4 Induction

Let X be any set of numbers that satisfies each of the axioms P1 through P12 (X might be the rational numbers, or the real numbers, or any number of other possibilities). Inside X there is a copy of what we will think of as the "natural numbers", namely

$$\mathbb{N} = \{1, 1+1, 1+1+1, \ldots\} \quad \text{or} \quad \mathbb{N} = \{1, 2, 3, \ldots\}.$$

(I'm going to assume that everyone is familiar with the standard naming convention of Arabic numbers!) Notice that we have

$$1 < 1+1 < 1+1+1 < \cdots \quad \text{or} \quad 1 < 2 < 3 < \cdots,$$

since $1 > 0$, so adding one more "1" to a sum of a collection of 1's increases the sum.

This definition of the natural numbers is somewhat informal (what exactly does that "..." mean?), but it will work perfectly well for us while we introduce the most important property of the natural numbers, the principle of mathematical induction. In this section we'll discuss induction first in this informal setting. We'll then present a more formal definition of $\mathbb{N}$, and indicate how (in principle at least) we could establish all of $\mathbb{N}$'s expected properties in this formal setting.

4.1 The principle of mathematical induction (informally)

We have already encountered a number of situations in which we would like to be able to prove that some predicate, that depends on a natural number n, is true for *every* $n \in \mathbb{N}$. Examples include:

- if $a_1, \ldots, a_n$ are n arbitrary reals, then the sum $a_1 + a_2 + \cdots + a_n$ does not depend on the order in which parentheses are put around the a_i's, and
- if $a_1, \ldots, a_n$ are n arbitrary reals, then the sum of the a_i's does not depend on the order in which the a_i's are arranged in the sum.

We know that we can, in principle, use the axioms of the real numbers to prove each of these statements *for any particular* n, but it seems like this case-by-case approach would require *infinite* time to prove either of the statements for *every* n.

There's a fix. Let's pick one of these predicates, call it $p(n)$. Suppose we can prove

A that $p(1)$ is true

and we can also give an argument that shows that

B for any arbitrary natural number n, $p(n)$ implies $p(n+1)$.

Armed with these two weapons, we have a convincing argument that $p(n)$ is true for *every* n. Indeed, if a friend were to challenge us to provide them with a proof of $p(7)$, we would tell them:

- well, $p(1)$ is true (that's **A**), so
- since $p(1)$ is true, and $p(1)$ implies $p(2)$ (that's **B**, in the specific case $n = 1$), we conclude via modus ponens that $p(2)$ is true, so
- since $p(2)$ is true, and $p(2)$ implies $p(3)$ (**B** for $n = 2$), modus ponens again tells us that $p(3)$ is true, so
- since $p(3)$ is true, and $p(3)$ implies $p(4)$, $p(4)$ is true, so
- since $p(4)$ is true, and $p(4)$ implies $p(5)$, $p(5)$ is true, so
- since $p(5)$ is true, and $p(5)$ implies $p(6)$, $p(6)$ is true, so
- since $p(6)$ is true, and $p(6)$ implies $p(7)$, $p(7)$ is true.

And if instead they challenged us to prove $p(77)$, we would do the same thing, just with many more lines. There's a *uniform* proof of $p(n)$ for *any* n — one that doesn't require a specific examination of $p(n)$, but simply one appeal to **A** followed by $n - 1$ identical appeals to **B** and modus ponens. Because of this uniformity, we can simply present **A** and **B** as a proof of $p(n)$ for *all* n. If our friend wants a specific proof of $p(777)$ from this, they are free to supply the 777 required steps themselves!

As long as **A** and **B** can both be given finite length proofs, this gives a finite length proof of $p(n)$ for infinitely many n. We summarize this:

> **The principle of mathematical induction**: Let $p(n)$ be a predicate, with the universe of discourse for n being natural numbers. If $p(1)$ is true, and if, for arbitrary n, $p(n)$ implies $p(n + 1)$, then $p(n)$ is true for all n.

Some notation:

- a proof using the principle of mathematical induction is commonly called a *proof by induction*;
- the step in which $p(1)$ is verified is called the *base case* of the induction; and
- the step in which it is established that for arbitrary n, $p(n)$ implies $p(n + 1)$ (a step which will almost always involve symbolic manipulations of expressions involving n, where no *specific* properties of n are used), is called the *induction step*.

Here is a very tangible illustration of what's going on with induction:

> **The principle of mathematical induction, ladder version**: If you have a way of getting on a ladder, and if you have a way of going from any rung of the ladder to the next rung up, then you can get as high up the ladder as you wish.

Proving identities via induction

Let's have an example. What's the sum of the first n natural numbers? Well:

- $1 = 1$,
- $1 + 2 = 3$,
- $1 + 2 + 3 = 6$,
- $1 + 2 + 3 + 4 = 10$,
- $1 + 2 + 3 + 4 + 5 = 15$,
- $1 + 2 + 3 + 4 + 5 + 6 = 21$,
- $1 + 2 + 3 + 4 + 5 + 6 + 7 = 28$,
- $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 = 36$,
- $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = 45$,
- $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$.

A pattern seems to be emerging: it appears that $1 + 2 + \ldots + n = n(n + 1)/2$.

Claim 4.1. *For every natural number n,*

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}.$$

Proof: Let $p(n)$ be the predicate

$$p(n) : \text{“}1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}\text{”}$$

(where the universe of discourse for n is natural numbers). We prove that $p(n)$ is true for all n, by induction.

Base case: $p(1)$ is the assertion that $1 = 1(2)/2$, or $1 = 1$, which is true.

Induction step: Let n be an arbitrary natural number. We want to establish the implication

$$p(n) \text{ implies } p(n + 1),$$

that is to say, we want the establish that this statement has truth value T. By definition of implication, this is the same as showing that the statement

$$\text{either (not } p(n)) \text{ or } p(n + 1)(\star)$$

has truth value T.

If $p(n)$ is false, then (not $p(n)$) is true, so $(\star)$ is indeed true. If $p(n)$ is true, then (not $p(n)$) is false, so to establish that $(\star)$ is true we need to show that $p(n+1)$ is true. *But*, we don't have to start an argument establishing $p(n+1)$ from scratch — we are in the case where $p(n)$ is true, so *we get to assume* $p(n)$ *as part of our proof of* $p(n+1)$.

$p(n+1)$ is the assertion

$$1+2+3+\cdots+n+(n+1)=\frac{(n+1)((n+1)+1)}{2}$$

or

$$(1+2+3+\cdots+n)+(n+1)=\frac{(n+1)((n+2)}{2}.(\star\star)$$

Since $p(n)$ is being assumed to be true, we get the assume that

$$1+2+3+\cdots+n=\frac{n(n+1)}{2},$$

and so $(\star\star)$ (the statement whose truth we are trying to establish) becomes

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

Multiplying both sides by 2, and expanding out the terms, this becomes

$$n^2+n+2n+2=n^2+3n+2,$$

which is true.

We have established the truth of the implication "$p(n)$ implies $p(n+1)$", for arbitrary n, and so we have showing that the induction step is valid.

Conclusion: By the principle of mathematical induction, $p(n)$ is true for all natural numbers n, that is,

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}.$$

□

Of course, this write-up was filled with overkill. In particular, in proving the truth of the implication $p \Rightarrow q$ we almost never explicitly write that if the premise p is false then the implication is true; so it is very typical to start the induction step with "Assume $p(n)$. We try to deduce $p(n+1)$ from this." Also, we very often don't even explicitly introduce the predicate $p(n)$. Here is a more condensed write-up of the proof, that should act as template for other proofs by induction.

