MATHEMATICAL INDUCTION, POWER SUMS, AND DISCRETE CALCULUS

PETE L. CLARK

1. Something interesting to say about uninteresting induction proofs

I am currently teaching mathematical induction in a "transitions" course for prospective math majors. Inevitably the first, and apparently least interesting, examples of induction proofs are identities like the following:

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

$$1^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2}$$

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

It is well known to both instructors and students that such induction proofs quickly boil down to an algebraic computation. For instance, to show (1), we verify that $1 = \frac{1(1+1)}{2}$, then assume that for some fixed n,

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

Then we add $n+1$ to both sides, getting

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

Many beginning students would like to know what is "really going on here". One answer is that we are simply applying the principle of mathematical induction via a procedure which will be more clear after viewing other, similar examples. This is a perfectly good answer but not the only possible one. We might get a better answer by asking a more precise question, e.g. "Why is mathematical induction particularly well suited to proving closed-form identities involving finite sums?"

I have a quite different answer to that question. Here goes: the task in finding a closed form expression for a sum is to eliminate the "dot dot dot". This is exactly what induction does for us. In general, suppose f and g are functions from the positive integers to the real numbers, and our task is to prove that

$$f(1) + \ldots + f(n) = g(n).$$

Now let us contemplate proving this by induction.

First we verify that $f(1) = g(1)$.

Now assume the identity for a fixed positive integer n, and add $f(n)$ to both sides:

$$f(1) + \ldots + f(n) + f(n+1) = (1 + \ldots + f(n)) + f(n+1) \stackrel{\text{IH}}{=} g(n) + f(n+1).$$

Since we want to get $g(n+1)$ in the end, what remains to be shown is precisely that $g(n+1) = g(n) + f(n+1)$, or equivalently

$$g(n+1) - g(n) = f(n+1). \tag{4}$$

If f and g are both simple algebraic functions, then (assuming the result is actually true!) verifying (4) is a matter of high school algebra. For example, to prove (2), then – after checking that $1^2 = \frac{1(1+1)(2\cdot 1+1)}{6}$ – the identity we need to verify is

$$\frac{(n+1)(n+2)(2(n+1)+1)}{6} - \frac{n(n+1)(2n+1)}{6} = (n+1)^2,$$

and we need only expand out both sides and see that we get n^2+2n+1 either way.

I wish to suggest that this procedure is analogous to what happens in calculus when we have two functions f and g and wish to verify that $\int_1^x f(t) = g(x)$: it suffices, by the Fundamental Theorem of Calculus, to show that $g(1) = 0$ and $\frac{dg}{dx} = f$. This analogy may well seem farfetched at the moment, so let's leave it aside and press on.

But not just yet. First let us consider a slightly different framework: we have two functions $f, g : \mathbb{Z}^+ \rightarrow \mathbb{R}$, but instead of trying to show $f(1)+\ldots+f(n) = g(n)$, we are trying to show that

$$f(1) + \ldots + f(n) = g(1) + \ldots + g(n).$$

Let us write $F(n) = f(1) + \ldots + f(n)$ and $G(n) = g(1) + \ldots + g(n)$, so we want to show that $F(n) = G(n)$ for all n. Suppose we try to prove this by induction. We must show that $F(1) = G(1)$, and then we get to assume that for a given n, $F(n) = G(n)$, and we need to show $F(n+1) = G(n+1)$. Here's the point: given

$$F(n) = G(n), \tag{5}$$

the desired conclusion $G(n+1) = F(n+1)$ is equivalent to

$$F(n+1) - F(n) = G(n+1) - G(n). \tag{6}$$

Indeed, if (5) and (6) both hold, then adding them together, we get

$$F(n+1) = F(n+1) - F(n) + F(n) = G(n+1) - G(n) + G(n) = G(n+1),$$

and similarly, if we know $F(n+1) = G(n+1)$, then subtracting (5), we get (6). So our application of induction gives that it is necessary and sufficient to show that $F(n+1) - F(n) = G(n+1) - G(n)$. Note however that

$$F(n+1) - F(n) = (f(1) + \ldots + f(n) + f(n+1)) - (f(1) + \ldots + f(n)) = f(n+1),$$

and similarly $G(n+1) - G(n) = g(n+1)$. We need to show this for all n, i.e., we need to know that $f(n) = g(n)$ for all $n \geq 2$. Since we also needed this for $n = 1$, we see – perhaps somewhat sheepishly – that what we have shown is the following.

Proposition 1. *Let $f, g : \mathbb{Z}^+ \rightarrow \mathbb{R}$. The following are equivalent:*
(i) For all $n \in \mathbb{Z}^+$, $f(n) = g(n)$.
(ii) For all $n \in \mathbb{Z}^+$, $f(1) + \ldots + f(n) = g(1) + \ldots + g(n)$.

This is not earth-shattering, but the following minor variation is somewhat interesting. Namely, for any function f, define a new function Δf, by

$$(\Delta f)(n) = f(n+1) - f(n).$$

The point here is that $(\Delta f)(1) + \ldots + (\Delta f)(n) =$

$$(f(2)-f(1))+(f(3)-f(2))+\ldots+(f(n)-f(n-1))+(f(n+1)-f(n)) = f(n+1)-f(1).$$

Since Proposition 1 holds for all functions f and g, in particular it holds for Δf and Δg, and we get:

Proposition 2. *Let $f, g : \mathbb{Z}^+ \to \mathbb{R}$. The following are equivalent:*
(i) For all $n \in \mathbb{Z}^+$, $(\Delta f)(n) = f(n+1) - f(n) = g(n+1) - g(n) = (\Delta g)(n)$.
(ii) For all $n \in \mathbb{Z}^+$, $f(n+1) - f(1) = g(n+1) - g(1)$.

We easily deduce the following:

Theorem 3. *Let $f, g : \mathbb{Z}^+ \to \mathbb{R}$. Suppose :*
(i) For all n, $(\Delta f)(n) = (\Delta g)(n)$, and
(ii) $f(1) = g(1)$.
Then $f(n) = g(n)$ for all n.

But wait! This is directly reminiscent of the following theorem.

Theorem 4. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be differentiable functions. Suppose:*
(i) For all $x \in \mathbb{R}$, $f'(x) = g'(x)$, and
(ii) $f(1) = g(1)$.
Then $f(x) = g(x)$ for all x.

