Don't teach mathematical induction! (as a way of proving summation formulas to students in an intro calculus course)

James Propp

July 22, 2014

Slides on the web at http://jamespropp.org/mathpath14.pdf.

The Principle of Mathematical Induction

Let S_n be a statement involving the positive integer n.
Suppose that
1. S_1 is true.
2. If S_k is true, then S_{k+1} is true.

Then S_n is true for all positive integers n.

Schematically:

$1 \to 2 \to 3 \to 4 \to \ldots$

A formula from second-semester calculus

(used in proving that the area under the parabola $y = x^2$ between $x = 0$ and $x = 1$ is $1/3$):

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = n(n+1)(2n+1)/6$$

That is:

$$\begin{aligned} 1^2 &= (1)(2)(3)/6 \\ 1^2 + 2^2 &= (2)(3)(5)/6 \\ 1^2 + 2^2 + 3^2 &= (3)(4)(7)/6 \\ 1^2 + 2^2 + 3^2 + 4^2 &= (4)(5)(9)/6 \end{aligned}$$

etc.

The standard approach

To prove the formula using mathematical induction:

- Show that the formula is true for $n = 1$ (the "base case").
- Show that when the formula is true for $n = k$, the formula must also be true for $n = k + 1$ (the "induction step").
- Now appeal to the principle of mathematical induction to conclude that the formula is true for all positive integers n.

See http://jamespropp.org/stewart-induction.pdf.

What's good about this proof

- It takes just a few lines.
- It appeals to a general principle (mathematical induction) that is used throughout mathematics.

. . . . . .

What's bad about this proof

- It's disconnected from first semester calculus.
- It's disconnected from the rest of second semester calculus.
- It confuses many students with its nested conditionals.

Here's an alternative approach:

The Constant Sequence Principle

Suppose $s_1, s_2, s_3, \ldots$ is an (infinitely long) sequence of numbers.

Then $s_1, s_2, s_3, \ldots$ is CONSTANT
(that is, there is some number c such that $s_n = c$ for all n)
if and only if $s_1, s_2, s_3, \ldots$ is UNCHANGING
(that is, $s_n = s_{n+1}$ for all n).

They're equivalent

You can prove the Constant Sequence Principle using the Principle of Mathematical Induction.

You can also prove the Principle of Mathematical Induction using the Constant Sequence Principle!

So mathematically they're the same assertion, even though they feel different.

New sequences from old

Given a sequence

$$s_1, s_2, s_3, \ldots,$$

define a new sequence

$$\Delta s_1, \Delta s_2, \Delta s_3, \ldots$$

by the rule

$$\Delta s_n = s_{n+1} - s_n.$$

This new sequence is called the sequence of differences, or the difference sequence.
Example: If

$$s_1, s_2, s_3, s_4, \cdots = 1, 4, 9, 16, \ldots,$$

then

$$\Delta s_1, \Delta s_2, \Delta s_3, \cdots = 3, 5, 7, \ldots.$$

New sequences from old

Notice that saying that a sequence is unchanging
($s_n = s_{n+1}$ for all n)
is equivalent to saying that its difference sequence is constantly zero
($s_{n+1} - s_n = 0$ for all n).

So the Constant Sequence Principle can be restated as follows:
$s_1, s_2, s_3, \ldots$ is constant if and only if $\Delta s_n = 0$ for all n.

To apply Δ to problems like proving that
$1^2 + 2^3 + \cdots + n^2 = n(n+1)(2n+1)/6$, we need one more tool.

A useful lemma

Lemma: $s_1, s_2, \ldots$ and $t_1, t_2, \ldots$ differ by a constant
(that is, there exists a number c such that $s_n - t_n = c$ for all n)
if and only if
$s_1, s_2, s_3, \ldots$ and $t_1, t_2, t_3, \ldots$ have the same difference sequence.