Claim 4.2. *For every natural number n,*

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}.$$

Proof: We proceed by induction on n.

Base case: The base case $n = 1$ is obvious.

Induction step: Let n be an arbitrary natural number. Assume

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}.$$

From this we get

$$\begin{aligned} 1 + 2 + 3 + \ldots + n + (n+1) &= (1 + 2 + 3 + \ldots + n) + (n+1) \\ &= \frac{n(n+1)}{2} + n + 1 \\ &= \frac{n^2 + n + 2n + 2}{2} \\ &= \frac{n^2 + 3n + 2}{2} \\ &= \frac{(n+1)(n+2)}{2} \\ &= \frac{(n+1)((n+1)+1)}{2}. \end{aligned}$$

The equality of the first and last expressions in this chain is the case $n + 1$ of the assertion, so we have verified the induction step.[57]

By induction the assertion is true for all n. □

In proving an identity — an equality between two expressions, both depending on some variable(s) — by induction, it is often very helpful to start with one side of the $n + 1$ case of the identity, and manipulate it via a sequence of equalities in a way that introduces one side of the n case of the identity into the mix; this can then be replaced with the *other* side of the n case, and then the whole thing might be massage-able into the other side of the $n + 1$ identity. That's exactly how we proceeded above.

Now is a good time to introduce *summation notation*. We write

$$\sum_{k=1}^{n} a_k$$

as shorthand for

$$a_1 + a_2 + a_3 + \cdots + a_{k-1} + a_k.$$

k is called the *index of summation*, and the a_k's are the *summands*. For example, we have

$$\sum_{k=1}^{7} k^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2,$$

[57] Notice that the induction step is presented here as a complete english-language paragraph, even though it involves a lot of mathematics. Read it aloud!

$$\sum_{k=1}^{2} f(k) = f(1) + f(2)$$

and

$$\sum_{k=1}^{n} 1 = 1 + 1 + \ldots + 1 = n,$$

where there are n 1's in the sum (so the summand doesn't actually have to change as k changes).

More generally $\sum_{k=\ell}^{u} a_k$ means $a_\ell + a_{\ell+1} + \cdots + a_{u-1} + a_u$, so

$$\sum_{j=-3}^{2} 2^j = \frac{1}{8} + \frac{1}{4} + \frac{1}{2} + 1 + 2 + 4.$$

If there happen to be no numbers in the range between ℓ and u inclusive, then the sum is called *empty*, and by convention is declared to be 0, so, for example,

$$\sum_{k=3}^{1} a_k = 0$$

(starting from 3 and working upwards along the number line, no numbers between 3 and 1 are encountered).

If "$\sum$" is replaced with "$\prod$", then we replace addition with multiplication, so

$$\prod_{i=1}^{5} i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.$$

The empty product is by convention declared to be equal to 1.

In summation notation, the statement of Claim 4.2 is

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

There are similar formulas for the sums of the first n squares, cubes, et cetera. The following are good exercises in proof by induction:

- $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$,
- $\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$.

Recursively defined sequences

Hand-in-glove with proof by induction goes definition by recursion. A sequence of numbers $(a_1, a_2, a_3, \ldots)$ is defined *recursively* if

- the values of the a_i for some small indices are specified, and

- for all other indices i, a procedure is given for calculating a_i, in terms of a_{i-1}, a_{i-2}, et cetera.

Properties of sequences defined recursively are often proved by induction, as we will now see.

The most famous example of a recursively defined sequence is the *Fibonacci numbers*. Define a sequence $(f_0, f_1, f_2, \ldots)$ by[58]

- $f_0 = 0$, $f_1 = 1$ and
- for $n \geq 2$, $f_n = f_{n-1} + f_{n-2}$.

The sequence begins $(0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots)$. Fibonacci numbers count many different things, for example:

- f_{n+1} is the number of ways of tiling a 1 by n strip with 1 by 1 and 1 by 2 tiles;
- f_{n+1} is the number of hopscotch boards that can be made using n squares[59];
- f_{n+1} in the number of ways of covering 2 by n strip with 2 by 1 dominoes;
- f_{n+2} is the number of words of length n that can be formed from the letters a and b, if two a's are not ever allowed to appear consecutively; and
- the Fibonacci numbers count the number of pairs of rabbit on an island after a certain amount of time has passed, under some very contrived conditions.[60]

The Fibonacci numbers exhibit many nice patterns. For example, define s_n to be the sum of all the Fibonacci numbers up to and including f_n, that is, $s_n = f_0 + f_1 + \cdots + f_n$, or $s_n = \sum_{k=0}^{n} f_k$. Here is a table of some values of s_n, compared to f_n:

n	0	1	2	3	3	5	6	7	8
f_n	0	1	1	2	3	5	8	13	21
s_n	0	1	2	4	7	12	20	33	54.

There seems to be a pattern: $s_n = f_{n+2} - 1$. We can prove this by induction on n. The base case $n = 0$ is clear, since $s_0 = 0 = 1 - 1 = f_2 - 1$. For the induction step, suppose that for some $n \geq 0$ we have $s_n = f_{n+2} - 1$. Then

$$\begin{aligned} s_{n+1} &= s_n + f_{n+1} \\ &= (f_{n+2} - 1) + f_{n+1} \quad \text{(inductive hypothesis)} \\ &= (f_{n+2} + f_{n+1}) - 1 \\ &= f_{n+3} - 1 \quad \text{(recursive definition of Fibonacci numbers)} \\ &= f_{(n+1)+2} - 1, \end{aligned}$$

[58] Notice that here I'm starting indexing at 0, rather than 1.

[59] See https://en.wikipedia.org/wiki/Hopscotch.

[60] The Fibonacci numbers are named for Leonardo of Pisa, nicknamed "Fibonacci", who discussed them in his book *Liber Abaci* in 1202, in the context of rabbits on an island. They had already been around for a while, though, having been studied by the Indian mathematician Pingala as early as 200BC.

so, by induction, the claimed identity is proven.

Other sum identities satisfied by the Fibonacci numbers include the following, that you can try to prove by induction:

- (Sum of odd-indexed Fibonacci numbers) $\sum_{k=0}^{n} f_{2k+1} = f_{2n+2}$;
- (Sum of even-indexed Fibonacci numbers) $\sum_{k=0}^{n} f_{2k} = f_{2n+1} - 1$; and
- (Sum of squares of Fibonacci numbers) $\sum_{k=0}^{n} f_k^2 = f_n f_{n+1}$ (hard!).

Many important mathematical operations are defined recursively. For example, although it is tempting simply to define a^n, for real a and natural number n, by

$$\text{“}a^n = a \cdot a \cdot \cdots \cdot a\text{”}$$

where there are n a's in the product on the right, this somewhat informal definition is an awkward one to use when trying to establish basic properties of powers. If instead (as we do) we define a^n recursively, via:

$$a^n = \begin{cases} a & \text{if } n = 1 \\ a \cdot a^{n-1} & \text{if } n \geq 2 \end{cases}$$

then proving all the expected properties becomes a fairly straightforward exercise in induction. For example, on the homework you will be asked to prove that for all natural numbers n, m, it holds that $a^{n+m} = (a^n)(a^m)$, and this should be done via induction.

We can also define $a^0 = 1$ for all non-zero a. We do not define 0^0[61].

Now that we've defined powers, it's possible to present another application of induction, the *Bernoulli inequality*. In the future (not this year) the content of the inequality will be quite useful; right now, it's just an example of an *inequality* proved inductively.

