induction, or more exactly mathematical induction, is a particularly useful method of proof for dealing with families of statements which are indexed by the natural numbers, such as the last three statements above. We shall prove both statements B and C using induction (see below and Example 6). Statement B (and likewise statement C) can be approached with induction because in each case knowing that the nth statement is true helps enormously in showing that the $(n+1)$th statement is true — this is the crucial idea behind induction. Statement D, on the other hand, is a famous problem known as *Goldbach's Conjecture* (Christian Goldbach (1690–1764), who was a professor of mathematics at St. Petersburg, made this conjecture in a letter to Euler in 1742.and it is still an open problem). If we let $D(n)$ be the statement that $2n + 4$ can be written as the sum of two primes, then it is currently known that $D(n)$ is true for $n < 4 \times 10^{14}$. What makes statement D different, and more intractable to induction, is that in trying to verify $D(n+1)$ we can't generally make much use of knowledge of $D(n)$ and so we can't build towards a proof. For example, we can verify $D(17)$ and $D(18)$ by noting that

$$38 = 7 + 31 = 19 + 19, \quad \text{and} \quad 40 = 3 + 37 = 11 + 29 = 17 + 23.$$

Here, knowing that 38 can be written as a sum of two primes, is no help in verifying that 40 can be, as none of the primes we might use for the latter was previously used in splitting 38.

*These pages are produced by Richard Earl, who is the Schools Liaison and Access Officer for mathematics, statistics and computer science at Oxford University. Any comments, suggestions or requests for other material are welcome at earl@maths.ox.ac.uk

By way of an example we shall prove statement B by induction, before giving a formal definition of just what induction is. For any $n \in \mathbb{N}$, let $B(n)$ be the statement

$$0+1+2+\cdots+n=\frac{1}{2}n(n+1).$$

We shall prove two facts:

$$\text{(i) } B(0) \text{ is true and (ii) for any } n \in \mathbb{N}, \text{ if } B(n) \text{ is true then } B(n+1) \text{ is also true.}$$

The first fact is the easy part as we just need to note that

$$\text{LHS of } B(0)=0=\frac{1}{2}\times 0\times 1=\text{RHS of } B(0).$$

To verify (ii) we need to prove *for each* n that $B(n+1)$ is true *assuming* $B(n)$ to be true. Now

$$\text{LHS of } B(n+1)=0+1+\cdots+n+(n+1).$$

But, assuming $B(n)$ to be true, we know that the terms from 0 through to n add up to $n(n+1)/2$ and so

$$\begin{aligned}\text{LHS of } B(n+1) &= \frac{1}{2}n(n+1)+(n+1)\\ &= (n+1)\left(\frac{n}{2}+1\right)\\ &= \frac{1}{2}(n+1)(n+2)=\text{RHS of } B(n+1).\end{aligned}$$

This verifies (ii). Be sure that you understand the above calculation, it contains the important steps common to any proof by induction. Note in the final step that we have retrieved our original formula of $n(n+1)/2$, but with $n+1$ now replacing n everywhere; this was the expression that we always had to be working towards.

With induction we now know that B is true, i.e. that $B(n)$ is true for any $n \in \mathbb{N}$. How does this work? Well suppose we want to be sure $B(2)$ is correct — above we have just verified the following three statements:

$$B(0) \text{ is true}, \quad \text{if } B(0) \text{ is true then } B(1) \text{ is true}, \quad \text{if } B(1) \text{ is true then } B(2) \text{ is true}$$

and so putting the three together, we see that $B(2)$ is true: the first statement tells us that $B(0)$ is true and the second two are stepping stones, first to the truth about $B(1)$, and then on to proving $B(2)$.

Formally then the Principle of Induction is as follows:

Theorem 2 *(THE PRINCIPLE OF INDUCTION) Let $P(n)$ be a family of statements indexed by the natural numbers. Suppose that*

$$\begin{aligned}&P(0) \textit{ is true, and}\\ &\textit{for any } n \in \mathbb{N}, \textit{ if } P(n) \textit{ is true then } P(n+1) \textit{ is also true,}\end{aligned}$$

then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. Let S denote the subset of $\mathbb{N}$ consisting of all those n for which $P(n)$ is *false*. We aim to show that S is empty, i.e. that no $P(n)$ is false.

Suppose for a contradiction that S is non-empty. Any non-empty subset of $\mathbb{N}$ has a minimum element; let's write m for the minimum element of S. As $P(0)$ is true then $0 \notin S$, and so m is at least 1.

Consider now $m-1$. As $m \geq 1$ then $m-1 \in \mathbb{N}$ and further, as $m-1$ is smaller than the minimum element of S, then $m-1 \notin S$, i.e. $P(m-1)$ is true. But, as $P(m-1)$ is true, then induction tells us that $P(m)$ is also true. This means $m \notin S$, which contradicts m being the minimum element of S. This is our required contradiction, an absurd conclusion. We see that S being non-empty just doesn't hold water. If S being non-empty leads to a contradiction, then S must be empty. ■

It is not hard to see how we might amend the hypotheses of the theorem above to show

Corollary 3 *Let $N \in \mathbb{N}$ and let $P(n)$ be a family of statements for $n = N, N+1, N+2, \ldots$ Suppose that*

$$\begin{array}{l} P(N) \text{ is true, and} \\ \text{for any } n \geq N, \text{ if } P(n) \text{ is true then } P(n+1) \text{ is also true,} \end{array}$$

then $P(n)$ is true for all $n \geq N$.