Proof. We define the new function $h(x) = f(x) - g(x)$. Our hypotheses are that $h'(x) = (f(x)-g(x))' = f'(x)-g'(x) = 0$ for all x and that $h(1) = f(1)-g(1) = 0$, and we want to show that $h(x) = 0$ for all x. So suppose not, i.e., there exists some x_0 with $h(x_0) \neq 0$. Certainly $x_0 \neq 1$, and we may assume without loss of generality that $x_0 > 1$. Now we apply the Mean Value Theorem to $h(x)$ on the interval $[1, x_0]$: there exists a real number c, $1 < c < x_0$, such that

$$h'(c) = \frac{h(x_0) - h(1)}{x_0 - 1}.$$

Thus $h'(c) \neq 0$, contradicting the hypothesis that $h'(x)$ for all $x \in \mathbb{R}$. ☐

If we think instead of two functions $x(t)$ and $y(t)$ giving the position of a moving particle at time t, the Theorem states the following physically plausible result: two moving bodies with identical instantaneous velocity functions and which have the same position at time $t = 1$ will have the same position for all times t. Theorem 3 applies to $x(t)$ and $y(t)$ as follows: we look only at positive integer values of x and y, and we can interpret $(\Delta x)(n)$ as the average velocity of $x(t)$ between times n and $n+1$. The result then says that if $x(t)$ and $y(t)$ start out at the same position at time $t = 1$ and their average velocities on each interval $[n, n+1]$ agree, then x and y have the same positions at all positive integer values.

Note that we needed a deep theorem from calculus to show Theorem 4, but for the analogous Theorem 3 we only needed mathematical induction.

2. Another natural question

There is something else in the proofs by induction of identities (1), (2), (3) that confuses many students. Namely, how do we figure out what goes on the right hand side?!? The schematic introduced in the previous section allows us to rephrase this more crisply, as follows: given a function $f(n)$ like $f(n) = n$, $f(n) = n^2$ or $f(n) = n^3$, how do we find the function $g(n)$ such that $g(n) = f(1) + \ldots + f(n)$?

Again, there is a simple answer which is perfectly good and probably indeed should be the first answer given. Namely, we should clarify our task: we were not claiming to be able to find – and still less, asking the student to find! – the right hand side of these identities. It is important to understand that induction is never used to discover a result; it is only used to prove a result that one already either suspects to be true or has been asked to show. In other words, the simple answer to the question "How do we figure out what goes on the right hand side?" is: we don't. It is given to us as part of the problem.

But this is a very disappointing answer. I feel that the disappointment this answer engenders in students is of pedagogical significance, so forgive me while I digress on this point (or skip ahead, of course). In other words, it often happens in university level math classes that we present certain techniques and advertise them as giving solutions to certain problems, but we often do not discuss the limitations of these techniques, or more positively, try to identify the range of problems to which the techniques can be successfully applied. For instance, after learning several integration techniques, many calculus students become anxious when they realize that they may not know which technique or combination of techniques to apply to a given proble. They often ask for simple rules like, "Can you tell us when to use integration by parts?" I at least have found it tempting as an instructor to brush off such questions, or answer them by saying that much of the point is for them to gain enough experience with the various techniques so as to be able to figure out (or guess) which techniques will work on a given problem. But calculus instructors know something that the students don't: many functions, like e^{x^2}, simply do not have elementary antiderivatives. It would be a terrible disservice not to point this out to the students, as well as not to clue them into the truth: we carefully select the integration problems we give the students so that (i) elementary antiderivatives exist and (ii) they can indeed be found using the set of tools we have taught them.

There are real dangers that such practices will dampen or kill off students' mathematical curiosity. Most students initially think they are being asked to solve a robust class of problems – and thus, they think that they should know how to solve these problems, and are disturbed that they don't – but eventually they learn that *less* knowledge than they thought is actually needed to solve the exercise. This is intellectually deadening. What use it it to know how to prove $1^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$ without knowing how to figure out what should go on the right hand side? The answer is that there is no inherent use that I can see (other than being able to compute $\int_a^b x^2 dx$ using Riemann sums); it is just an opportunity to demonstrate a mastery of a very narrow skill, which we identified in the last section as being able to verify that $g(n + 1) - g(n) = f(n)$.

Of course, sometimes there are necessary reasons for not answering the natural questions: the answer may be very complicated! For instance, although I tell my calculus students that the reason they cannot integrate e^{x^2} is that it is provably impossible, I do not give any indication why this is true: such arguments are well beyond the scope of the course.

But such is not the case for $1^2 + \ldots + n^2$, and we now present a simple method to derive formulas for the power sums

$$S_d(n) = 1^d + \ldots + n^d.$$

We begin with the sum

$$S = \sum_{i=1}^{n} \left((i+1)^{d+1} - i^{d+1}\right),$$

which we evaluate in two different ways. First, writing out the terms gives

$$S = 2^{d+1} - 1^{d+1} + 3^{d+1} - 2^{d+1} + \ldots + n^{d+1} - (n-1)^{d+1} + (n+1)^{d+1} - n^{d+1} = (n+1)^{d+1} - 1.$$

Second, by first expanding out the binomial $(i+1)^{d+1}$ we get

$$S = \sum_{i=1}^{n} \left((i+1)^{d+1} - i^{d+1}\right) = \sum_{i=1}^{n} \left(i^{d+1} + \binom{d+1}{1} i^d + \ldots + \binom{d+1}{d} i + 1 - i^{d-1}\right) =$$

$$\sum_{i=1}^{n} (\binom{d+1}{1} i^d + \ldots + \binom{d+1}{d} i) = \binom{d+1}{1} \sum_{i=1}^{n} i^d + \ldots + \binom{d+1}{d} \sum_{i=1}^{n} i + \sum_{i=1}^{n} 1 =$$

$$\sum_{j=0}^{d} \binom{d+1}{d+1-j} S_j(n) = \sum_{j=0}^{d} \binom{d+1}{j} S_j(n).$$

Equating our two expressions for S, we get

$$(n+1)^{d+1} - 1 = \sum_{j=0}^{d} \binom{d+1}{j} S_j(n).$$

Solving this equation for $S_d(n)$ gives

$$S_d(n) = \frac{(n+1)^{d+1} - \left(\sum_{j=0}^{d-1} \binom{d+1}{j} S_j(n)\right) - 1}{(d+1)}. \tag{7}$$

This formula allows us to compute $S_d(n)$ recursively: that is, given exact formulas for $S_j(n)$ for all $0 \leq j < d$, we get an exact formula for $S_d(n)$. And getting the ball rolling is easy: $S_0(n) = 1^0 + \ldots + n^0 = 1 + \ldots 1 = n$.