Claim 4.3. *For all $x \geq -1$ and all $n \in \mathbb{N}$, $(1+x)^n \geq 1 + nx$.*

Proof: We proceed by induction on n. We could if we wished start the induction at $n = 0$, where the assertion is that for all $x \geq -1$, $(1+x)^0 \geq 1 + 0 \cdot x$. This *seems* true enough: it's “$1 \geq 1$”. But, it's not always that, because at $x = -1$ we are required to interpret 0^0, which we have chosen not to do. So we'll start our induction (as the claim suggests) at $n = 1$, where the assertion is that for all $x \geq -1$, $(1+x)^1 \geq 1 + 1 \cdot x$, or $1 + x \geq 1 + x$, which is true not only for $x \geq -1$ but for all x.

[61] A very strong case can be made for $0^0 = 1$, because for natural numbers a and b, a^b counts the number of functions from a set of size b to a set of size a. When $b = 0$ and $a \neq 0$, there should be one function from the empty set to a set of size a, namely the “empty function” that does nothing, and this agrees with $a^0 = 1$ for $a \neq 0$; and when both a and b are 0, there is again one function from the empty set to itself, again the empty function, justifying setting 0^0 to be 1. If none of this makes sense, that's fine, as we haven't yet said what a function is. It might make more sense after we do.

We now move on to the induction step. Assuming $(1+x)^n \geq 1+nx$ holds for all $x \geq -1$, we consider how $(1+x)^{n+1}$ compares with $1+(n+1)x$ for $x \geq -1$. We have

$$\begin{aligned}
(1+x)^{n+1} &= (1+x)(1+x)^n \text{ by definition of powers} \\
&\geq (1+x)(1+nx) \text{ (by induction hypothesis)} \\
&= 1+(n+1)x+nx^2 \text{ (by some algebra)} \\
&\geq 1+(n+1)x \text{ (since } nx^2 \geq 0).
\end{aligned}$$

This proves the validity of the induction step, and so the claim is proved by induction. □

But wait ... where did we use $x \geq -1$ in the proof? The result is *false* without this assumption — for example, if $x = -4$ and $n = 3$, then $(1+x)^n = -27$ while $1+nx = -11$, so $(1+x)^n < 1+nx$. I'll leave it as an exercise to identify where the hypothesis got used.

4.2 A note on variants of induction

The principle of induction says that for $p(n)$ a predicate, with the universe of discourse for n being natural numbers, if $p(1)$ is true, and if, for arbitrary n, $p(n)$ implies $p(n+1)$, then $p(n)$ is true for all n. There are numerous natural variants, too numerous to possibly mention, and too similar to the basic principle for use to need to mention. I'll say a few here, so you can get the idea; looking at these examples you should realize that induction can be quite flexible. In all cases, $p(n)$ is a predicate with universe of discourse for n being natural numbers.

- If, for some natural number k, $p(k)$ is true, and if, for arbitrary $n \geq k$, $p(n)$ implies $p(n+1)$, then $p(n)$ is true for all $n \geq k$.
- If $p(0)$ is true, and if, for arbitrary $n \geq 0$, $p(n)$ implies $p(n+1)$, then $p(n)$ is true for all $n \geq 0$.
- If $p(-5)$ is true, and if, for arbitrary $n \geq -5$, $p(n)$ implies $p(n+1)$, then $p(n)$ is true for all $n \geq -5$.
- If $p(2)$ is true, and if, for arbitrary $n \geq 2$, $p(n)$ implies $p(n+2)$, then $p(n)$ is true for all positive even numbers.
- $\cdots$.

4.3 Binomial coefficients and the binomial theorem

We all know that $(x+y)^2$ expands out to $x^2+2xy+y^2$. What about $(x+y)^3$, $(x+y)^4$, et cetera? Here is a table showing the various expansions of $(x+y)^n$ for some small values of n.

For completeness I've included $n = 0$ ($(x+y)^0 = 1$ as long as $x + y \neq 0$) and $(x+y)^1 = x + y$.

$$\begin{array}{rcc}
(x+y)^0 & = & 1 \\
(x+y)^1 & = & x+y \\
(x+y)^2 & = & x^2 + 2xy + y^2 \\
(x+y)^3 & = & x^3 + 3x^2y + 3xy^2 + y^3 \\
(x+y)^4 & = & x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\
(x+y)^5 & = & x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5 \\
(x+y)^6 & = & x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6 \\
\cdots & & \cdots .
\end{array}$$

There is a pattern here. Of course, $(x+y)^n$, when expanded out, has some x^n terms, some $x^{n-1}y$ terms, some $x^{n-2}y^2$ terms, and so on, down to some xy^{n-1} and some y^n terms, but by "a pattern" I mean a pattern in the coefficients of each of these terms. That pattern is best spotted when the table is re-written as a triangle of coefficients, without the terms $x^n y^{n-k}$ that are just cluttering up the picture:

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

Notice that there are 1's down the outside, and all other numbers are obtained by summing their neighbor above and to right and above and to the left. If this pattern were to continue, the next row would be

$$1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1,$$

and sure enough, after a lot of computation, we discover that

$$(x+y)^7 = x^7 + 7x^6y + 21x^5y^2 + 35x^4y^3 + 35x^3y^4 + 21x^2y^5 + 7xy^6 + y^7.$$

We'll prove, by induction, that this pattern holds for all n. But first we have to be a little more precise about what the pattern is. Define the number "$n!$" (read this as "n factorial") recursively, via:

$$n! = \begin{cases} 1 & \text{if } n = 1 \\ n \cdot (n-1)! & \text{if } n \geq 2 \\ 1 & \text{if } n = 0. \end{cases}$$

So, informally, $n!$ is the product of all the whole numbers from n down to 1:

- $3! = 3 \cdot 2! = 3 \cdot 2 \cdot 1! = 3 \cdot 2 \cdot 1 = 6$,

- $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, et cetera.

We can interpret $n!$ as the number of ways of arranging n objects in order — there are n options for the first object; for each of those n options there are a further $n-1$ options for the second, so $n \cdot (n-1)$ options for the first two objects; for each of those $n(n-1)$ options there are a further $n-2$ options for the third, so $n(n-1)(n-2)$ options for the first three objects, and so on down to $n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1 = n!$ options for ordering all n objects.

That 0! is defined to be "1" may seem a little strange, as it is hard to imagine counting the number of ways of arranging 0 objects in order; but as we will see there are sound reasons for this, and also it adheres to our convention that an empty product is 1 (the product of all the numbers, starting at 0, and going *down* to 1, is empty).

Next define the expression "$\binom{n}{k}$", (read this as "n choose k") via

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

Notice that because of the way we have defined the factorial function (never 0, and making sense exactly for the natural numbers including 0) this expression is well-defined for all $n \geq 0$, and for all k satisfying $0 \leq k \leq n$. That means we can form a triangle from the numbers $\binom{n}{k}$:

$$\begin{array}{ccccccccccccc}
 & & & & & & \binom{0}{0} & & & & & & \\
 & & & & & \binom{1}{0} & & \binom{1}{1} & & & & & \\
 & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & \\
 & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & \\
 & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & \\
 & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & \\
\binom{6}{0} & & \binom{6}{1} & & \binom{6}{2} & & \binom{6}{3} & & \binom{6}{4} & & \binom{6}{5} & & \binom{6}{6}.
\end{array}$$

Calculating the numerical values, this triangle becomes:

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

which looks just like the triangle of coefficients we got when we expanded out $(x+y)^n$.