Example:

$$\begin{aligned} s_1, s_2, s_3, \ldots &= 1, 4, 9, 16, \ldots \\ \Delta s_1, \Delta s_2, \Delta s_3, \ldots &= 3, 5, 7, \ldots \\ t_1, t_2, t_3, \ldots &= 0, 3, 8, 15, \ldots \\ \Delta t_1, \Delta t_2, \Delta t_3, \ldots &= 3, 5, 7, \ldots \end{aligned}$$

A useful lemma

Lemma: $s_1, s_2, \ldots$ and $t_1, t_2, \ldots$ differ by a constant (that is, there exists c such that $s_n - t_n = c$ for all n)
if and only if
$s_1, s_2, s_3, \ldots$ and $t_1, t_2, t_3, \ldots$ have the same difference sequence.

Proof:

$$
\begin{aligned}
& s_n - t_n \text{ is constant} \\
\leftrightarrow \quad & \Delta(s_n - t_n) = 0 \text{ for all } n \\
\leftrightarrow \quad & (s_{n+1} - t_{n+1}) - (s_n - t_n) = 0 \text{ for all } n \\
\leftrightarrow \quad & (s_{n+1} - s_n) - (t_{n+1} - t_n) = 0 \text{ for all } n \\
\leftrightarrow \quad & \Delta s_n - \Delta t_n = 0 \text{ for all } n \\
\leftrightarrow \quad & \Delta s_n = \Delta t_n \text{ for all } n \\
\leftrightarrow \quad & s_n \text{ and } t_n \text{ have the same differences.}
\end{aligned}
$$

Proving that formula about sums of squares

Theorem: Let $s_n = 1^2 + 2^2 + \cdots + n^2$ and $t_n = n(n+1)(2n+1)/6$. Then $s_n = t_n$ for all n.
Proof:

$$\begin{aligned} \Delta s_n &= \quad (1^2 + 2^2 + \cdots + n^2 + (n+1)^2) \\ &\quad -(1^2 + 2^2 + \cdots + n^2) \\ &= (n+1)^2 \end{aligned}$$

while

$$\begin{aligned} \Delta t_n &= (n+1)(n+2)(2n+3)/6 - n(n+1)(2n+1)/6 \\ &= \ldots \\ &= (n+1)^2 \end{aligned}$$

so $\Delta s_n = \Delta t_n$ for all n.
So by our Lemma, s_n and t_n differ by a constant.

Proving that formula about sums of squares

We've shown that $s_n - t_n =$ some constant c, for all n, where
$s_n = 1^2 + 2^2 + \cdots + n^2$ and $t_n = n(n+1)(2n+1)/6$.
How do we nail down c?
$s_1 - t_1 = c$
$s_1 - t_1 = 1^2 - (1)(2)(3)/6 = 1 - 1 = 0$.
So $c = 0$.
So $s_n - t_n = 0$ for all n.
So $s_n = t_n$ for all n.

Comparison with proof by induction

Something like induction is going on in the second proof.
The two parts of the proof correspond to the induction step and the base case.
But the philosophy is different.

What's good about the second proof

- The second method of proof ties in with a solution-method used throughout calculus: determine an unknown function of x by a formula that incorporates some unknown constant, and then determine the constant.
- In physical terms: First use the equations of motion to find the general solution, and then plug in the initial conditions to find the particular solution you care about.
- The principle "A sequence is constant if and only if its difference sequence is 0" has an analogue in first semester calculus: "A function is constant if and only if its derivative function is 0".
- This approach is closer to what computers do when they prove formulas.

. . . . . .

What's bad about both proofs

- "Where did that $n(n+1)(2n+1)/6$ come from?"
- In both proofs, the right hand side is a given.
- Proving an equation in which the right hand side comes from nowhere reinforces the idea that math has a magical component, where someone smarter than you pulls something out of thin air.

After a short break, I'll show you how to derive $n(n+1)(2n+1)/6$ without knowing it ahead of time.

What kind of math is this?

The fancy name for “this whole Δ business” is “the difference calculus”.

In ordinary calculus, you have an operator D (for differentiation or derivative) that turns one function into a new function.