Example ($d = 1$): Our formula gives

$$1 + \ldots + n = S_1(n) = (\frac{1}{2})((n+1)^2 - S_0(n) - 1) = (\frac{1}{2})(n^2 + 2n + 1 - n - 1) = \frac{n(n+1)}{2}.$$

Example ($d = 2$): Our formula gives $1^2 + \ldots + n^2 = S_2(n) =$

$$\frac{(n+1)^3 - S_0(n) - 3S_1(n) - 1}{3} = \frac{n^3 + 3n^2 + 3n + 1 - n - \frac{3}{2}n^2 - \frac{3}{2}n - 1}{3} =$$

$$\frac{2n^3 + 3n^2 + n}{6} = \frac{n(n+1)(2n+1)}{6}.$$

We leave it as an exercise for the reader to derive (3) using this method.

This approach is not ideally suited to rapid calculation. In particular, if we wanted a formula for $S_{10}(n)$ then our method requires us to first derive formulas for $S_1(n)$ through $S_9(n)$, which would be rather time-consuming. On the other hand (7) has theoretical applications: with it in hand we can harness induction to a much more worthy goal, namely the proof of the following result.

Theorem 5. *For every positive integer d, there exist $a_1, \ldots, a_d \in \mathbb{Q}$ such that for all $n \in \mathbb{Z}^+$ we have*

$$1^d + \ldots + n^d = \frac{n^{d+1}}{d+1} + a_d n^d + \ldots + a_1 n.$$

We leave this **interesting** induction proof as an exercise for the reader.

Again though, let us not neglect the natural question: the method presented gives a **recursive** method for evaluating the power sums $S_d(n)$. Can we find a closed form expression for $S_d(n)$ in general? The answer is *yes*, and we will derive it as an application of the discrete calculus, a topic to which we turn to next.

3. Discrete calculus

Let $f : \mathbb{Z}^+ \to \mathbb{R}$ be any function. We define the new function Δf as follows:

$$(\Delta f)(n) = f(n+1) - f(n).$$

We view Δ as being an operator on the set $V = \{f : \mathbb{Z}^+ \to \mathbb{R}\}$ of all functions from $\mathbb{Z}^+$ to $\mathbb{R}$. Specifically, it is called the **forward difference operator**.

There is also a **backward difference operator** ∇, which takes f to

$$(\nabla f)(n) = f(n) - f(n-1).$$

Something must be said about $(\nabla f)(1) = f(1) - f(0)$, since by our hypothesis f need not be defined at 0. Let us make the convention of "extension by zero": whenever a formula or definition calls upon us to evaluate a function at a value outside its stated domain, we assign it the value 0. In particular this means that

$$(\nabla f)(1) = f(1) - f(0) = f(1).$$

Above we remarked upon an analogy between the Δ operator and the operator $\frac{d}{dx}$ of differentiation. We will further develop the analogy here to the extent that it will seem reasonable to refer to think of Δ (and also ∇) as the **discrete derivative** of f. But we note one nonanalogy : the operator $\frac{d}{dx}$ is not well-defined for any function $f : \mathbb{R} \to \mathbb{R}$; it is only defined for differentiable functions (which is in fact sort of a circular definition, but that's in the nature of things). Certainly Δ and ∇ are defined for all $f \in V$: every discrete function is "discretely differentiable".

Recall that the usual derivative satisfies the following properties:

(D1) The derivative of a constant function is the zero function.
(D2) $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$.
(D3) $\frac{d}{dx}(\alpha f) = \alpha \frac{df}{dx}$.
(D4) $\frac{d}{dx}(fg) = \frac{df}{dx} g + f \frac{dg}{dx}$.

(D5) $\frac{d}{dx}\frac{f}{g} = \frac{g\frac{df}{dx} - f\frac{dg}{dx}}{g^2}$.
(D6) $\frac{dx^n}{dx} = nx^{n-1}$.

It is easy to check that Δ satisfies (D1) through (D3):

(D1Δ) If f is constant, then $\forall n, f(n+1) = f(n)$, so $\forall n, (\Delta f)(n) = 0$.

(D2Δ) $\Delta(f+g)(n) = (f+g)(n+1) - (f+g)(n) = f(n+1) + g(n+1) - f(n) - g(n) =$

$$f(n+1) - f(n) + g(n+1) - g(n) = (\Delta f)(n) + (\Delta g)(n).$$

(D3Δ) $\Delta(\alpha f)(n) = \alpha f(n+1) - \alpha f(n) = \alpha(f(n+1) - f(n)) = (\alpha \Delta f)(n)$.

Similarly, ∇ satisfies (D2) and (D3), but not quite (D1): if $f(n) = C$ for all n, then for all $n \geq 2$ we have $(\nabla f)(n) = f(n) - f(n-1) = C - C = 0$, but $(\nabla f)(1) = f(1) - f(0) = C - 0 = C$.

On the other hand, if you try to verify the direct analogue of (D4) – namely $\Delta(fg) = \Delta(f)g + f\Delta(g)$ – you soon see that it does not work out: the left hand side has two terms, the right hand side has four terms, and no cancellation is possible. Another way to see that this formula cannot be correct is as follows: if x denotes the function $n \mapsto n$, then as with the usual derivative we have

$$\Delta(x)(n) = n + 1 - n = 1;$$

i.e., the discrete derivative of the identity function x is the constant function 1. But from this and the product rule (D4), the power rule (D6) follows by mathematical induction. However, let us calculate $\Delta(x^2)$:

$$\Delta(x^2)(n) = (n+1)^2 - n^2 = n^2 + 2n + 1 - n^2 = 2n + 1,$$

whereas

$$((\Delta x)x + x\Delta x)(n) = n + n = 2n.$$

Something slightly different does hold:

$$(\Delta fg)(n) = (\Delta f)(n)g(n) + f(n+1)(\Delta g)(n).$$

This formula looks a bit strange: the left hand side is symmetric in f and g, whereas the right hand side is not. Thus there is another form of the product rule:

$$(\Delta fg)(n) = \Delta(gf)(n) = f(n)(\Delta g)(n) + (\Delta f)(n)g(n+1).$$

A more pleasant looking form of the product rule is

$$(\Delta fg) = f\Delta g + (\Delta f)g + \Delta f \Delta g. \tag{8}$$

This formulation makes clear the relationship with the usual product rule for $\frac{d}{dx}$: if $f : \mathbb{R} \rightarrow R$ is a differentiable function, then defining Δf to be $f(x+h) - f(x)$, one checks that (8) remains valid. Now dividing both sides by $\Delta x = h$, we get

$$\frac{(\Delta fg)}{h} = f\frac{\Delta g}{\Delta x} + \frac{\Delta f}{\Delta x}g + \frac{\Delta f}{\Delta x}\Delta g.$$

As $h \rightarrow 0$, $\frac{\Delta f}{\Delta x} \rightarrow \frac{df}{dx}$, $\frac{\Delta g}{\Delta x} \rightarrow \frac{dg}{dx}$ and the last term approaches $\frac{df}{dx} \cdot 0 = 0$. So the product rule for $\frac{d}{dx}$ is a *simplification* of the discrete product rule, an approximation which becomes valid in the limit as $h \rightarrow 0$. Thinking back to calculus, it becomes clear that many identities in calculus are simplfications of corresponding discrete identities obtained by neglecting higher-order differentials.