In that case (coefficients in expansion of $(x+y)^n$) we *suspected* that each number in the triangle, other than the 1's on the outside, was obtained by summing its two neighbors above it. In this case (triangle of $\binom{n}{k}$), we can easily *prove* this relation.

Claim 4.4. *For $n \geq 0$, and for $0 \leq k \leq n$:*

- *if* $k = 0$ *or if* $k = n$ *then* $\binom{n}{k} = 1$*;*
- *otherwise,*

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

Proof: If $k = 0$ then since $0! = 1$, we have

$$\binom{n}{k} = \frac{n!}{0!(n-0)!} = \frac{n!}{1 \cdot n!} = 1,$$

and if $n = k$ for a similar reason we have $\binom{n}{k} = 1$.[62]

Otherwise, we must have $n \geq 2$ and $1 < k < n$. We have

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

(Notice that all steps above involve expressions that make sense, because $n \geq 2$ and $1 < k < n$). □

Just as there was a counting interpretation of $n!$, there's a counting interpretation of $\binom{n}{k}$. How many subsets of size k does a set of size n have? Well, we can select such a subset by choosing a first element, then a second, et cetera, leading to a count of $n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = n!/(n-k)!$; but each particular subset has been counted many times. In fact, a particular subset has been counted $k!$ times, once for each of the $k!$ ways in ways in which its k elements can be arranged in order. So our count of $n!/(n-k)!$ was off by a multiplicative factor of $k!$, and the correct count is $(n!/(n-k)!)/k!$, which is exactly $\binom{n}{k}$. So:

> $\binom{n}{k}$ is the number of subsets of size k of a set of size n.

This allows a alternate proof of Claim 4.4. When $k = n$, $\binom{n}{k}$ is the number of subsets of size n of a set of size n, and this is clearly 1 (the set itself). When $k = 0$, $\binom{n}{k}$ is the number of subsets of size 0 of a set of size n, and this is also 1 (the empty set is a subset of any set, and it is the only set with 0 elements). For $n \geq 2$, and $1 < k < n$, subsets of size k of a set X of size n fall into two classes:

[62] This is a strong justification for declaring $0! = 1$.

- those that include a particular fixed element x — there are $\binom{n-1}{k-1}$ of these, one for each subset of $X - \{n\}$ of size $k-1$, and
- those that *don't* include x — there are $\binom{n-1}{k}$ of these, one for each subset of $X - \{n\}$ of size k.

So X has $\binom{n-1}{k-1} + \binom{n-1}{k}$ subsets of size k; but it also (directly) has $\binom{n}{k}$ subsets of size k; so

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

This identity is called *Pascal's identity*[63]

We're now ready to formalize a theorem that captures the pattern we were noticing with $(x+y)^n$. It's called the *binomial theorem* (because the expansion of $(x+y)^n$ is a *binomial expansion* — an expansion of an expression involving two (*bi*) named (*nomial*) things, x and y), and the numbers $\binom{n}{k}$ that come up in it are often called *binomial coefficients*.

Theorem 4.5. *Except in the case when $n = 0$ and at least one of $x, y, x+y = 0$, for all $n \in \mathbb{N}^0$ and for all real x, y,*

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{k}x^{n-k}y^k + \cdots + \binom{n}{n-1}xy^{n-1} + y^n,$$

or, more succinctly,

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

Proof: When $n = 0$, as long as all of $x, y, x+y$ are non-zero both sides of the identity are 1, so they are equal.

For $n \geq 1$ we proceed by induction on n (with predicate:

$$p(n) : \text{“for all real } x, y,\ (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k\text{”}).$$

The base case $p(1)$ asserts $(x+y)^1 = \binom{1}{0}x + \binom{1}{1}y$, or $x + y = x + y$, which is true for all real x, y.

For the induction step, we assume that for some $n \geq 1$ we have

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-2}x^2y^{n-2} + \binom{n}{n-1}xy^{n-1} + y^n$$

for all real x, y. Multiplying both sides by $x+y$, this yields

$$(x+y)^{n+1} = (x+y)\left(x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-2}x^2y^{n-2} + \binom{n}{n-1}xy^{n-1} + y^n\right).$$

[63] It's named for the French polymath Blaise Pascal. The triangle of values of $\binom{n}{k}$ is called *Pascal's triangle*, and has many lovely properties. It is easily googled.

Now the right-hand side above is

$$\begin{gathered} x^{n+1}+ \\ \tbinom{n}{1}x^n y + \tbinom{n}{2}x^{n-1}y^2 + \cdots + \tbinom{n}{n-2}x^3y^{n-2} + \tbinom{n}{n-1}x^2y^{n-1} + xy^n + \\ x^n y + \tbinom{n}{1}x^{n-1}y^2 + \cdots + \tbinom{n}{n-3}x^3y^{n-2} + \tbinom{n}{n-2}x^2y^{n-1} + \tbinom{n}{n-1}xy^n + \\ y^{n+1}. \end{gathered}$$

or

$$\begin{gathered} x^{n+1}+ \\ \tbinom{n}{1}x^n y + \tbinom{n}{2}x^{n-1}y^2 + \cdots + \tbinom{n}{n-2}x^3y^{n-2} + \tbinom{n}{n-1}x^2y^{n-1} + \tbinom{n}{n}xy^n + \\ \tbinom{n}{0}x^n y + \tbinom{n}{1}x^{n-1}y^2 + \cdots + \tbinom{n}{n-3}x^3y^{n-2} + \tbinom{n}{n-2}x^2y^{n-1} + \tbinom{n}{n-1}xy^n + \\ y^{n+1}. \end{gathered}$$

Applying Claim 4.4 to each pair of terms in matching columns in the second and third rows, this becomes

$$\begin{gathered} x^{n+1}+ \\ \tbinom{n+1}{1}x^n y + \tbinom{n+1}{2}x^{n-1}y^2 + \cdots + \tbinom{n}{n-2}x^3y^{n-2} + \tbinom{n}{n-1}x^2y^{n-1} + \tbinom{n+1}{n}xy^n + \\ y^{n+1} \end{gathered}$$

(for example,

$$\binom{n}{2}x^{n-1}y^2 + \binom{n}{1}x^{n-1}y^2 = \left(\binom{n}{2} + \binom{n}{1}\right)x^{n-1}y^2 = \binom{n+1}{2}x^{n-1}y^2)$$

Using $\binom{n+1}{0} = \binom{n+1}{n+1} = 1$, this last expression is exactly

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

So we have shown that $(1+x)^{n+1} = \sum_{k=0}^{n+1}\binom{n+1}{k}x^{(n+1)-k}y^k$ for all real x, y, which is $p(n+1)$. The induction step is complete, as is the proof of the theorem. □

At the end of Spivak Chapter 2, there are plenty of exercises that explore the many properties of the numbers $\binom{n}{k}$.

4.4 Complete, or strong, induction (informally)

Sometimes induction is not enough to verify a proposition that at first glance seems tailor-made for induction. For example, consider the recursively defined sequence

$$a_n = \begin{cases} 2 & \text{if } n = 0 \\ 3 & \text{if } n = 1 \\ 3a_{n-1} - 2a_{n-2} & \text{if } n \geq 2. \end{cases}$$

This table shows the first few values of a_n:

n	0	1	2	3	4	5	6	7
a_n	2	3	5	9	17	33	65	129.