This is really just induction again, but we have started the ball rolling at a later stage. Here is another version of induction, which is usually referred to a the *Strong Form Of Induction*:

Theorem 4 *(STRONG FORM OF INDUCTION) Let $P(n)$ be a family of statements for $n \in \mathbb{N}$. Suppose that*

$$\begin{array}{l} P(0) \text{ is true, and} \\ \text{for any } n \in \mathbb{N}, \text{ if } P(0), P(1), P(2), \ldots, P(n) \text{ are all true then so is } P(n+1), \end{array}$$

then $P(n)$ is true for all $n \in \mathbb{N}$.

To reinforce the need for proof, and to show how patterns can at first glance delude us, consider the following example. Take two points on the circumference of a circle and take a line joining them; this line then divides the circle's interior into two regions. If we take three points on the perimeter then the lines joining them will divide the disc into four regions. Four points can result in a maximum of eight regions — surely then, we can confidently predict that n points will maximally result in 2^{n-1} regions. Further investigation shows our conjecture to be true for $n = 5$, but to our surprise, however we take six points on the circle, the maximum number of regions attained is 31. Indeed the maximum number of regions attained from n points on the perimeter is given by the formula [2, p.18]

$$\frac{1}{24}\left(n^4 - 6n^3 + 23n^2 - 18n + 24\right).$$

Our original guess was way out!

There are other well-known 'patterns' that go awry in mathematics: for example, the number

$$n^2 - n + 41$$

is a prime number for $n = 1, 2, 3, \ldots, 40$ (though this takes some tedious verifying), but it is easy to see when $n = 41$ that $n^2 - n + 41 = 41^2$ is not prime. A more amazing example comes from the study of Pell's equation $x^2 = py^2 + 1$ in number theory, where p is a prime number and x and y are natural numbers. If $P(n)$ is the statement that

$$991n^2 + 1 \text{ is not a perfect square (i.e. the square of a natural number),}$$

then the first counter-example to $P(n)$ is staggeringly found at [1, pp. 2–3]

$$n = 12,055,735,790,331,359,447,442,538,767.$$

2 Examples

On a more positive note though, many of the patterns found in mathematics won't trip us at some later stage and here are some further examples of proof by induction.

Example 5 *Show that n lines in the plane, no two of which are parallel and no three meeting in a point, divide the plane into $n(n+1)/2+1$ regions.*

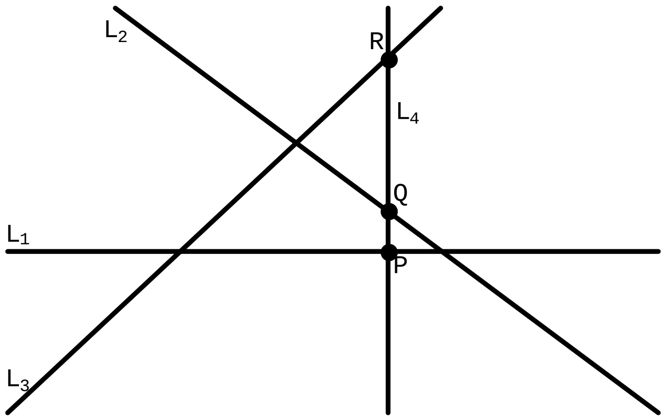
Proof. When we have no lines in the plane then clearly we have just one region, as expected from putting $n=0$ into the formula $n\,(n+1)\,/2+1$.

Suppose now that we have n lines dividing the plane into $n\,(n+1)\,/2+1$ regions and we will add a $(n+1)$th line. This extra line will meet each of the previous n lines because we have assumed it to be parallel with none of them. Also, it meets each of these n lines in a distinct point, as we have assumed that no three lines are concurrent.

These n points of intersection divide the new line into $n+1$ segments. For each of these $n+1$ segments there are now two regions, one on either side of the segment, where previously there had been only one region. So by adding this $(n+1)$th line we have created $n+1$ new regions. In total the number of regions we now have is

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

This is the correct formula when we replace n with $n+1$, and so the result follows by induction. ■



An example when $n=3$.

Here the four segments, 'below P', PQ, QR and 'above R' on the fourth line L_4, divide what were four regions previously, into eight new ones.

Example 6 *Prove for $n\in\mathbb{N}$ that*

$$\int_0^{\pi}\sin^{2n}\theta\ d\theta=\frac{(2n)!}{(n!)^2}\frac{\pi}{2^{2n}}. \tag{1}$$

Proof. Let's denote the integral on the LHS of equation (1) as I_n. The value of I_0 is easy to calculate, because the integrand is just 1, and so $I_0=\pi$. We also see

$$\text{RHS}\,(n=0)=\frac{0!}{(0!)^2}\frac{\pi}{2^0}=\pi,$$

verifying the initial case.