Moral: the conventional calculus is more *analytically complex* than discrete calculus: in the former, one must deal rigorously and correctly with limiting processes, whereas in the latter no such processes exist. Conversely, discrete calculus can be more *algebraically complex* than conventional calculus. More on this later.

There are corresponding, but slightly different, identities for ∇:

$$(\nabla fg)(n) = (\nabla f)(n)g(n-1) + f(n)(\nabla g)(n) = (\nabla f)(n)g(n) + f(n-1)(\nabla g)(n).$$

Similarly, in place of (D5) we have something slightly different:

$$\Delta(f/g)(n) = \frac{(\Delta f)(n)g(n) - f(n)(\Delta g)(n)}{g(n)g(n+1)},$$

assuming that for all n, $g(n) \neq 0$. In place of (D6) we get something significantly more complicated, which we will discuss in detail in the next section.

3.1. **Higher discrete derivatives.**

Of course there is no difficulty in defining the second, and higher-order, discrete derivatives. Namely, we put $\Delta^2 f = \Delta(\Delta f)$, and for $k \in \mathbb{Z}$, we define $\Delta^k f$ as $\Delta(\Delta^{k-1} f)$. Explicitly

$$(\Delta^2 f)(n) = (\Delta f(n+1) - f(n)) = f(n+2) - f(n+1) - (f(n+1) - f(n)) = f(n+2) - 2f(n+1) + f(n).$$

Similarly,

$$(\Delta^3 f)(n) = (\Delta f(n+2) - 2f(n+1) + f(n))$$

$$= f(n+3) - f(n+2) - 2f(n+2) + 2f(n+1) + f(n+1) - f(n) = f(n+3) - 3f(n+2) + 3f(n+1) - f(n).$$

This suggests the following general formula for the kth discrete derivative:

Theorem 6. *Let $f : \mathbb{Z}^+ \to \mathbb{R}$ be a discrete function. For any $k \in \mathbb{Z}^+$, we have*

$$(\Delta^k f)(n) = \sum_{i=0}^{n} (-1)^k \binom{k}{i} f(n+k-i).$$

We leave the proof to the reader as an exercise in mathematical induction.

3.2. **The discrete antiderivative.**

In the usual calculus, one studies the inverse process to differentiation, namely antidifferentiation. The plausible candidate for the discrete antiderivative is just the summation operator $\Sigma : f \to \Sigma f$ defined as

$$(\Sigma f)(n) = f(1) + \ldots + f(n).$$

Let us now calculate the composite operators $\Delta \circ \Sigma$ and $\Sigma \circ \Delta$ applied to an arbitary discrete function f:

$$(\Delta \circ \Sigma)(f)(n) = \Delta(n \mapsto f(1) + \ldots + f(n)) = f(1) + \ldots + f(n+1) - (f(1) + \ldots + f(n)) = f(n+1).$$

Similarly,

$$(\Sigma \circ \Delta)(f)(n) = \Sigma(n \mapsto f(n+1) - f(n)) = f(2) - f(1) + \ldots + f(n+1) - f(n) = f(n+1) - f(1).$$

So Σ and Δ are very close to being inverse operators, but there is something slightly off with the indexing. Now the ∇ operator proves its worth: we have

$$(\Sigma\circ\nabla)(f)(n) = \Sigma(n \mapsto f(n)-f(n-1)) = f(1)-f(0)+\ldots+f(n)-f(n-1) = f(n)-f(0) = f(n)$$

and

$$(\nabla\circ\Sigma)(f)(n) = \nabla(n \mapsto f(1)+\ldots+f(n)) = (f(1)+\ldots+f(n))-(f(1)+\ldots+f(n-1)) = f(n).$$

So indeed ∇ and Σ are inverse operators. This is even better than in the usual calculus, where the antiderivative is only well-determined up to the addition of a constant. In other words:

Theorem 7. *(Fundamental Theorem of Discrete Calculus, v. 1) For functions $f, g : \mathbb{Z}^+, \mathbb{R}$, the following are equivalent:*
(i) For all $n \in \mathbb{Z}^+$, $f(n) = g(n) - g(n-1)$ $(f = \nabla g)$.
(ii) For all $n \in \mathbb{Z}^+$, $g(n) = f(1) + \ldots + f(n)$ $(\Sigma f = g)$.

In other words, if we want to find a closed form expression for $\sum_{i=1}^{n} f(i)$, it suffices to find a function g such that $\nabla g = f$.

3.3. **The discrete definite integral.**

Thinking carefully on traditional calculus reveals a certain discrepancy between Theorem 7 and the usual fundamental theorem. Namely, in usual calculus the fundamental theorem relates two notions of integrals: the indefinite and the definite integral. The indefinite integral of f is by definition any function F such that $\frac{dF}{dx} = F$. In this sense, we found that Σf is an antiderivative of f: $(\nabla\Sigma f)(n) = f(n)$ for all $n \in \mathbb{Z}^+$. But what about the definite integral? Traditionally, the definite integral of a (say, continuous) function $f : [a, b] \to \mathbb{R}$ is a real number $\int_a^b f(x)dx$ which is supposed to represent the signed area under the curve $y = f(x)$ (and is formalized as a limit of Riemann sums).

It is not difficult to come up with a discrete definite integral. There is a small choice to be made, as to whether we use Δ or ∇ as our differentiation operator. Since in the previous section we developed a ∇-theory, to show that it doesn't make any essential difference, this time we set things up so as to work well with Δ.