There seems to be a pattern: it seems that $a_n = 2^n + 1$ for each n. If we try to prove this by induction, though, we run into a number of problems. The base case $n = 0$ is evident. The first problem arises when we think about the induction step: we assume, for some arbitrary $n \geq 0$, that $a_n = 2^n + 1$, and try to deduce that $a_{n+1} = 2^{n+1} + 1$.

Our inclination is to use the recursive definition $a_{n+1} = 3a_n - 2a_{n-1}$. But already in the very first instance of the induction step, we are stuck, because at $n = 0$ the recursive definition we would like to use is $a_1 = 3a_0 - 2a_{-1}$. This makes no sense (there is no a_{-1}). And indeed, it shouldn't make sense, because the clause $a_n = 3a_{n-1} - 2a_{n-2}$ of the definition of a_n kicks in only when $n \geq 2$. To say anything about a_1, we have to appeal to a different clause in the definition, namely $a_1 = 3$. Since $3 = 2^1 + 1$, this is still consistent with the general pattern we are trying to prove.

One way to think of this is that we are verifying *two* base cases ($n = 0$ *and* $a = 1$) before going on to the induction step; another way to think of it is that we are treating the induction step "$p(0) \Rightarrow p(1)$" as a special case, and showing that it is a true implication by showing that both $p(0)$ and $p(1)$ are simply true, always, so the implication is true; the remainder of the induction step, "$p(n) \Rightarrow p(n + 1)$ for every $n \geq 1$" will be dealt with in a different, more general, way. However we choose to think of it, this issue arises frequently in proofs by induction, especially when dealing with recursively defined sequences.

Having dealt with the first instance of the induction step, lets move on to the general inductive step, $p(n) \Rightarrow p(n+1)$ for $n \geq 1$. Here we can legitimately write $a_{n+1} = 3a_n - 2a_{n-1}$, because for $n \geq 1$, this is the correct clause for defining a_{n+1}. We would like to say that

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

using that $a_n = 2^n + 1$. But we can't: the best we can say is

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

because in trying to verify $p(n) \Rightarrow p(n + 1)$ we can assume nothing about $p(n - 1)$.

There's a fix: presumable, in getting this far in the induction, we have already established not just $p(n)$, but also $p(n - 1), p(n - 2), p(n - 3)$, et cetera. If we have, then we can, as well as using $a_n = 2^n + 1$, use $a_{n-1} = 2^{n-1} + 1$. Then we get

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

as we need to show to establish $p(n + 1)$.

We can formalize this idea in the *principle of complete induction*, also called the *principle of strong induction*:

> **The principle of complete mathematical induction**: Let $p(n)$ be a predicate, with the universe of discourse for n being natural numbers. If $p(1)$ is true, and if, for arbitrary n, the conjunction of $p(1), p(2), \ldots, p(n)$ implies $p(n + 1)$, then $p(n)$ is true for all n.

Going back through the discussion that we gave to justify the principle of induction, it should be clear that complete or strong induction is an equally valid proof technique. We can in fact argue that strong induction is *exactly* as strong as regular induction:

- Suppose that we have access to the principle of strong induction. Suppose that $p(n)$ is a predicate (with n a natural number) and that we know
  - $p(1)$ and
  - for arbitrary $n \geq 1$, $p(n)$ implies $p(n+1)$.

  Then we *also* know $p(1) \wedge p(2) \wedge \cdots \wedge p(n)$ implies $p(n+1)$ (if we can infer $p(n+1)$ from $p(n)$, we can certainly infer it from $p(1), p(2), \ldots, p(n)$!). So by strong induction, we can conclude that $p(n)$ is true for all n. In other words, if we have access to the principle of strong induction, we also have access to the principle of induction.

- Suppose that we have access to the principle of induction. Suppose that $p(n)$ is a predicate (with n a natural number) and that we know
  - $p(1)$ and
  - for arbitrary $n \geq 1$, $p(1) \wedge p(2) \wedge \cdots \wedge p(n)$ implies $p(n+1)$.

  We would like to conclude that $p(n)$ is true for all n; but we can't simple say that $p(n)$ implies $p(n+1)$, and use induction; we don't know whether $p(n)$ (on its own) implies $p(n+1)$. Here's a fix: consider the predicate $Q(n)$ define by

  $$Q(n) : \text{“}p(1) \wedge p(2) \wedge \cdots \wedge p(n).\text{”}$$

  We know $Q(1)$ (it's just $p(1)$. Suppose, for some arbitrary n, we know $Q(n)$. Then we know $p(1) \wedge p(2) \wedge \cdots \wedge p(n)$, and we can deduce $p(n+1)$. But, again since we know $p(1) \wedge p(2) \wedge \cdots \wedge p(n)$, we can now deduce $p(1) \wedge p(2) \wedge \cdots \wedge p(n) \wedge p(n+1)$, that is, we can deduce $Q(n+1)$. So we can apply induction to Q to conclude $Q(n)$ for all n. But a consequence of this is that $p(n)$ holds for all n (remember, $Q(n)$ is $p(1) \wedge p(2) \wedge \cdots \wedge p(n)$). In other words, if we have access to the principle of induction, we also have access to the principle of strong induction.

Here's an important application of complete induction, from elementary number theory. A natural number $n \geq 2$ is said to be *composite* if there are natural numbers a and b, both at least 2, such that $ab = n$. It is said to be *prime* if it is not composite. We can use strong (complete) induction to show that every natural number $n \geq 2$ can be written as a product of prime numbers.[64]

[64]The *fundamental theorem of arithmetic* states that the prime factorization of any number is *unique* up to the order in which the primes in the factorization are listed (note that this would not be true if 1 was considered a prime number, for then 3.2.1 and 3.2.1.1.1 would be different prime factorizations of 6). The fundamental theorem of arithmetic is also proven by induction, but takes a lot more work than the result we are about to prove, establishing the existence of *a* prime factorization.

Indeed, let $p(n)$ be the predicate "n can be written as the product of prime numbers". We prove that $p(n)$ is true for all $n \geq 2$ by complete induction.

Base case $n = 2$: This is trivial since 2 is a prime number.

Inductive step: Suppose that for some $n \geq 3$, we know that $p(m)$ is True for all m in the range $2 \leq m \leq n-1$[65]. We consider two cases.

- Case 1: n is prime. In this case $p(n)$ is trivial.
- Case 2: n is composite. In this case $n = ab$ for some natural numbers a and b with $2 \leq a \leq n-1$ and $2 \leq b \leq n-1$. Since $p(a)$ and $p(b)$ are both true (by the complete induction hypothesis) we have

  $$a = p_1 p_2 \cdots p_k$$

  and

  $$b = q_1 q_2 \cdots q_\ell$$

  where $p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_\ell$ are all prime numbers. But that implies that n can be written as a product of prime numbers, via

  $$n = ab = p_1 p_2 \cdots p_k q_1 q_2 \cdots q_\ell.$$

  This shows that $p(n)$ follows from $p(2), p(3), \ldots, p(n-1)$.

By complete induction, we conclude that $p(n)$ is true for all $n \geq 2$.

Note that this proof would have gone exactly *nowhere* if all we were able to assume, when trying to factorize n, was the existence of a factorization of $n-1$.

We now give a more substantial example of complete induction. The associativity axiom for multiplication says that for all reals a, b, c, we have $a(bc) = (ab)c$ (note that I'm using juxtaposition for multiplication here, as is conventional, rather than the "$\cdot$" that I've been using up to now). Presumably, there is an "associativity axiom" for the product of n things, too, for all $n \geq 3$ (we've already seen the version for $n = 4$). Let GAA(n) be the predicate "for any set of n real numbers $a_1, \ldots, a_n$ the order in which the product $a_1 \cdots a_n$ is parenthesized does not affect the final answer", and let GAA be the generalized associativity axiom, that is, the statement that GAA(n) holds for all $n \geq 1$[66].