We now prove a *reduction formula* connecting I_n and I_{n+1}, so that we can use this in our induction.

$$\begin{aligned}
I_{n+1} &= \int_0^\pi \sin^{2(n+1)}\theta \, \mathrm{d}\theta \\
&= \int_0^\pi \sin^{2n+1}\theta \times \sin\theta \, \mathrm{d}\theta \\
&= \left[\sin^{2n+1}\theta \times (-\cos\theta)\right]_0^\pi - \int_0^\pi (2n+1)\sin^{2n}\theta\cos\theta \times (-\cos\theta)\ \mathrm{d}\theta \quad \text{[using integration by parts]} \\
&= 0 + (2n+1)\int_0^\pi \sin^{2n}\theta \left(1-\sin^2\theta\right)\ \mathrm{d}\theta \quad \text{[using } \cos^2\theta = 1-\sin^2\theta\text{]} \\
&= (2n+1)\int_0^\pi \left(\sin^{2n}\theta - \sin^{2(n+1)}\theta\right)\ \mathrm{d}\theta \\
&= (2n+1)\left(I_n - I_{n+1}\right).
\end{aligned}$$

Rearranging gives

$$I_{n+1} = \frac{2n+1}{2n+2} I_n.$$

Suppose now that equation (1) gives the right value of I_k for some natural number k. Then, turning to equation (1) with $n = k+1$, and using our assumption and the reduction formula, we see:

$$\begin{aligned}
\text{LHS} &= I_{k+1} = \frac{2k+1}{2(k+1)} \times \frac{(2k)!}{(k!)^2} \times \frac{\pi}{2^{2k}} \\
&= \frac{2k+2}{2(k+1)} \times \frac{2k+1}{2(k+1)} \times \frac{(2k)!}{(k!)^2} \times \frac{\pi}{2^{2k}} \\
&= \frac{(2k+2)!}{((k+1)!)^2} \times \frac{\pi}{2^{2(k+1)}},
\end{aligned}$$

which equals the RHS of equation (1) with $n = k+1$. The result follows by induction. ■

Example 7 *Show for $n = 1, 2, 3\ldots$ and $k = 1, 2, 3, \ldots$ that*

$$\sum_{r=1}^{n} r(r+1)(r+2)\cdots(r+k-1) = \frac{n(n+1)(n+2)\cdots(n+k)}{k+1}. \tag{2}$$

Remark 8 *This problem differs from our earlier examples in that our family of statements now involves two variables n and k, rather than just the one variable. If we write $P(n,k)$ for the statement in equation (2) then we can use induction to prove all of the statements $P(n,k)$ in various ways:*

- *we could prove $P(1,1)$ and show how $P(n+1,k)$ and $P(n,k+1)$ both follow from $P(n,k)$ for $n, k = 1, 2, 3, \ldots$;*
- *we could prove $P(1,k)$ for all $k = 1, 2, 3, \ldots$ and show how knowledge of $P(n,k)$ for all k, leads to proving $P(n+1,k)$ for all k — effectively this reduces the problem to one application of induction, but to a family of statements at a time*
- *we could prove $P(n,1)$ for all $n = 1, 2, 3, \ldots$ and show how knowing $P(n,k)$ for all n, leads to proving $P(n,k+1)$ for all n — in a similar fashion to the previous method, now inducting through k and treating n as arbitrary.*

What these different approaches rely on, is that all the possible pairs (n,k) are somehow linked to our initial pair (or pairs). Let

$$S = \{(n,k) : n, k \geq 1\}$$

be the set of all possible pairs (n,k).

The first method of proof uses the fact that the only subset T of S satisfying the properties

$$\begin{array}{rcll} (1,1) & \in & T, & \\ \text{if} \quad (n,k) & \in & T & \text{then} \quad (n,k+1) \in T, \\ \text{if} \quad (n,k) & \in & T & \text{then} \quad (n+1,k) \in T, \end{array}$$

is S itself. Starting from the truth of $P(1,1)$, and deducing further truths as the second and third properties allow, then every $P(n,k)$ must be true. The second and third methods of proof rely on the fact that the whole of S is the only subset having similar properties.

Proof. In this case is the second method of proof seems easiest, that is: we will prove that $P(1,k)$ holds for each $k = 1,2,3,\ldots$ and show that assuming statements $P(N,k)$ for a particular N and all k, is sufficient to prove the statements $P(N+1,k)$ for all k. Firstly we note

$$\begin{aligned} \text{LHS of } P(1,k) &= 1 \times 2 \times 3 \times \cdots \times k, \text{ and} \\ \text{RHS of } P(1,k) &= \frac{1 \times 2 \times 3 \times \cdots \times (k+1)}{k+1} = 1 \times 2 \times 3 \times \cdots \times k, \end{aligned}$$

are equal, proving $P(1,k)$ for all $k \geq 1$. Then, assuming $P(N,k)$ for all $k = 1,2,3,\ldots,$ we have

$$\begin{aligned} \text{LHS of } P(N+1,k) &= \sum_{r=1}^{N+1} r(r+1)(r+2)\cdots(r+k-1) \\ &= \frac{N(N+1)\ldots(N+k)}{k+1} + (N+1)(N+2)\cdots(N+k) \\ &= (N+1)(N+2)\cdots(N+k)\left(\frac{N}{k+1}+1\right) \\ &= \frac{(N+1)(N+2)\cdots(N+k)(N+k+1)}{k+1} = \text{RHS of } P(N+1,k), \end{aligned}$$

proving $P(N+1,k)$ simultaneously for each k. This verifies all that is required for the second method. ■