Namely, for positive integers $a \leq b$ and a discrete function $f : \mathbb{Z}^+ \to \mathbb{R}$, we define

$$S_a^b f = f(a) + \ldots + f(b-1).$$[1]

This has the following property: we can define a new function F as a definite integral with variable upper limit:

$$F(n) = S_1^n f = f(1) + \ldots + f(n-1).$$

Then we have a perfect analogue of the fundamental theorem of calculus, in two parts. First,

$$(\Delta F)(n) = (f(1) + \ldots + f(n)) - (f(1) + \ldots f(n-1)) = f(n),$$

[1]Stopping the sum at $f(b-1)$ rather than $f(b)$ is the correct normalization for Δ, as we will shortly see.

so that F is a function whose discrete derivative – this time Δ and not ∇ – is equal to f. Second, for $1 \leq a \leq b$, we have

$$F(b) - F(a) =$$

$$(f(1) + \ldots + f(b-1)) - (f(1) + \ldots + f(a-1)) = f(a) + \ldots + f(b-1) = S_a^b f.$$

In summary:

Theorem 8. *(Discrete fundamental theorem of calculus, v. 2) Let $f : \mathbb{Z}^+ \to \mathbb{R}$ be any function. For any two integers $a \leq b$, define $\Sigma_a^b f = f(a) + \ldots + f(b-1)$, and define a function $F : \mathbb{Z}^+ \to \mathbb{R}$ by $F(n) = S_1^n f$. Then:*
a) $\Delta F = f$.
b) $S_a^b f = F(b) - F(a)$.

Remark: Of course in the traditional calculus the definite integral $\Sigma_a^b f(x)dx$ has an area interpretation. This can be given to the discrete definite integral $\Sigma_a^b f$ as well. Namely, for a discrete function $f : \mathbb{Z}^+ \to \mathbb{R}$, extend it to a function $f : [1, \infty) \to \mathbb{R}$ by $f(x) = f(\lfloor x \rfloor)$. Thus f is the unique step function whose restriction to $\mathbb{Z}^+$ is f and is left-continuous at each integer value. Then one has

$$\int_a^b f dx = S_a^b f,$$

i.e., the area under the step function $f : [1, n]$ is precisely $f(1) + \ldots + f(n-1)$.

Remark: We hope the reader has by now appreciated that the distinction between a discrete calculus based on Δ versus one based on ∇ is very minor: in the former case, we define the antiderivative to be $f \mapsto (n \mapsto f(1) + \ldots + f(n-1))$ (and similarly for the definite integral) $\Sigma_a^b f$ and the in the latter case we define the antiderivative to be $f \mapsto (n \mapsto f(1) + \ldots + f(n))$. So we could make do by choosing either one once and for all as the discrete derivative; on the other hand, there is no compelling need to do so.

Remark: It would also make sense to consider functions $f : \mathbb{Z} \to \mathbb{R}$. Then one defines Δ as before, and ∇ even more simply than before, namely $(\nabla f)(n) = f(n) - f(n-1)$ for all $n \in \mathbb{Z}$: no convention about $f(0)$ is necessary. In this context the definite integral $\Sigma_a^b f$ still makes perfect sense, and it is more clear that the notion of an indefinite integral well-defined only up to a constant. For instance, as we did aove we can define a discrete antiderivative F of f by

$$F(n) = \Sigma_a^{n-1} f(n)$$

for any fixed $a \in \mathbb{Z}$. Conversely, it is easy to see that all discrete antiderivatives of f differ from each other by an additive constant. Equivalently, the choice of a particular antiderivative F is determined by specifying its value at a single integer.

3.4. **Discrete integration by parts.**

Let $f, g : \mathbb{Z} \to \mathbb{R}$ be discrete functions. Summing the product rule

$$(\Delta fg)(k) = (\Delta f)(k)g(k) + f(k+1)(\Delta g)(k)$$

from $k = 1$ to n yields

$$f(n+1)g(n+1) - f(1)g(1) = \sum_{k=1}^{n} (\Delta f)(k)g(k) + \sum_{k=1}^{n} f(k+1)(\Delta g)(k),$$

or

$$\sum_{k=1}^{n} (\Delta f)(k)g(k) = f(n+1)g(n+1) - f(1)g(1) - \sum_{k=1}^{n} f(k+1)(\Delta g)(k).$$

Applying this with $f(n) = n$ gives

$$\sum_{k=1}^{n} g(k) = (n+1)g(n+1) - g(1) \sum_{k=1}^{n} (k+1)(\Delta g)(k). \tag{9}$$

Taking $g(n) = n^d$, we get

$$S_d(n) = \sum_{k=1}^{n} k^d = (n+1)^{d+1} - 1 - \sum_{k=1}^{n} (k+1)((k+1)^d - k^d).$$

$$\ldots$$

RELATE THIS TO THE CALCULATION OF §2.

Example: Let $h(n) = n2^n$. Evaluate $H(n) = g(1) + \ldots + g(n)$.

Solution: If instead we are asked to find the antiderivative of xe^x, we would use integration by parts, writing $h(x) = g(x)df(x)$, with $g(x) = x$, $df(x) = e^x dx$. Then

$$\int xe^x dx = xe^x - \int e^x = xe^x - e^x + C = (x-1)e^x + C.$$

Let's try the discrete analogue: put $(\Delta f)(n) = 2^n$, $g(n) = n$, so that $h(n) = (\Delta f)(n)g(n)$. Then $f(n) = 2^n$, $\Delta g = 1$, so

$$H(n) = \sum_{k=1}^{n} (\Delta f)(k)g(k) = (n+1)2^{n+1} - 2 - \sum_{k=1}^{n} 2^{k+1} = (n+1)2^{n+1} - 2 - (2^{n+2} - 4)$$
$$= 2^{n+1}(n+1-2) = (n-1)2^{n+1} + 2.$$

Thus the result is closely analogous but not precisely what one might guess: note the 2^{n+1} in place of e^x.

3.5. **Some discrete differential equations.**

Example: Let us find all discrete functions $f : \mathbb{Z}^+ \to \mathbb{R}$ with $(\Delta f) = f$. In other words, for all $n \geq 1$, we have $f(n+1) - f(n) = f(n)$, or $f(n+1) = 2f(n)$. Evidently we need $f(n) = 2^{n-1}f(1)$, so that the general solution is $f(n) = C2^n$, for any real number C. Thus in some sense 2 is the discrete analogue of e!

Example: For $\alpha \in \mathbb{R}$, let us find all discrete functions f with $(\Delta f) = \alpha f$. For all $n \geq 1$, we have $f(n+1) - f(n) = \alpha f(n)$ or $f(n+1) = (\alpha+1)f(n)$. The general solution is then $f(1) = C$ (arbitrary) and $f(n) = (\alpha+1)^{n-1}C$. Here we are using the convention that $0^0 = 1$, so that the general solution to $\Delta f = -f$ is given by $C\delta_1$, where $\delta_1(1) = 1$, $\delta_1(n) = 0$ for all $n > 1$.

Example: The general solution to $\Delta^2 f = f$ is $n \mapsto C_1 2^n + C_2 \delta_1(n)$.

Example: The general solution to $\Delta^k f = f$ is the set of all real-valued functions in the complex span of $\langle (\zeta_k^i + 1)^n \rangle$, where $\zeta_k = e^{2\pi\sqrt{-1}/k}$ and $0 \leq i < k$.