Claim 4.6. GAA *is true.*

Proof: Among all the ways of parenthesizing the product $a_1 \cdots a_n$ we identify one special one, the *right-multiply*:

$$R(a_1, \ldots, a_n) = (\cdots(((a_1 a_2)a_3)a_4)\cdots)a_n.$$

[65] Note that when proving things by induction, you can either deduce $p(n+1)$ from $p(n)$, or deduce $p(n)$ from $p(n-1)$; similarly, when proving things by strong induction you can either deduce $p(n+1)$ from $p(1) \wedge \cdots \wedge p(n)$, or deduce $p(n)$ from $p(1) \wedge \cdots \wedge p(n-1)$; it's a matter of taste or convenience

[66] Really the result is only interesting for $n \geq 3$, but it makes sense, and is true, for $n = 1, 2$ as well, so we'll throw those into the mix, too.

We will prove, by strong induction on n, that for all $n \geq 1$, the predicate "for any set of n real numbers $a_1, \ldots, a_n$, all the ways of parenthesizing the product $a_1 \cdots a_n$ lead to the answer $R(a_1, \ldots, a_n)$." This will show that GAA is true.

The base case $n = 1$ is trivial — with only one number, there is one one possible product. The same goes for the base case $n = 2$. The base case $n = 3$ is axiom P5.

For the inductive step, let $n \geq 4$ be arbitrary, and suppose that the predicate we are trying to prove (GAA(k)) is true for all values k of the variable between 1 and $n - 1$. Let P be an arbitrary parenthesizing of the product $a_1 \cdots a_n$. P has a final, outer, product, the last pair of numbers multiplied together before P is fully evaluated. We consider cases.

Case 1 The final product is of the form Aa_n. By induction (variable value $n - 1$) we have $A = R(a_1, \ldots, a_{n-1})$, so

$$P = Aa_n = R(a_1, \ldots, a_{n-1})a_n = R(a_1, \ldots, a_n).$$

Case 2 The final product is of the form AB where A is a parenthesizing of $a_1, \ldots, a_k$ and B is a parenthesizing of $a_{k+1}, \ldots, a_n$, where $1 \leq k \leq n - 2$. If $k = n - 2$ then we have

$$P = A(a_{n-1}a_n) = (Aa_{n-1})a_n$$

(by P5), and we are back in case 1, so $P = R(a_1, \ldots, a_n)$. If $k \leq n-3$ then by induction (variable value $n - k$) we have

$$B = R(a_{k+1} \cdots a_n) = R(a_{k+1} \cdots a_{n-1})a_n$$

and so, once again by P5,

$$P = AB = A(R(a_{k+1} \cdots a_{n-1})a_n) = (AR(a_{k+1} \cdots a_{n-1}))a_n,$$

and we are back in case 1, so $P = R(a_1, \ldots, a_n)$.[67]

[67]Here's an alternate way of presenting the two cases:
Case 1: $P =$ (SOMETHING)a_n. By induction (GAA($n-1$)), this is the same as

$$P = R(a_1, \ldots, a_{n-1})a_n = R(a_1, \ldots, a_n).$$

Case 2:

$$\begin{aligned}
P &= \underbrace{\text{(SOMETHING)}}_{\text{involving } x_1, \ldots, x_k,\ 1 \leq k \leq n-2} \cdot \underbrace{\text{(SOMETHING ELSE)}}_{\text{involving } x_{k+1}, \ldots, x_n} \\
&= \underbrace{R(a_1, \ldots, a_k)}_{\text{GAA}(k)} \underbrace{R(a_{k+1}, \ldots, a_n)}_{\text{GAA}(n-k)} \\
&= R(a_1, \ldots, a_k) \cdot (R(a_{k+1}, \ldots, a_{n-1}) \cdot a_n) \\
&= (R(a_1, \ldots, a_k)R(a_{k+1}, \ldots, a_{n-1}) \cdot) \cdot a_n \quad \text{(GAA(3))} \\
&= R(a_1, \ldots, a_{n-1})a_n \quad \text{(GAA}(n-1)\text{)} \\
&= R(a_1, \ldots, a_n).
\end{aligned}$$

In either case, $P = R(a_1, \ldots, a_n)$, and so the claim is proven by (strong) induction. $\square$

Notice that we needed the induction hypothesis for *all* values of the variable below n, so we really needed strong induction.

Of course, there is also an analogous generalized associativity for addition. Strong induction is in general a good way to extend arithmetic identities from a few terms to arbitrarily many terms. You should do some of the following as exercises:

Generalized commutativity For $n \geq 2$, for any set of n reals, the result of adding the n reals does not depend on the order in which the numbers are written down; and the same for multiplication.

Generalized distributivity For $n \geq 2$, and for any set of real numbers $a, b_1, b_2, \ldots, b_n$,

$$a(b_1 + \cdots b_n) = ab_1 + \cdots ab_n.$$

Generalized triangle inequality For $n \geq 2$, and for any set of real numbers $b_1, b_2, \ldots, b_n$,

$$|b_1 + \ldots b_n| \leq |b_1| + \cdots + |b_n|.$$

Generalized Euclid's rule For $n \geq 2$, and for any set of real numbers $b_1, b_2, \ldots, b_n$, if $b_1 b_2 \cdots b_n = 0$ then at least one of $b_1, b_2, \ldots, b_n$ must be 0.

4.5 The well-ordering principle (informal)

A set S has a *least element* if there is an element s *in the set* S with $s \leq s'$ for every $s' \in S$. Not every set has a least element: there is no least positive number (for every positive number p, $p/2$ is a smaller positive number), and there is no least negative number (for every negative number q, $q - 1$ is a smaller negative number).

The set of natural numbers, on the other hand (at least as we have informally defined it), has a least element element, namely 1. Moreover, it seems intuitively clear that every subset of $\mathbb{N}$ has a least element; or rather, every *non-empty* subset of $\mathbb{N}$ has a least element (the empty set has no least element). We formulate this as the *well-ordering principle* of the natural numbers:

Claim 4.7. *(The well-ordering principle of the natural numbers) If E is a non-empty subset of the natural numbers, then E has a least element.*

Proof: Suppose E is a subset of the natural numbers with no least element. We will show that E is empty; this is the contrapositive of, and equivalent to, the claimed statement.

Let $p(n)$ be the predicate "$n \notin E$". We will show, by strong induction, that $p(n)$ is true for all n, which will show that E is empty.

The base case $p(1)$ asserts $1 \notin E$, which is true; if $1 \in E$ then 1 would be the least element in E.

For the induction step, assume that $p(1), \ldots p(n-1)$ are all true, for some arbitrary natural number $n \geq 2$. Then none of $1, 2, \ldots, n-1$ are in E, so neither is n, since if $n \in E$ then would be the least element in E. So $p(n)$ is true, assuming $p(1), \ldots p(n-1)$ are all true, and by strong induction $p(n)$ is true for all n. $\square$

As an application of well-ordering, we give an alternate proof of the irrationality of $\sqrt{2}$.

Suppose (for a contradiction) $\sqrt{2}$ is rational. Let E be set of all natural numbers x such that $x^2 = 2y^2$ for some natural number y. Under the assumption that $\sqrt{2}$ is rational, E is non-empty, and so by well-ordering it has a least element, a say, with $a^2 = 2b^2$ for some natural number b.