We end with one example which makes use of the Strong Form of Induction. Recall that a natural number $n \geq 2$ is called *prime* if the only natural numbers which divide it are 1 and n. (Note that 1 is not considered prime.) The list of prime numbers begins $2, 3, 5, 7, 11, 13, \ldots$ and has been known to be infinite since the time of Euclid. (Euclid was an Alexandrian Greek living c. 300 B.C. His most famous work is *The Elements*, thirteen books which present much of the mathematics discovered by the ancient Greeks, and which was a hugely influential text on the teaching of mathematics even into the twentieth century. The work presents its results in a rigorous fashion, laying down basic assumptions, called *axioms,* and carefully proving his theorems from these axioms.) The prime numbers are, in a sense, the atoms of the natural numbers under multiplication as every natural number $n \geq 2$ can be written as a product of primes in what is essentially a unique way — this fact is known as the Fundamental Theorem of Arithmetic. Here we just prove the existence of such a product.

Example 9 *Every natural number $n \geq 2$ can be written as a product of prime numbers.*

Proof. We begin at $n = 2$ which is prime. As our inductive hypothesis we assume that every number $2 \leq k \leq N$ is a prime number or can be written as a product of prime numbers. Consider then $N+1$; we need to show this is a prime number, or else a product of prime numbers. Either $N+1$ is prime or it is not. If $N+1$ is prime then we are done. If $N+1$ is not prime, then it has a factor $2 \leq m < N+1$ which divides $N+1$. Note that $m \leq N$ and $(N+1)/m \leq N$, as m is at least 2. So, by hypothesis, we know m and $(N+1)/m$ are both either prime or the product of prime numbers. Hence we can write

$$m = p_1 \times p_2 \times \cdots \times p_k, \quad \text{and} \quad \frac{N+1}{m} = P_1 \times P_2 \times \cdots \times P_K,$$

where $p_1, \ldots, p_k$ and $P_1, \ldots, P_K$ are prime numbers. Finally we have that

$$N+1 = m \times \frac{N+1}{m} = p_1 \times p_2 \times \cdots \times p_k \times P_1 \times P_2 \times \cdots \times P_K,$$

showing $N+1$ to be a product of primes. The result follows using the strong form of induction. ■

3 The Binomial Theorem

All of you will have met the identity

$$(x+y)^2 = x^2 + 2xy + y^2$$

and may even have met identities like

$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3.$$

It may even have been pointed out to you that these coefficients $1, 2, 1$ and $1, 3, 3, 1$ are simply the numbers that appear in *Pascal's Triangle.* This is the infinite triangle of numbers that has 1s down both sides and a number internal to some row of the triangle is calculated by adding the two numbers above it in the previous row. So the triangle grows as follows:

$$\begin{array}{lccccccccccccc}
n=0 & & & & & & & 1 & & & & & & \\
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
n=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
n=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

From the triangle we could say read off the identity

$$(x+y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6.$$

Of course we haven't *proved* this identity yet — these identities, for general n other than just $n = 6$, are the subject of the Binomial Theorem. We introduce now the binomial coefficients; their connection with Pascal's triangle will become clear soon.

Definition 10 *The (n, k)th binomial coefficient is the number*

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

where $n = 0, 1, 2, 3, \ldots$ and $0 \le k \le n$. It is read as 'n choose k' and in some books is denoted as ${}^nC_k$. As a convention we set $\binom{n}{k}$ to be zero when $n < 0$ or when $k < 0$ or $k > n$.

Note some basic identities concerning the binomial coefficients

$$\binom{n}{k} = \binom{n}{n-k}, \quad \binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n.$$

The following lemma demonstrates that the binomial coefficients are precisely the numbers that appear in Pascal's triangle.

Lemma 11 *Let $n \in \mathbb{N}$ and $1 \le k \le n$. Then*

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

Proof. Putting the LHS over a common denominator

$$\begin{aligned}
\text{LHS} &= \frac{n!}{(k-1)!\,(n-k+1)!} + \frac{n!}{k!\,(n-k)!} \\
&= \frac{n!}{k!\,(n-k+1)!}\{k + (n-k+1)\} \\
&= \frac{(n+1)!}{k!\,(n+1-k)!} \\
&= \binom{n+1}{k} = \text{RHS}.
\end{aligned}$$

■

Corollary 12 *The kth number in the nth row of Pascal's triangle is $\binom{n}{k}$ (remembering to count from $n = 0$ and $k = 0$). In particular the binomial coefficients are whole numbers.*

Proof. We shall prove this by induction. Note that $\binom{0}{0} = 1$ gives the 1 at the apex of Pascal's triangle, proving the initial step.

Suppose now that the numbers $\binom{N}{k}$ are the numbers that appear in the Nth row of Pascal's triangle. The first and last entries of the next, $(N+1)$th, row (associated with $k = 0$ and $k = N+1$) are

$$1 = \binom{N+1}{0}, \quad \text{and} \quad 1 = \binom{N+1}{N+1}$$

as required. For $1 \le k \le N$, then the kth entry on the $(N+1)$th row is formed by adding the $(k-1)$th and kth entries from the Nth row. By our hypothesis about the Nth row their sum is

$$\binom{N}{k-1} + \binom{N}{k} = \binom{N+1}{k},$$

using the previous lemma, and this verifies that the $(N+1)$th row also consists of binomial coefficients. So the $(N+1)$th row checks out, and the result follows by induction. ■