4. Linear algebra of the discrete derivative

Let us reconsider the set V of all functions $f : \mathbb{Z}^+ \rightarrow \mathbb{R}$ from the perspective of linear algebra. As we saw above, we can naturally add two elements of V and also multiply any element of V by any real number. Moreover these operations satisfy all the usual algebraic properties, like commutativity and associativity of addition, and so forth. What we are trying to say is:

Proposition 9. *The set $V = \{f : \mathbb{Z}^+ \rightarrow \mathbb{R}\}$ of discrete functions has the structure of a vector space over $\mathbb{R}$.*

It is however a very large $\mathbb{R}$-vector space. In particular it is not finite-dimensional: for each $n \in \mathbb{Z}^+$, define a function δ_n which maps n to 1 and every other positive integer to 0. Then $\{\delta_n\}_{n=1}^{\infty}$ is an infinite linearly independent set.[2]

Moreover, the operators Δ, ∇ and Σ are all linear operators on the vector space V, and ∇ and Σ are mutually inverse.

Again, that's nice, but to compute things we would rather have a finite dimensional vector space. In fact we can define an infinite family of finite dimensional subspaces $\mathcal{P}_d \subset V$, as follows.

First, let $\mathcal{P} \subset V$ be the set of all polynomial functions. In other words, it is the span of the set of power functions $\{x^d \mid d \in \mathbb{N}\}$, where as before

$$x^d : n \mapsto n^d.$$

It is easy to check that the set of power functions is linearly independent, hence a basis for $\mathcal{P}$, so that $\mathcal{P}$ is again infinite dimensional. However, for $d \in \mathbb{N}$, define $\mathcal{P}_d$ to be the span of $1, x, \ldots, x^d$, i.e., the set of polynomial functions of degree at most d: evidently $\mathcal{P}_d$ is a $d+1$-dimensional subspace of V. Now they key fact:

Proposition 10. *Let $d \in \mathbb{Z}^+$. If $P(x)$ is a degree d polynomial function, then its discrete derivative $\Delta P : n \mapsto P(n+1) - P(n)$ is a degree $d-1$ polynomial function.*

Proof. First compute the discrete derivative of a monomial function $x^d : n \mapsto n^d$:

$$(\Delta x^d)(n) = (n+1)^d - n^d =$$

$$n^d + \binom{d+1}{1} n^{d-1} + \ldots + \binom{n+1}{n} n + 1 - n^d = \binom{d+1}{d} n^{d-1} + \ldots + \binom{n+1}{n} n + 1.$$

This is, as claied, a polynomial of degree $d-1$. From the linearity of Δ it follows that the discrete derivative of any polynomial function $P(x) = a_d x^d + \ldots + a_1 x + a_0$ is a sum of polynomials of degree at most $d-1$, so certainly it is a polynomial of degree *at most* $d-1$. What remains to be seen is that – assuming $a_d \neq 0$, of course – the coefficient of n^{d-1} in the polynomial $(\Delta P)(n) = P(n+1) - P(n)$ is

[2]Moreover, this set is not a basis: its span is the set of all functions which are zero for all sufficiently large n. But this is not a key point for us.

nonzero. Direct calculation shows that this coefficient is $(d+1)a_d - a_d = da_d$, which is nonzero since we assumed that $d > 0$ and $a_d \neq 0$. $\square$

An immediate consequence of Proposition 10 is

$$\Delta(\mathcal{P}_d) \subseteq \mathcal{P}_{d-1} \subseteq \mathcal{P}_d.$$

In particular, Δ is a linear operator on the finite-dimensional vector space $\mathcal{P}_d$. Moreover its kernel is the set of constant functions. Indeed, if for any discrete function f we have $f(n+1) - f(n) = 0$ for all n, then by induction we have $f(n) = f(1)$ for all n. (This is the discrete analogue of the fact that a function with identically zero derivative must be constant.) Therefore the kernel of Δ on $\mathcal{P}_d$ is one-dimensional. Recall the following fundamental fact of linear algebra: for any linear operator L on a finite-dimensional vector space W, we have

$$\dim(\ker(L)) + \dim(L(W)) = \dim L.$$

Therefore we find that $\dim(\Delta(\mathcal{P}_d)) = d$. On the other hand, we know from the proposition that the image $\Delta(\mathcal{P}_d)$ is contained in the d-dimensional subspace $\mathcal{P}_{d-1}$. Therefore we must have equality:

Theorem 11. *We have* $\Delta(\mathcal{P}_d) = \mathcal{P}_{d-1}$.

What is the significance of this? Applied to the function $x^{d-1} \in \mathcal{P}_d$, we get that there exists a degree d polynomial $P_d(x)$ such that for all $n \in \mathbb{Z}^+$,

$$\Delta(P_d)(n) = P_d(n+1) - P_d(n) = n^d.$$

Since $\Delta(P_d)(n) = \nabla(P_d)(n+1)$, we have also

$$\nabla(P_d)(n+1) = n^d,$$

and applying the summation operator Σ to both sides we get that for all $n \in \mathbb{Z}^+$,

$$P_d(n+1) = 1^d + \ldots + n^d.$$

Thus we've shown that there must be a nice closed form expression for the sum of the first n dth powers: indeed, it must be a polynomial of degree $d+1$.

Now we have undeniably done something worthwhile. But of course we won't stop here: we would like to actually compute the polynomials P_d!

But first we should address the following concern: given that it is ∇ and not Δ which is the inverse to Σ on V, why did we work first with Δ and only at the end "shift variables" to get back to ∇?

The answer is that the linear map ∇ does not map the space of polynomial functions to itself. It's close, but remember that ∇ of the nonzero constant function C is not zero: rather it is the function which is C at $n = 1$ and 0 for all larger values of n, but this is not a polynomial function of any degree. (Recall that a degree d polynomial function can be zero for at most d distinct real numbers.) So Δ it is.

Now we have a finite dimensional vector space $\mathcal{P}_d$, a linear operator Δ on $\mathcal{P}_d$, and a fixed basis $1, x, \ldots, x^d$ of $\mathcal{P}_d$. So we can write down a matrix representing Δ.

Note that the usual derivative operator $\frac{d}{dx}$ also carries $\mathcal{P}_d$ linearly onto $\mathcal{P}_{d-1}$ so gives a linear operator on $\mathcal{P}_d$. In this classical case the corresponding matrix is very simple: it is the matrix (**INSERT ME**).