Here we have an operator Δ (for difference) that turns one sequence into a new sequence.

Another operator

A companion operator is Σ (for sum) defined by

$$\Sigma s_n = s_1 + s_2 + \cdots + s_n.$$

We call Σ the partial summation operator (or summation operator), and call $\Sigma s_1, \Sigma s_2, \Sigma s_3, \ldots$ the sequence of partial sums of the sequence $s_1, s_2, s_3, \ldots$.

Example

If

$$s_1, s_2, s_3, s_4, \dots = 1, 4, 9, 16, \dots$$

then

$$\Sigma s_1, \Sigma s_2, \Sigma s_3, \Sigma s_4 = 1, 5, 14, 30, \dots$$

which is exactly the sequence we're giving a formula for.

How could we ever come up with the formula

$$\Sigma n^2 = n(n+1)(2n+1)/6?$$

Idea: Express n^2 as a sum of suitable "building blocks" (polynomials s_n for which it's easier to compute Σs_n) and combine.

Pascal's triangle to the rescue

For what sequences s_n is it easy to compute Σs_n?
Look at downward sloping lines in Pascal's triangle:

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

$1, 1, 1, 1, \ldots$
$1, 2, 3, 4, \ldots$ (counting numbers)
$1, 3, 6, 10, \ldots$ (triangular numbers)
$1, 4, 10, 20, \ldots$ (tetrahedral numbers)
Etc.

Pascal's triangle to the rescue

For what sequences s_n is it easy to compute Σs_n?
Look at downward sloping lines in Pascal's triangle:

						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

$1,\ 1,\ 1,\ 1,\ldots \stackrel{\Sigma}{\to} 1,2,3,4,\ldots$
$1,\ 2,\ 3,\ 4,\ldots \stackrel{\Sigma}{\to} 1,3,6,10,\ldots$
$1,\ 3,\ 6,10,\ldots \stackrel{\Sigma}{\to} 1,4,10,20,\ldots$
Applying Σ to a diagonal gives the next diagonal!
(This is equivalent to the "Hockey Stick Pattern".)

The building blocks

If you want to apply Σ to a polynomial function $p(n)$, you want to express $p(n)$ as a combination of the right building blocks.

The right building blocks to use are binomial coefficients.

You can use mathematical induction (or the constant sequence principle) to show that

$$\begin{aligned} \Sigma 1 &= n \\ \Sigma n &= n(n+1)/2 \\ \Sigma n(n+1)/2 &= n(n+1)(n+2)/6 \end{aligned}$$

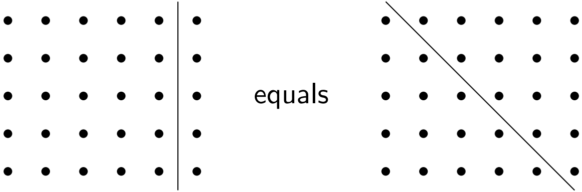
etc.

Breaking things down into building blocks

Claim: $n^2 = n(n+1)/2 \; + \; n(n+1)/2 \; - \; n$.

Algebraic proof:
$n(n+1)/2 + n(n+1)/2 - n = n(n+1) - n = n^2 + n - n = n^2$.

Picture proof: Here's a picture proof that $n^2 + n$ is the sum of the nth triangle number with itself, in the case $n = 5$:



Breaking things down into building blocks

$$\begin{array}{rcrcrcr} 1 & = & 1 & + & 1 & - & 1 \\ 4 & = & 3 & + & 3 & - & 2 \\ 9 & = & 6 & + & 6 & - & 3 \\ 16 & = & 10 & + & 10 & - & 4 \\ 25 & = & 15 & + & 15 & - & 5 \end{array}$$