Finally, we come to the binomial theorem:

Theorem 13 *(THE BINOMIAL THEOREM): Let $n \in \mathbb{N}$ and x, y be real numbers. Then*

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

Proof. Let's check the binomial theorem first for $n = 0$. We can verify this by noting

$$\text{LHS} = (x+y)^0 = 1, \qquad \text{RHS} = \binom{0}{0} x^0 y^0 = 1.$$

We aim now to show the theorem holds for $n = N+1$ assuming it to be true for $n = N$. In this case

$$\text{LHS} = (x+y)^{N+1} = (x+y)(x+y)^N = (x+y)\left(\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k}\right)$$

writing in our assumed expression for $(x+y)^N$. Expanding the brackets gives

$$\sum_{k=0}^{N} \binom{N}{k} x^{k+1} y^{N-k} + \sum_{k=0}^{N} \binom{N}{k} x^k y^{N+1-k},$$

which we can rearrange to

$$x^{N+1} + \sum_{k=0}^{N-1} \binom{N}{k} x^{k+1} y^{N-k} + \sum_{k=1}^{N} \binom{N}{k} x^k y^{N+1-k} + y^{N+1}$$

by taking out the last term from the first sum and the first term from the second sum. In the first sum we now make a change of variable. We set $k = l - 1$, noting that as k ranges over $0, 1, ..., N-1$, then l ranges over $1, 2, ..., N$. So the above equals

$$x^{N+1} + \sum_{l=1}^{N} \binom{N}{l-1} x^l y^{N+1-l} + \sum_{k=1}^{N} \binom{N}{k} x^k y^{N+1-k} + y^{N+1}.$$

We may combine the sums as they are over the same range, obtaining

$$x^{N+1} + \sum_{k=1}^{N} \left\{ \binom{N}{k-1} + \binom{N}{k} \right\} x^k y^{N+1-k} + y^{N+1}$$

which, using Lemma 11, equals

$$x^{N+1}+\sum_{k=1}^{N}\binom{N+1}{k}x^{k}y^{N+1-k}+y^{N+1}=\sum_{k=0}^{N+1}\binom{N+1}{k}x^{k}y^{N+1-k}=\text{RHS}.$$

The result follows by induction. ■

There is good reason why $\binom{n}{k}$ is read as 'n choose k' — there are $\binom{n}{k}$ ways of choosing k elements from the set$\{1,2,\ldots,n\}$ (when showing no interest in the order that the k elements are to be chosen). Put another way, there are $\binom{n}{k}$ subsets of $\{1,2,\ldots,n\}$ with k elements in them. To show this, let's think about how we might go about choosing k elements.

For our 'first' element we can choose any of the n elements, but once this has been chosen it can't be put into the subset again. So for our second element any of the remaining $n-1$ elements may be chosen, for our third any of the $n-2$ that are left, and so on. So choosing a set of k elements from $\{1,2,\ldots,n\}$ in a particular order can be done in

$$n\times(n-1)\times(n-2)\times\cdots\times(n-k+1)=\frac{n!}{(n-k)!}\quad\text{ways.}$$

But there are lots of different orders of choice that could have produced this same subset. Given a set of k elements there are $k!$ ways of ordering this (see Exercise **??** below) — that is to say, for each subset with k elements there are $k!$ different orders of choice that will lead to that subset. So the number $n!/(n-k)!$ is an 'overcount' by a factor of $k!$. Hence the number of subsets of size k equals

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

as required.

Remark 14 *There is a Trinomial Theorem and further generalisations of the binomial theorem to greater numbers of variables. Given three real numbers x,y,z and a natural number n we can apply the binomial theorem twice to obtain*

$$\begin{aligned}(x+y+z)^{n} &= \sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}x^{k}\,(y+z)^{n-k}\\ &= \sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}\sum_{l=0}^{n-k}\frac{(n-k)!}{l!\,(n-k-l)!}x^{k}y^{l}z^{n-k-l}\\ &= \sum_{k=0}^{n}\sum_{l=0}^{n-k}\frac{n!}{k!l!\,(n-k-l)!}x^{k}y^{l}z^{n-k-l}.\end{aligned}$$

This is a somewhat cumbersome expression; it's easier on the eye, and has a nicer symmetry, if we write $m=n-k-l$ and then we can rewrite the above as

$$(x+y+z)^{n}=\sum_{\substack{k+l+m=n\\ k,l,m\geq 0}}\frac{n!}{k!\,l!\,m!}x^{k}\,y^{l}\,z^{m}.$$

Again the number $n!/(k!\,l!\,m!)$, where $k+l+m=n$ and $k,l,m\geq 0$, is the number of ways that n elements can be apportioned into three subsets associated with the numbers x,y and z.