Now it is an easy check that $b < a < 2b$ (indeed, since $a^2 = 2b^2$ it follows that $b^2 < a^2 < 4b^2$, from which $b < a < 2b$, via a homework problem).

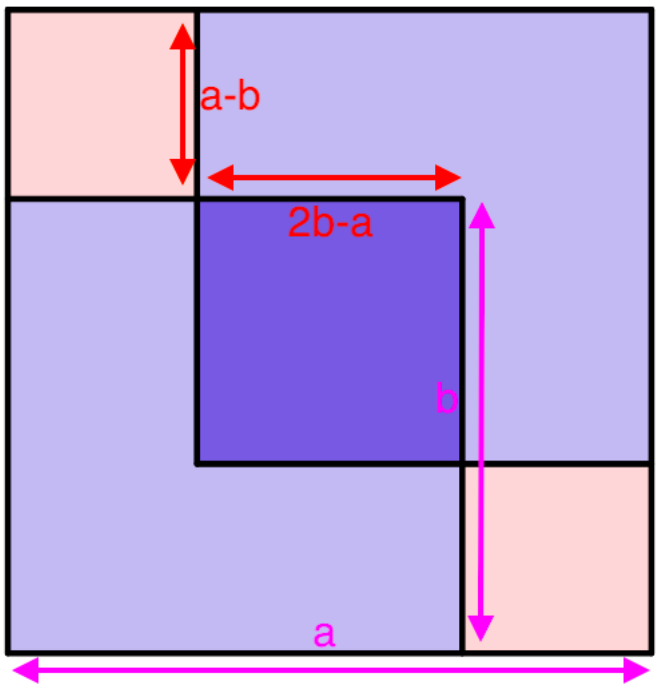
Set $a' = 2b - a$ and $b' = a - b$. By the relations $b < a < 2b$, both natural numbers, and since $b < a$ we have $a' < a$. But now note that

$$2(b')^2 = 2(a-b)^2 = 2a^2 - 4ab + 2b^2 = a^2 - 4ab + 4b^2 = (2b-a)^2 = (a')^2,$$

so $a' \in E$, contradicting that a is smallest element of E.

We conclude that E' is empty, so $\sqrt{2}$ is irrational.

This lovely proof was discovered by Stanley Tennenbaum; here is a visual illustration of it, that I have taken (and augmented) from https://divisbyzero.com/2009/10/06/tennenbaums-proof-of-the-irrationality-of-the-square-root-of-2/.



A visual illustration of Tennenbaum's proof of the irrationality of $\sqrt{2}$

4.6 Inductive sets

The purpose of the rest of this section is to make the "..." in

$$\mathbb{N} = \{1, 1+1, 1+1+1, \ldots\},$$

and the principle of mathematical induction, a little more formal.

Say that a set $S \subseteq X$ is *inductive* if it satisfies both of these properties:

1. 1 is in S and
2. $k+1$ is in S whenever k is in S.

So, for example:

- X is inductive.
- The set of positive numbers in X is inductive.
- The set of positive numbers excluding 5 is *not* inductive; it fails the second condition, since 4 is in S but not 5.
- The set of positive numbers, excluding $3/2$ is *not* inductive; it fails the second condition, since $1/2$ is in S but not $3/2$.
- The set of positive numbers that are at least 1, excluding $3/2$ is inductive; the absence of $3/2$ is not an obstacle, since $1/2$ is not in S, so the implication "If $1/2$ is in S then $3/2$ is in S" is true.
- The set of positive numbers that are greater than 1 is *not* inductive; it fails the first condition.
- If S_1 and S_2 are two inductive sets, then the set of elements that are in both S_1 and S_2 is also inductive.

It feels like the set $\{1, 1+1, 1+1+1, \ldots\}$ should be in *every* inductive set, because 1 is in every inductive set, so $1+1$ is also, and so on. To formalize that "and so on", we make the following definition.

Definition 4.8. A number n is a *natural number* if it is in every inductive set. We denote by $\mathbb{N}$ the set of all natural numbers.

So, for example, 1 is a natural number (because it is in every inductive set), and so is $1+1$, and so is $1+1+\ldots+1$ where there are 1876 1's in the sum. More generally if k is in $\mathbb{N}$ then k is in every inductive set, so (by definition of inductive sets) $k+1$ is in every inductive set, so $k+1$ is in $\mathbb{N}$. In other words, $\mathbb{N}$ is an inductive set itself.

By its definition, $\mathbb{N}$ is contained in every inductive set. Moreover, it is the only inductive set that is contained in every inductive set. To see this, consider an inductive set E that is contained in every inductive set. Since $\mathbb{N}$ is inductive, we have that E is contained in $\mathbb{N}$. Suppose that E is not equal to $\mathbb{N}$. Then there is some number k with k in $\mathbb{N}$ but k not in E. But if k is in $\mathbb{N}$ then by the definition of $\mathbb{N}$ we have that k is E, since being in $\mathbb{N}$ means being in *every* inductive set, including E. The contradiction — k is not in E and k is in E — shows that E is not equal to $\mathbb{N}$ is False, and so we conclude $E = \mathbb{N}$. We summarize what we have just proven in a claim.

Claim 4.9. *The natural numbers form an inductive set, and* $\mathbb{N}$ *is the* unique *minimal inductive set — it is contained in every inductive set, and no other inductive set has this property. In particular if* E *is a subset of* $\mathbb{N}$ *and* E *is inductive then* $E = \mathbb{N}$.

4.7 The principle of mathematical induction

Re-phrasing the last sentence of Claim 4.9 we obtain the important *principle of mathematical induction.*

Theorem 4.10. *Suppose that* E *is a set of natural numbers satisfying*

1. 1 *is in* E *and*
2. $k+1$ *is in* E *whenever* k *is.*

Then $E = \mathbb{N}$.

There is no need for a proof of this — it really is just a direct re-phrasing of the last sentence of Claim 4.9. To get a first hint of the power of Theorem 4.10 we use it to derive the following result, which is precisely the form of induction that we are by now familiar with.

Theorem 4.11. *Suppose that* $p(n)$ *is a predicate (a statement that is either True or False, depending on the value of* n*), where the universe of discourse for the variable* n *is all natural numbers. If*

- $p(1)$ *is true and*
- $p(k+1)$ *is true whenever* $p(k)$ *is true*

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

Proof: Let E be the set of all n for which $p(n)$ is True. We immediate have that 1 is in E and that $k+1$ is in E whenever k is. That $E = \mathbb{N}$, that is that $p(n)$ is True for all n in $\mathbb{N}$, now follows from Theorem 4.10. □

Slightly informally Theorem 4.11 says that if $p(n)$ is some proposition about natural numbers, and if we can show that

Base case $p(1)$ is True and

Induction step for all n the truth of $p(n)$ (the **induction hypothesis**) implies the truth of $p(n+1)$

then we can conclude that $p(n)$ is True for all natural numbers. The power here, that you should see from some examples, is that the principle of mathematical induction allows us to prove *infinitely many things* ($p(1)$, $p(2)$, $p(3)$, et cetera), with only a *finite amount of work* (proving $p(1)$ and proving the single implication $p(n) \Rightarrow p(n+1)$, involving a variable).

More informally still, induction says (repeating a previous observation) that if you can get onto the first rung of a ladder ($p(1)$), and you know how to climb from any one rung to any other ($p(n) \Rightarrow p(n+1)$), then you can climb as high up the ladder as you wish, by first getting on the ladder and then moving up as many rungs as you wish, one rung at a time.