The linear algebra of Δ is more complicated (and more interesting!). E.g., since

$$\Delta(1) = 0,$$

$$\Delta(x) = 1,$$

$$\Delta(x^2) : n \mapsto (n+1)^2 - n^2 = 2n + 1,$$

$$\Delta(x^3) : n \mapsto (n+1)^3 - n^3 = 3n^2 + 3n + 1,$$

$$\Delta(x^4) : n \mapsto (n+1)^4 - n^4 = 4n^3 + 6n^2 + 4 + 1,$$

the matrix representation of Δ on $\mathcal{P}_4$ with respect to the basis $1, x, x^2, x^3, x^4$ is

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For any d, let M_d be the matrix for Δ on the basis $1, x, x^2, \ldots, x^d$ of $\mathcal{P}_d$. Then

$$(M_d)_{ij} = \begin{cases} 0, & \text{if } i \geq j \\ \binom{j}{i}, & \text{if } i < j \end{cases}$$

To find the discrete antiderivative of the function x^{d-1} then, it suffices to solve the matrix equation $M[a_d, a_{d-1}, \ldots, a_0]^t = [0, \ldots, 0, 1, 0]^t$ for $a_0, \ldots, a_d$.

Moreover, because the matrix is in upper triangular form, we can easily solve the linear system by back substitution. For example, when $d = 4$ we get:

$$a_1 + a_2 + a_3 + a_4 = 0,$$

$$+2a_2 + 3a_3 + 4a_4 = 0,$$

$$3a_3 + 6a_4 = 0,$$

$$4a_4 = 1.$$

(The last equation reads $0 = 0$.) So we have $a_4 = \frac{1}{4}$, and then

$$a_3 = \frac{1}{3}(-6a_4) = \frac{-1}{2},$$

$$a_2 = \frac{1}{2}(-3a_3 - 4a_4) = \frac{1}{4},$$

$$a_1 = -a_2 - a_3 - a_4 = 0.$$

Note that the constant term a_0 is undetermined, as it should be. It follows from the above analysis that it doesn't matter what constant term we take, so we may as well take $a_0 = 0$. Thus

$$P_4(x) = \frac{1}{4}x^4 - \frac{1}{2}x^3 + \frac{1}{4}x^2,$$

and we easily calculate

$$P_4(n+1) = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4} = \left(\frac{n(n+1)}{2}\right)^2.$$

Since $(\Sigma\ x^3)(n) = P_4(n+1)$ for all n, we get the identity

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

5. More linear algebra of the discrete derivative

5.1. Finding a better basis.

As we have seen, the basis $1, x, \ldots, x^d$ is a very nice one for the linear operator $\frac{d}{dx}$ on the space $\mathcal{P}_d$: the matrix representation is ....

There is in fact an even better basis for $\frac{d}{dx}$ on $\mathcal{P}_d$: namely $1, x, \frac{x^2}{2}, \ldots, \frac{x^d}{d}$, because with respect to this basis the matrix is ..., a **shift operator**.

An important lesson in linear algebra is to find the best basis for the problem at hand. More specifically, given a linear operator T on a finite-dimensional vector space V over a field k, then under the assumption that all of the eigenvalues of V are elements of k, there exists a **Jordan basis** for V with respect to T, i.e., a basis in which V decomposes as a direct sum of T-stable subspaces W_i such that T_i restricted to each W_i is the sum of a shift operator and a scalar.

The fact that the discrete derivative of a degree k polynomial is a degree $k-1$ polynomial implies that the only eigenvalue of Δ on $\mathcal{P}_d$ is zero and that the 0-eigenspace is one-dimensional. This means that we have only one Jordan block, so that there exists a basis $p_0, \ldots, p_d$ of $\mathcal{P}_d$ with respect to which Δ is a shift operator: $\Delta(p_i) = p_{i-1}$ for $i > 0$, $\Delta(p_0) = 0$.

The constant polynomial 1 generates the 0-eigenspace, so we can take $p_0 = 1$. We want $p_1(x)$ such that $\Delta(p_1) = 1$; for this we can take $p_1(x) = x$. Next we want $p_2(x)$ such that $\Delta(p_2) = p_1$, and as we have seen, here $p_2 = \frac{1}{2}x^2$ is not quite right: $\Delta(\frac{1}{2}x^2)(n) = \frac{1}{2}(2x+1) = x + \frac{1}{2}$. So instead we need to take $p_2(x) = \frac{1}{2}x^2 - \frac{1}{2}x = \frac{x(x-1)}{2}$. Similarly, one finds that we can take $p_3(x) = \frac{x(x-1)(x-2)}{6}$. After enough computation, the pattern becomes clear:

Proposition 12. *Define $p_0(x) = 1$, and for $k \geq 1$, define $p_k(x) = \frac{x(x-1)\cdots(x-k+1)}{k!}$. Then $\Delta(p_0) = 0$ and for all $k \geq 1$, $\Delta(p_k) = p_{k-1}$.*

Proof. Notice that $p_k(x)$ is nothing else than the binomial coefficient $\binom{x}{k}$, viewed as a polynomial in x! The result we are trying to prove is then $\binom{x+1}{k} - \binom{x}{k} = \binom{x}{k-1}$, which is equivalent to the well-known binomial coefficient identity

$$\binom{x+1}{k} = \binom{x}{k-1} + \binom{x}{k}, \tag{10}$$

which for instance can be verified combinatorially: it is enough to show it for $x \in \mathbb{Z}^+$, in which case the left hand side is the number N of k-element subsets of $\{1, \ldots x+1\}$. Because every such subset either contains $x+1$ or it doesn't (and not both!), we have $N = N_1 + N_2$, where N_1 is the number of $k-1$ element subsets of $\{1, \ldots, x\}$ and N_2 is the number of k element subsets of $\{1, \ldots, x\}$. That is to say, $N_1 = \binom{x}{k-1}$ and $N_2 = \binom{x}{k}$. $\square$

Corollary 13. *a)* $\sum_1^{n-1} p_0(x) = p_1(n) - p_1(1) = p_1(x) - 1$.
b) For all $k \geq 1$, $\sum_1^{n-1} p_k(x) = p_{k+1}(n) - p_{k+1}(1) = p_{k+1}(n)$.

To stress the analogy between Δ and $\frac{d}{dx}$, it is common to define the **falling powers**

$$x^{\underline{k}} = (k!)p_k(x) = x(x-1)\cdots(x-k+1).$$

So we have found what linear algebra tells us is the optimal basis for $\mathcal{P}_d$ (and, in fact, for the infinite dimensional vector space $\mathcal{P}$). What benefits do we reap?

5.2. **The discrete Taylor series of a polynomial function.**
Another aspect of the philosophy of the optimal basis $\{b_1, \ldots, b_n\}$ for a vector space V is that, upon expressing v as a linear combination of the basis vectors:

$$v = a_1 b_1 + \ldots + a_n b_n, \ a_i \in \mathbb{R},$$

we expect that the coefficients a_i will be natural functions of v.