4 Difference Equations

In this chapter we shall be mainly interested in solving linear difference equations with constant coefficients — that is finding an expression for numbers x_n defined recursively by a relation such as

$$x_{n+2} = 2x_{n+1} - x_n + 2 \quad \text{for } n \geq 0, \quad \text{with } x_0 = 1, x_1 = 1.$$

We see that the x_n can be determined by applying this relation sufficiently many times from our initial values of $x_0 = 1$ and $x_1 = 1$. So for example to find x_7 we'd calculate

$$\begin{aligned}
x_2 &= 2x_1 - x_0 + 1 = 2 - 1 + 2 = 3; \\
x_3 &= 2x_2 - x_1 + 1 = 6 - 1 + 2 = 7; \\
x_4 &= 2x_3 - x_2 + 1 = 14 - 3 + 2 = 13; \\
x_5 &= 2x_4 - x_3 + 1 = 26 - 7 + 2 = 21; \\
x_6 &= 2x_5 - x_4 + 1 = 42 - 13 + 2 = 31; \\
x_7 &= 2x_6 - x_5 + 1 = 62 - 21 + 2 = 43.
\end{aligned}$$

If this was the first time we had seen such a problem, then we might try pattern spotting or qualitatively analysing the sequence's behaviour, in order to make a guess at a general formula for x_n. Simply looking at the sequence x_n above no obvious pattern is emerging. However we can see that the x_n are growing, roughly at the same speed as n^2 grows. We might note further that the differences between the numbers $0, 2, 4, 6, 8, 10, 12, \ldots$ are going up linearly. Even if we didn't know how to sum an arithmetic progression, it would seem reasonable to *try* a solution of the form

$$x_n = an^2 + bn + c, \tag{3}$$

where a, b, c are constants, as yet undetermined. We can can find a, b, c using the first three cases, so that

$$\begin{aligned}
x_0 &= 1 = a0^2 + b0 + c \text{ and so } c = 1; \\
x_1 &= 1 = a1^2 + b1 + 1 \text{ and so } a + b = 0; \\
x_2 &= 3 = a2^2 - a2 + 1 \text{ and so } a = 1.
\end{aligned}$$

So the unique solution of the form (3) which gives the right answer in the $n = 0, 1, 2$ cases is

$$x_n = n^2 - n + 1. \tag{4}$$

If we put $n = 3, 4, 5, 6, 7$ into (4) then we get the correct values of x_n calculated above. This is, of course, not a proof, but we could prove this formula to be correct for all values of $n \geq 0$ using induction.

Alternatively having noted the differences go up as $0, 2, 4, 6, 8, 10, 12, \ldots$ we can write, using statement B from the start of this article,

$$\begin{aligned}
x_n &= 1 + 0 + 2 + 4 + \cdots + (2n - 2) \\
&= 1 + \sum_{k=0}^{n-1} 2k \\
&= 1 + 2\frac{1}{2}(n-1)\, n \\
&= n^2 - n + 1.
\end{aligned}$$

To make this proof water-tight we need to check that the pattern $0, 2, 4, 6, 8, 10, 12, \ldots$ of the differences carries on forever, and that it wasn't just a fluke, but this follows if we note

$$x_{n+2} - x_{n+1} = x_{n+1} - x_n + 2$$

and so the difference between consecutive terms is increasing by 2 each time.

Of course if a pattern to x_n is difficult to spot then the above methods won't apply. We will show now how to solve a difference equation of the form

$$ax_{n+2} + bx_{n+1} + cx_n = 0$$

where a, b, c are real (or complex) constants. The theory extends to constant coefficient difference equations of any order. We will later treat some inhomogeneous examples where the RHS is non-zero. As with constant coefficient differential equations this involves solving the corresponding homogeneous difference equation and finding a particular solution of the inhomogeneous equation. We will see that the theory has much in common with the differential equations material we met in week 3 — this is because the underlying linear algebra behind solving the two sets of problems is identical.

Theorem 15 *Suppose that the sequence x_n satisfies the difference equation*

$$ax_{n+2} + bx_{n+1} + cx_n = 0 \quad \text{for } n \geq 0, \tag{5}$$

and that α and β be the roots the auxiliary equation

$$a\lambda^2 + b\lambda + c = 0.$$

The general solution of (5) has the form

$$x_n = A\alpha^n + B\beta^n \quad (n \geq 0),$$

when α and β are distinct, and has the form

$$x_n = (An + B)\,\alpha^n \quad (n \geq 0),$$

when $\alpha = \beta \neq 0$. In each case the values of A and B are uniquely determined by the values of x_0 and x_1.

Proof. Firstly we note that the sequence x_n defined in (5) is uniquely determined by the initial values x_0 and x_1. Knowing these values (5) gives us x_2, knowing x_1 and x_2 gives us x_3 etc. So if we can find a solution to (5) for certain initial values then we have *the* unique solution; if we can find a solution for arbitrary initial values then we have the general solution.

Note that if $\alpha \neq \beta$ then putting $x_n = A\alpha^n + B\beta^n$ into the LHS of (5) gives

$$\begin{aligned} ax_{n+2} + bx_{n+1} + cx_n &= a\left(A\alpha^{n+2} + B\beta^{n+2}\right) + b\left(A\alpha^{n+1} + B\beta^{n+1}\right) + c\left(A\alpha^n + B\beta^n\right) \\ &= A\alpha^n\left(a\alpha^2 + b\alpha + c\right) + B\beta^n\left(a\beta^2 + b\beta + c\right) \\ &= 0 \end{aligned}$$

as α and β are both roots of the auxiliary equation.