We've already seen many examples of induction at work, in the informal setting, and of course all of those examples go through perfectly in the more formal setting we've given here. We give a few more examples of induction at work now, mostly to establish some very fundamental properties of the natural numbers, that will be useful later. You should get the sense that every property of numbers that you are already familiar with can be established formally in the context of the definition of natural numbers that we have given.

Claim 4.12. *For all natural numbers n, $n \geq 1$.*

Proof: Let $p(n)$ be the predicate "$n \geq 1$", where the universe of discourse for the variable n is all natural numbers. We prove that $p(n)$ is true for all n by induction.

Base case: $p(1)$ is the assertion $1 \geq 1$, which is true.

Induction step: Assume that for some n, $n \geq 1$. Then $n + 1 \geq 1 + 1 \geq 1 + 0 = 1$. So the truth of $p(n)$ implies the truth of $p(n+1)$.

By induction, $p(n)$ is true for all n, that is, for all natural numbers n, $n \geq 1$. □

Corollary 4.13. *There is no natural number x with $0 < x < 1$.*

Proof: Such an x would be a natural number that does not satisfy $x \geq 1$, contradicting Claim 4.12. □

Claim 4.14. *For every natural number n other than 1, $n - 1$ is a natural number.*

Proof: Let $p(n)$ be the predicate "$(n \neq 1) \implies (n - 1 \in \mathbb{N})$". We prove $(\forall n)p(n)$ by induction (with, as usual, the universe of discourse being $\mathbb{N}$).

Base case: $p(1)$ is the assertion $(1 \neq 1) \implies (1 - 1 \in \mathbb{N})$, which is true, since the premise $1 \neq 1$ is false.

Induction step: Assume that for some n, $(n \neq 1) \implies (n - 1 \in \mathbb{N})$. Then $n + 1 \neq 1$, for if $n + 1 = 1$ then $n = 0$, which is not a natural number. Also, $(n + 1) - 1 = n$, which is a

natural number. So both the premise and the hypothesis of $p(n+1)$ is are true, so $p(n+1)$ is true

By induction, $p(n)$ is true for all n, that is, for all natural numbers n, if $n \neq 1$ then $n-1 \in \mathbb{N}$. $\square$

Corollary 4.15. *There is no natural number x with $1 < x < 2$.*

Proof: Such an x would be a natural number other than 1, so $x - 1 \in \mathbb{N}$ by Claim 4.14. But $0 < x - 1 < 1$, contradicting Corollary 4.13. $\square$

All of these results have been leading up to the following, an "obvious" statement that requires a (somewhat sophisticated) proof. It captures in very concrete way that the natural numbers are indeed a set of the form $\{1, 2, 3, \ldots\}$.

Claim 4.16. *For every natural number n, there is no natural number x with $n < x < n + 1$.*

Proof: We proceed by induction on n, with the base case $n = 1$ being Claim 4.15. For the induction step, suppose that for some n there is no natural number x with $n < x < n+1$, but there is a natural number y with $n + 1 < y < n + 2$. Since $n \neq 0$ we have $y \neq 1$ so $y - 1 \in \mathbb{N}$, and since $n < y - 1 < n + 1$ this contradicts the induction hypothesis. We conclude that there is no such y, and so by induction the claim is true. $\square$

4.8 The principle of complete, or strong, induction

Sometimes it is helpful in an induction argument to be able to assume not just $p(n)$ when trying to prove $p(n + 1)$, but instead to assume $p(k)$ for all $k \leq n$. Here are the two forms of the method of *strong* or *complete* induction that this leads to.

Theorem 4.17. *Suppose that E is a set of natural numbers satisfying*

1. *1 is in E and*
2. *$k + 1$ is in E whenever every j with $j \leq k$ is.*

Then $E = \mathbb{N}$.

Theorem 4.18. *Suppose that $p(n)$ is a predicate with universe of discourse for n being all natural numbers. If*

- *$p(1)$ is True and*
- *$p(k + 1)$ is True whenever $p(j)$ is True for all $j \leq k$*

then $p(n)$ is True for all n in $\mathbb{N}$.

As we have observed earlier, complete induction (Theorem 4.18) and ordinary induction (Theorem 4.11) are equivalent, in the sense that any proof that can be carried out using one can be transformed into a proof that use the other. We repeat the justification of this claim here, in slightly different language.

Suppose we have a proof of the truth of some predicate $p(n)$ for all natural numbers n, that uses ordinary induction. Then the argument used to deduce the truth of $p(k+1)$ from that of $p(k)$, is exactly an argument that deduces the truth of $p(k+1)$ from the truth $p(j)$ for all $j \leq k$ (just one that never needs to use any of the hypotheses of the implication except $p(k)$). So any prove using ordinary induction can be transformed into one using complete induction, somewhat trivially.

On the other hand, suppose we have a proof of the truth of some predicate $p(n)$ for all natural numbers n, that uses complete induction. Let $q(n)$ be the predicate "$p(m)$ holds for all $m \leq n$". If $q(n)$ is True for all n then $p(n)$ is True for all n, and vice-versa, so to prove that $p(n)$ is True for all n it is enough to show that $q(n)$ is true for all n. This can be proved by ordinary induction: $q(1)$ is True because $p(1)$ is True, and if we assume that $q(k)$ is True for some $k \geq 1$ then we know $p(j)$ for all $j \leq k$, so we know $p(k+1)$ (by our complete induction proof of $p(n)$ for all n), so we know $p(j)$ for all $j \leq k+1$ (here we need that there are no natural numbers strictly between k and $k+1$, which is Claim 4.16) so we know $q(k+1)$, and now ordinary induction can be used to infer that $q(n)$ is true for all n.

4.9 The well-ordering principle

A *least element* of a set S of numbers is an element x_0 of S such that for all $x \in S$ we have $x_0 \leq x$. None of the set of all real numbers, or all rational numbers, or all positive numbers, or all integers, has a least element. But it seems "obvious" that the set of natural numbers has a least element, namely 1, and indeed it can be proven by induction that $n \geq 1$ for every natural number n. More generally, it should be equally obvious that every *non-empty* subset of the natural numbers has a least element (the empty set does not have any elements, so in particular does not have a least element). This "obvious" fact is hard to pin down precisely, because there are so many subsets to consider. However, it is a true fact, called the well-ordering principle.

Theorem 4.19. *Every non-empty subset of the natural numbers has a least element.*

Proof: We use the principle of complete induction. Let S be a subset of the natural numbers with no least element, and let T be the complement of S (the set of all natural numbers not in S).

We have that $1 \in T$, because if $1 \in S$ then S would have a least element, namely 1.

Suppose, for some $k \geq 1$, that for all $j \leq k$ we have $j \in T$. Then $k+1$ is in T. Indeed, suppose $k+1$ is in S. Then $k+1$ would be a least element of S, since no natural number j with $j \leq k$ is in S, so if n is in S then $n > k$, so $n \geq k+1$ (this last by Claim 4.16).

By the principle of complete induction $T = \mathbb{N}$ and so S is empty.

We have proven that a subset of $\mathbb{N}$ with no least element is empty, which is the contrapositive of the assertion we wanted to prove. $\square$

In the other direction, one can also prove the principle of complete induction using the well-ordering principle, and so, remembering that ordinary and complete induction are equivalent, we conclude that the three principles

the principle of mathematical induction
the principle of complete induction
the well-ordering principle

are equivalent (and all follow from the axioms of real numbers). We will use the three interchangeably.