So

$$1+4+9+16+25 = 2(1+3+6+10+15) - (1+2+3+4+5).$$

Breaking things down into building blocks

Write $s_n = n$, $t_n = n(n+1)/2$, $u_n = n(n+1)(n+2)/6$.
We have $n^2 = 2t_n - s_n$, so

$$\begin{aligned} \Sigma n^2 &= \Sigma(2t_n - s_n) \\ &= 2\Sigma t_n - \Sigma s_n \\ &= 2u_n - t_n \\ &= 2n(n+1)(n+2)/6 - n(n+1)/2 \\ &= n(n+1)(2n+1)/6. \end{aligned}$$

What's good about this proof

- This way of looking at polynomials sneakily prepares students for linear algebra. Ordinarily we write a quadratic function of n as a combination of n^2, n, and 1, but when we're applying the Σ operator it's handier to write a quadratic function of n as a combination of $n(n+1)/2$, n, and 1.
- College students learn to see the set of polynomials of degree at most 2 as a three-dimensional vector space, and learn that there are lots of ways to choose a basis for a vector space!

. . . . . .

What's bad about this proof

- "Where did $n^2 = 2(n(n+1)/2) - (n)$ come from?"
- How would we generalize it if we wanted to express n^3 as a sum of binomial coefficients (as a step toward computing Σn^3)?

Being systematic

Here's how things work for Σn^3.
We want coefficients A, B, C, and D so that

$$n^3 = A[n(n+1)(n+2)/6] + B[n(n+1)/2] + C[n] + D[1].$$

We can choose A so that the cubic term of $A[n(n+1)(n+2)/6]$ matches the cubic term on the LHS. Now write

$$n^3 - A[n(n+1)(n+2)/6] = B[n(n+1)/2] + C[n] + D[1].$$

Because of our choice of A, the LHS is a quadratic function of n. We can choose B so that the quadratic term of $B[n(n+1)/2]$ matches the quadratic term on the LHS. Now write

$$n^3 - A[n(n+1)(n+2)/6] - B[n(n+1)/2] = C[n] + D[1].$$

Because of our choice of A and B, the LHS is a linear function of n. Choose C and D to make equality hold, and we're done.

Being systematic

Once we find A, B, C, D so that

$$n^3 = A[n(n+1)(n+2)/6] + B[n(n+1)/2] + C[n] + D[1]$$

we can apply Σ to both sides of the equation to get

$$\begin{aligned} \Sigma n^3 &= A[n(n+1)(n+2)(n+3)/24] + B[n(n+1)(n+2)/6] \\ &\quad + C[n(n+1)/2] + D[n] \end{aligned}$$

Homework: Do it, and show that $\Sigma n^3 = (n(n+1)/2)^2$.

. . . . . .

What's bad about this proof

- Why are we even proving $\Sigma n^2 = (2n^3 + 3n^2 + n)/6$ in the first place?
- We end up throwing most of the formula away! For the application to calculus, all we really need is that first coefficient, $2/6$.

What else is bad about this proof

- It's too long.
- Students won't have a chance to use the method later in the course.

So what is the best proof?

Since I'm running out of time, I'd better answer the question I keep asking. What is the best proof?

There is no such thing as THE best proof!

There are potentially as many different "best proofs" as there are learners.

Becoming a good guide to mathematical subjects like calculus requires finding lots of different learning-paths, and being able to match learning paths with learners, knowing that in a classroom setting, you can't please everybody; you aim to please as big a plurality as you can.

Um...

So, why do I say “Don’t teach induction in calculus classes”?

Aren’t there some students who’ll love it?

Absolutely! But I think that for most students, a proof that uses some of the features of the difference calculus is better, because it ties in better with some ideas they’ve seen already and other ideas they’re about to see.

There’s a tendency for students to see math as a hodgepodge: algebra, geometry, calculus, discrete math, etc. We math teachers should take advantage of every opportunity to display for our students the underlying unity of the subject.

The years ahead

The video of this talk will be on the web for years to come, as will the slides, so come back to them when you're taking your first calculus class.

And if any of you go on to write textbooks, or online course materials, consider using some of the ideas from this talk.

Video: Stay tuned! To be posted soon.

Slides: http://jamespropp.org/mathpath14.pdf.

Thanks for listening!