Similarly if $\alpha = \beta$ then putting $x_n = (An + B)\,\alpha^n$ into the LHS of (5) gives

$$\begin{aligned} ax_{n+2} + bx_{n+1} + cx_n &= a\left(A(n+2) + B\right)\alpha^{n+2} + b\left(A(n+1) + B\right)\alpha^{n+1} + c\left(An + B\right)\alpha^n \\ &= A\alpha^n\left(n\left(a\alpha^2 + b\alpha + c\right) + (2a\alpha + b)\,\alpha\right) + B\alpha^n\left(a\alpha^2 + b\alpha + c\right) \\ &= 0 \end{aligned}$$

because α is a root of the auxiliary equation and also of the derivative of the auxiliary equation, being a repeated root (or, if you prefer, you can show that $2a\alpha + b = 0$ by remembering that $ax^2 + bx + c = a(x-\alpha)^2$, comparing coefficients and eliminating c).

So in either case we see that we have a set of solutions. But further the initial equations

$$A + B = x_0, \quad A\alpha + B\beta = x_1,$$

are uniquely solvable for A and B when $\alpha \neq \beta$, whatever the values of x_0 and x_1. Similarly when $\alpha = \beta \neq 0$ then the initial equations

$$B = x_0, \quad (A + B)\,\alpha = x_1,$$

also have a unique solution in A and B whatever the values of x_0 and x_1. So in each case our solutions encompassed the general solution. ■

Remark 16 *When $\alpha = \beta = 0$ then (5) clearly has solution x_n given by*

$$x_0, x_1, 0, 0, 0, 0, \ldots$$

Probably the most famous sequence defined by such a difference equation is the sequence of Fibonacci numbers. The Fibonacci numbers F_n are defined recursively

$$F_{n+2} = F_{n+1} + F_n, \quad \text{for} \quad n \geq 0,$$

with initial values $F_0 = 0$ and $F_1 = 1$. So the sequence begins as

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

and continues to grow, always producing whole numbers and increasing by a factor of roughly 1.618 each time.

This sequence was first studied by Leonardo of Pisa (c.1170–c.1250), who called himself Fibonacci. (the meaning of the name 'Fibonacci' is somewhat uncertain; it may have meant 'son of Bonaccio' or may have been a nickname meaning 'lucky son'.) The numbers were based on a model of rabbit reproduction: the model assumes that we begin with a pair of rabbits in the first month, which every month produces a new pair of rabbits, which in turn begin producing when they are one month old. If we look at the F_n pairs we have at the start of the nth month, then these consist of $F_{n-1} - F_{n-2}$ pairs which have just become mature but were immature the previous month, and F_{n-2} pairs which were already mature and their new F_{n-2} pairs of offspring. In other words

$$F_n = (F_{n-1} - F_{n-2}) + 2F_{n-2}$$

which rearranges to the recursion above.

Proposition 17 *For every integer $n \geq 0$,*

$$F_n = \frac{\alpha^n - \beta^n}{\sqrt{5}} \tag{6}$$

where

$$\alpha = \frac{1+\sqrt{5}}{2} \quad \textit{and} \quad \beta = \frac{1-\sqrt{5}}{2}.$$

Proof. From our previous theorem we know that

$$F_n = A\alpha^n + B\beta^n \quad \text{for } n \geq 0,$$

where α and β are the roots of the auxiliary equation

$$x^2 - x - 1 = 0,$$

that is the α and β given in the statement of the proposition, and where A and B are constants uniquely determined by the equations

$$\begin{aligned} A + B &= 0, \\ A\alpha + B\beta &= 1. \end{aligned}$$

So

$$A = -B = \frac{1}{\alpha - \beta} = \frac{1}{\sqrt{5}},$$

concluding the proof. ■

We end with two examples of inhomogeneous difference equations.

Example 18 *Find the solution of the following difference equation*

$$x_{n+2} - 4x_{n+1} + 4x_n = 2^n + n, \tag{7}$$

with initial values $x_0 = 1$ *and* $x_1 = -1$.

Proof. The auxiliary equation

$$\lambda^2 - 4\lambda + 4 = 0$$

has repeated roots $\lambda = 2, 2$. So the general solution of the homogeneous equation

$$x_{n+2} - 4x_{n+1} + 4x_n = 0 \tag{8}$$

we know, from Theorem 15 to be $x_n = (An + B)\, 2^n$ where A and B are undetermined constants.

In order to find a particular solution of the recurrence relation (7) we need to try a solution $x_n = f(n)$ where $f(n)$ is a function similar in nature to $2^n + n$. We can deal with the n on the RHS by contributing a term $an+b$ to $f(n)$. But trying to deal with the 2^n on the RHS with contributions to $f(n)$ that consist of some multiple of 2^n or $n2^n$ would be useless as 2^n and $n2^n$ are both solutions of the homogeneous equation (8), and so trying them would just yield a zero on the RHS — rather we need to try instead a multiple of $n^2 2^n$ to deal with the 2^n. So let's try a particular solution of the form

$$x_n = an + b + cn^2 2^n,$$

where a, b, c are constants, as yet undetermined. Putting this expression for x_n into the LHS of (7) we get

$$\begin{aligned} & a(n+2) + b + c(n+2)^2\, 2^{n+2} - 4a(n+1) - 4b - 4c(n+1)^2\, 2^{n+1} + 4an + 4b + 4cn^2 2^n \\ = \;& a(n + 2 - 4n - 4 + 4n) + b(1 - 4 + 4) + c2^n \left(4n^2 + 16n + 16 - 8n^2 - 16n - 8 + 4n^2\right) \\ = \;& an + (b - 2a) + 8c2^n. \end{aligned}$$

