

Finite Calculus: A Tutorial for Solving Nasty Sums

David Gleich

January 17, 2005

Abstract

In this tutorial, I will first explain the need for finite calculus using an example sum I think is difficult to solve. Next, I will show where this sum actually occurs and why it is important. Following that, I will present all the mathematics behind finite calculus and a series of theorems to make it helpful before concluding with a set of examples to show that it really is useful.

Contents

1	How to Evaluate $\sum_{x=1}^n x^2$?	2
2	The Computational Cost of Gaussian Elimination	3
3	The Finite Calculus	4
3.1	Discrete Derivative	5
3.2	The Indefinite Sum and the Discrete Anti-Derivative	7
3.3	Helpful Finite Calculus Theorems	8
4	Making Finite Calculus Useful: Stirling and His Numbers	10
4.1	Stirling Numbers (of the Second Kind)	12
4.2	Proving the Theorem	14
4.3	Computing Stirling Numbers	15
5	Examples Galore...	16
5.1	$\sum_{x=1}^n x^2$	16
5.2	A Double Sum	16
5.3	Average Codeword Length	17
6	Conclusion	18

1 How to Evaluate $\sum_{x=1}^n x^2$?

One of the problems we'll¹ learn to address during this tutorial is how to mechanically (i.e. without much thinking) compute the closed form value of the sum

$$\sum_{x=1}^n x^2. \tag{1}$$

While many students may already know the closed form answer to this particular question, our quest for the answer will lead us to techniques to easily evaluate nasty sums such as

$$\sum_{x=1}^n \sum_{y=1}^m (x+y)^2 \tag{2}$$

and

$$\sum_{x=0}^n x2^x. \tag{3}$$

Since we'll be using $\sum_{x=1}^n x^2$ as a motivating example, we'll first familiarize ourselves with the first few values of this function.

n	1	2	3	4	5	6
$\sum_{x=1}^n x^2$	1	5	14	30	55	91

Now, we'll try a few techniques to evaluate this sum before honing in on the correct answer.

The astute student reading this tutorial may have surmised that we'll make some connection with calculus at some point. Hence, let's see what happens if we just "pretend" that this was an integral from calculus instead. Switching the \sum to a \int sign gives

$$\sum_{x=1}^n x^2 \stackrel{?}{=} \int_1^n x^2 dx.$$

Using calculus, we can immediately solve the integral formulation.

$$\sum_{x=1}^n x^2 \stackrel{?}{=} \frac{n^3}{3} - \frac{1}{3},$$

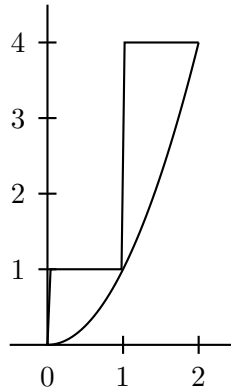
Sadly, trying $n = 2$ shows us that this is wrong.

$$\sum_{x=1}^2 x^2 \stackrel{?}{=} \frac{2^3}{3} - \frac{1}{3} = \frac{7}{3} \neq 5.$$

Graphically, we can see what went wrong in this derivation.

¹Since I like to keep things informal, if I say we in the text, I really mean you and I.

²Interesting trivia: The integral symbol was actually a long letter S for "summa" [3]. So such a substitution makes sense at least alphabetically.



We want the area underneath the stepped line, not the area underneath the smooth curve.

It is ironic that we can easily determine the area underneath the smooth curve using the infinite calculus but have trouble determining the much more simple area under the stepped line. In short, that is the goal of this tutorial – to derive a finite (or discrete) analogue of infinite calculus so the finite sum

$$\sum_{x=1}^2 x^2$$

is no more difficult to solve than the “infinite” sum

$$\int_1^2 x^2 dx.$$

Besides finite calculus, another way to compute the value of $\sum_{x=1}^n x^2$ is to look it up in a book, or try and guess the answer. Searching through CRC Standard Mathematical Tables and Formule [6], on p. 21 we find

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

I’m not sure I would have been able to guess that answer myself.

2 The Computational Cost of Gaussian Elimination

Now we’ll venture off onto a bit of a tangent. Because I like to keep things applied, I’d like to show that our example sum occurs in practice. It is not an equation that popped out of the aether as an interesting mathematical identity, but rather emerges during a computational analysis of Gaussian Elimination – a fundamental algorithm in linear algebra. The Gaussian Elimination algorithm accepts an $m \times m$ matrix A and computes a lower triangular matrix L and an upper triangular matrix U such that $A = LU$.

Algorithm 1 Gaussian Elimination

Require: $A \in \mathbb{R}^{m \times m}$ $U = A, L = I$ **for** $k = 1$ **to** $m - 1$ **do** **for** $j = k + 1$ **to** m **do** $l_{jk} = u_{jk}/u_{kk}$ $u_{j,k:m} \leftarrow u_{j,k:m} - l_{jk}u_{k,k:m}$ **end for****end for**

In this tutorial, I won't give a full derivation of Gaussian Elimination, but simply give the algorithm (from [4]) and analyze its computational cost.