Example: Suppose V is equipped with an inner product $\langle , \rangle$ and we choose an orthonormal basis $b_1, \ldots, b_n$, then $a_i = \langle v, b_i \rangle$.

Example: Suppose that $V = \mathcal{P}_d$ and we choose the basis $b_0 = 1$, $b_i = \frac{x^i}{i!}$. We know there are unique real numbers $a_0, \ldots, a_d$ such that

$$P(x) = a_0 + a_1 x + a_2 \frac{x^2}{2} + a_3 \frac{x^3}{3!} + \ldots + a_d \frac{x^d}{d!}.$$

But this is just the Taylor series for $P(x)$. Explicitly, repeated evaluation at 0 and differentiation gives $a_k = \frac{d^k P}{dx}(0)$ for all $0 \leq k \leq d$.

Now we keep $V = \mathcal{P}_d$ but consider instead the natural basis $p_0(x), \ldots, p_d(x)$ for the discrete derivative Δ. We get that for any $P \in \mathcal{P}_d$ unique real numbers $a_0, \ldots, a_d$ such that

$$P(x) = a_0 + a_1 x + a_2 \frac{x(x-1)}{2} + \ldots + a_n \frac{x(x-1)\cdots(x-d+1)}{d!}.$$

Again, evaluating at $x = 0$ we find

$$a_0 = P(0).$$

Taking Δ of both sides gives

$$(\Delta P)(x) = a_1 + a_2 p_1(x) + \ldots + a_d p_{d-1} x,$$

and evaluating at 0 gives $a_1 = (\Delta P)(x)$. Continuing on in this way, we find that for all $0 \leq k \leq d$, $a_k = (\Delta^k P)(0)$. So we have shown

Theorem 14. *For any* $P(x) \in \mathcal{P}_d$, *we have* $P(x) =$

$$\sum_{k=0}^{d} (\Delta^k P)(0) \frac{x(x-1)\cdots(x-k+1)}{k!} = \sum_{k=0}^{d} \sum_{j=0}^{k} (-1)^j \binom{k}{j} P(k-j) p_k(x).$$

Applying this theorem with $P(x) = x^d$, we get:

$$P(n) = n^d = \sum_{k=0}^{d} \left(\sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^d \right) p_k(n).$$

Taking $\sum_1^{n-1}$ of both sides and applying Corollary 13, we get at last a closed form expression for an arbitrary power sum:

$$1^d + \ldots + (n-1)^d = \sum_{k=1}^{d} \left(\sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^d \right) \frac{n(n-1)\cdots(n-k)}{(k+1)!}. \tag{11}$$

5.3. **Integer-valued polynomials.**

First, the polynomials p_k have the following interesting property:

Proposition 15. *For every $k \in \mathbb{N}$ and $n \in \mathbb{Z}$, $p_k(n) \in \mathbb{Z}$.*

Note that if the polynomials $p_k(x)$ had coefficients in $\mathbb{Z}$, this would be obvious. But the leading coefficient of $p_k(x)$ is $\frac{x^k}{k!}$, so this is certainly not the case for $k \geq 2$. But nevertheless, for any integer n, we have $p_2(n) = \frac{n(n-1)}{2}$. Since either n or $n-1$ is an even integer, $p_2(n)$ is still an integer.

It is a fairly standard exercise to prove Proposition 15 by a double induction on n and k, using the basic binomal coefficient identity (10). But with the discrete calculus in hand we can give a more thematic proof. First:

Lemma 16. *For $f : \mathbb{Z} \to \mathbb{R}$ a discrete function, the following are equivalent:*
(i) For all $n \in \mathbb{Z}$, $f(n) \in \mathbb{Z}$.
(ii) For all $n \in \mathbb{Z}$, $(\Delta f)(n) \in \mathbb{Z}$, and there exists $N_0 \in \mathbb{Z}$ such that $f(N_0) \in \mathbb{Z}$.

The proof of Lemma 16 is immediate and is left to the reader.

Moreover, Proposition 15 swiftly follows: certainly $p_0(x) = 1$ maps integers to integers. Since $\Delta p_k = p_{k-1}$ and $p_k(0) = 0$ for all k, we're done by induction on k.

Remark: For those of a number-theoretic bent, we note the following restatement of Proposition 15: the product of any k consecutive integers is divisible by $k!$.

In general, we say a polynomial $f(x) \in \mathbb{R}[x]$ is **integer-valued** if for all $n \in \mathbb{Z}$, $f(n) \in \mathbb{Z}$. It is not hard to see that this implies that $f(x) \in \mathbb{Q}[x]$ (i.e., all coefficients of f are rational numbers), using e.g. the Lagrange Interpolation Formula. But in fact a much more precise and beautiful result holds.

Theorem 17. *Let $f(x) \in \mathbb{R}[x]$ be a polynomial of degree d. There are unique $a_0, \ldots, a_d \in \mathbb{R}$ such that $f(x) = a_0 p_0(x) + \ldots + a_d p_d(x)$. Moreover, TFAE:*
(i) All of the coefficients $a_0, \ldots, a_d$ lie in $\mathbb{Z}$.
(ii) $f(x)$ is integer-valued.
(iii) There is $N_0 \in \mathbb{Z}$ such that $f(N_0)$, $f(N_0+1), \ldots, f(N_0+d)$ are all integers.

Proof. The existence of unique $a_0, \ldots, a_d$ such that $f(x) = a_0 p_0(x) + \ldots a_d p_d(x)$ merely expresses the fact that $p_0, \ldots, p_d$ gives a basis for $\mathcal{P}_d$.
(i) $\implies$ (ii): By Proposition 15 we know that each $p_d(x)$ is integer-valued, and a $\mathbb{Z}$-linear combination of integer-valued polynomials is integer-valued.
(ii) $\implies$ (iii) is immediate.
(iii) $\implies$ (i): Suppose first that $N_0 = 0$. By Theorem 14 we have that the a_k's are

precisely the Taylor coefficients at 0:

$$a_k = (\Delta^k f)(0) = \sum_{i=0}^{d} (-1)^k \binom{k}{i} f(k-i).$$

Since $f(0), \ldots, f(d) \in \mathbb{Z}$ by hypothesis, evidently $a_k \in \mathbb{Z}$. The case of arbitary N_0 follows easily by a change of variables argument: for instance, by making the evident generalization of Theorem 14 to Taylor series expansions about N_0. We leave the details to the interested reader. $\square$

5.4. Stirling numbers and a formula for $S_d(n)$.