This expression we need to equal $2^n + n$ and so we see that $a = 1, b = 2, c = 1/8$. Hence a particular solution is

$$x_n = n + 2 + \frac{n^2}{8} 2^n,$$

and the general solution of (7) is

$$x_n = (An + B)\, 2^n + n + 2 + \frac{n^2}{8} 2^n.$$

Recalling the initial conditions $x_0 = 1$ and $x_1 = -1$ we see

$$\begin{aligned} n &= 0: \quad B + 2 = 1; \\ n &= 1: \quad 2(A + B) + 1 + 2 + \frac{1}{4} = -1. \end{aligned}$$

The first line gives us $B = -1$ and the second that $A = -9/8$. Finally then the unique solution of (7) is

$$x_n = n + 2 + \frac{1}{8}\left(n^2 - 9n - 8\right) 2^n.$$

■

Example 19 *Find the solution of the difference equation*

$$x_{n+3} = 2x_n - x_{n+2} + 1,$$

with initial values $x_0 = x_1 = x_2 = 0$.

Proof. The auxiliary equation here is

$$\lambda^3 + \lambda^2 - 2 = 0,$$

which factorises as

$$\lambda^3 + \lambda^2 - 2 = (\lambda - 1)\left(\lambda^2 + 2\lambda + 2\right) = 0$$

and so has roots

$$\lambda = 1, -1 + i, -1 - i.$$

So the general solution of the homogeneous difference equation is

$$x_n = A + B\left(-1 + i\right)^n + C\left(-1 - i\right)^n.$$

At this point we know need to find a particular solution of the inhomogeneous equation. Because constant sequences are solutions of the homogeneous equation there is no point trying these as particular solutions; instead we try one of the form $x_n = kn$. Putting this into the difference equation we obtain

$$k\left(n + 3\right) = 2kn - k\left(n + 2\right) + 1 \quad \text{which simplifies to } 3k = -2k + 1$$

and so $k = \frac{1}{5}$. The general solution of the inhomogeneous difference equation has the form

$$x_n = \frac{n}{5} + A + B\left(-1 + i\right)^n + C\left(-1 - i\right)^n.$$

At first glance this solution does not necessarily look like it will be a real sequence, and indeed B and C will need to be complex constants for this to be the case. But if we remember that

$$\begin{aligned} \left(-1 + i\right)^n &= \left(\sqrt{2}e^{i3\pi/4}\right)^n = 2^{n/2}\left(\cos\frac{3n\pi}{4} + i\sin\frac{3n\pi}{4}\right) \\ \left(-1 - i\right)^n &= \left(\sqrt{2}e^{i5\pi/4}\right)^n = 2^{n/2}\left(\cos\frac{5n\pi}{4} + i\sin\frac{5n\pi}{4}\right) \end{aligned}$$

we can rearrange our solution in terms of overtly real sequences.

To calculate A, B and C then we use the initial conditions. We see that

$$\begin{aligned} n &= 0: \quad A + B + C = 0; \\ n &= 1: \quad A + B\left(-1 + i\right) + C\left(-1 - i\right) = \frac{-1}{5}; \\ n &= 2: \quad A + B\left(-2i\right) + C\left(2i\right) = \frac{-2}{5}. \end{aligned}$$

Substituting in $A = -B - C$ from the first equation we have

$$\begin{aligned} B\left(-2 + i\right) + C\left(-2 - i\right) &= \frac{-1}{5}; \\ B\left(-1 - 2i\right) + C\left(-1 + 2i\right) &= \frac{-2}{5}, \end{aligned}$$

and solving these gives

$$B = \frac{4 - 3i}{50} \quad \text{and} \quad C = \frac{4 + 3i}{50}, \quad \text{and} \quad A = -B - C = \frac{-8}{50}.$$

Hence the unique solution is

$$\begin{aligned} x_n &= \frac{n}{5} + \frac{-8}{50} + \frac{4-3i}{50}(-1+i)^n + \frac{4+3i}{50}(-1-i)^n \\ &= \frac{1}{50}\left(10n - 8 + (4-3i)(-1+i)^n + (4+3i)(-1-i)^n\right). \end{aligned}$$

The last two terms are conjugates of one another and so, recalling that

$$z + \overline{z} = 2\operatorname{Re} z$$

we have

$$\begin{aligned} x_n &= \frac{1}{50}\left(10n - 8 + 2\operatorname{Re}\left[(4-3i)(-1+i)^n\right]\right) \\ &= \frac{1}{50}\left(10n - 8 + 2 \times 2^{n/2}\operatorname{Re}\left[(4-3i)\left(\cos\frac{3n\pi}{4} + i\sin\frac{3n\pi}{4}\right)\right]\right) \\ &= \frac{1}{50}\left(10n - 8 + 2^{n/2+1}\left(4\cos\frac{3n\pi}{4} + 3\sin\frac{3n\pi}{4}\right)\right) \\ &= \frac{1}{25}\left(5n - 4 + 2^{n/2}\left(4\cos\frac{3n\pi}{4} + 3\sin\frac{3n\pi}{4}\right)\right) \end{aligned}$$

■

References

[1] Joseph Rotman, *Journey Into Mathematics — An Introduction to Proofs*, Prentice Hall, 1998.

[2] David Acheson, *1089 And All That,* Oxford University Press, 2002.