The computational cost of an algorithm is the number of addition, subtraction, multiplication, and division operations performed during the algorithm. To analyze this cost, we'll look at how much work is done at each step of Gaussian Elimination, that is, each iteration of the outer loop. In the case when $k = 1$, then we let j go between 2 and m . In the inner loop, we perform 1 division ($l_{jk} = u_{jk}/u_{kk}$), m multiplications ($l_{jk}u_{k,k:m}$), and m subtractions. Since we run this inner loop $m - 1$ times, we have

$$2m(m - 1) + (m - 1) = 2m^2 - 1 \leq 2m^2$$

work on the first step. On the second step, we let j go between 3 and m . In the inner loop, we now perform 1 division, $m - 1$ multiplications, and $m - 1$ subtractions, for a total of

$$2(m - 2)(m - 1) + (m - 2) \leq 2(m - 1)^2.$$

This pattern continues for each of the following outer steps.

Thus, the total runtime of Gaussian Elimination is less than

$$2m^2 + 2(m - 1)^2 + 2(m - 2)^2 + \dots + 1 = \sum_{x=1}^m 2x^2 = 2 \sum_{x=1}^m x^2.$$

Since we looked up the answer to this sum in the previous section, we know that Gaussian Elimination takes less

$$2 \sum_{x=1}^m x^2 = 2 \frac{m(m + 1)(2m + 1)}{6} = \frac{2}{3}m^3 + m^2 + \frac{1}{3}m$$

total additions, subtractions, multiplications, and divisions. Now, let's start determining how we can compute $\sum_{x=1}^n x^2$ ourselves.

3 The Finite Calculus

Before launching into the derivation of finite calculus, I should probably explain what it is. By analogy with infinite calculus, we seek a "closed form" expression for sums of the form

$$\sum_{x=a}^b f(x).$$

By closed form, we mean involving some set of fundamental operations, such as addition, multiplication, exponentiation, and even factorials. Whereas in infinite calculus, we needed to compute the area under a function exactly, in finite calculus we want to compute the area under a sequence exactly. We call this finite calculus because each sum is composed of a finite (rather than infinite) set of terms. One way to think of finite calculus is calculus on the integers rather than the real numbers.

In calculus, we used the notion of derivative and anti-derivative along with the fundamental theorem of calculus to write the closed form solution of

$$\int_a^b f(x) dx = F(b) - F(a),$$

where

$$\frac{d}{dx}F(x) = f(x).$$

Our immediate goal for finite calculus (and this section) is to develop a fundamental theorem of finite calculus of a similar form.

3.1 Discrete Derivative

Before we can hope to find the fundamental theorem of finite calculus, our first goal is to develop a notion of “derivative.” Recall from calculus that

$$\frac{d}{dx}f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Since we aren’t allowing real numbers, the closest we can get to 0 is 1 (that is, without actually getting to 0). Then, the discrete derivative is

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

(As an aid to those readers who are skimming this tutorial, we’ll repeat this and future definitions in a formal definition statement.)

Definition (Discrete Derivative). *The discrete derivative of $f(x)$ is defined as*

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

What can we do with our new derivative? From calculus, we found that it was simple to take the derivative of powers $f(x) = x^m$.

$$\frac{d}{dx}x^m = mx^{m-1}.$$

Hopefully, we’ll find such a simple derivative for finite powers as well.

$$\Delta x^m = (x+1)^m - x^m \stackrel{?}{=} mx^{m-1}.$$

A quick check shows that

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

Unfortunately, this simple form does not work for x^2 .

$$\Delta x^2 = (x + 1)^2 - x^2 = 2x + 1 \neq 2x.$$

Now, our derivate for x^2 was only off by 1. Is there any easy way we can fix this problem? If we rewrite our previous failed derivative as

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

we find ourselves must closer. Now, we can pull the -1 into the derivative by observing that $\Delta(x^2 - x) = 2x + 1 - 1 = 2x$. We can factor $x^2 - x$ as $x(x - 1)$. Thus, we have

$$\Delta(x(x - 1)) = 2x.$$

I'll leave it to the reader to verify that

$$\Delta(x(x - 1)(x - 2)) = 3x(x - 1).$$

It appears we've found a new type of power for discrete derivatives! These powers are known as "falling powers."

Definition (Falling Power). *The expression x to the m falling is denoted $x^{\underline{m}}$.³ The value of*

$$x^{\underline{m}} = x(x - 1)(x - 2) \cdots (x - (m - 1)).$$

Using the falling powers, we'll now prove that falling powers are analogous to regular powers in finite calculus.

Theorem 3.1. *The discrete derivative of a falling power is the exponent times the next lowest falling power. That is,*

$$\Delta x^{\underline{m}} = m x^{\underline{m-1}}.$$

Proof. The proof is just algebra.

$$\begin{aligned} \Delta x^{\underline{m}} &= (x + 1)^{\underline{m}} - x^{\underline{m}} \\ &= (x + 1)x(x - 1) \cdots (x - m + 2) - x(x - 1) \cdots (x - m + 2)(x - m + 1). \\ &= (x + 1 - x + m - 1)x(x - 1) \cdots (x - m + 2) \\ &= m x^{\underline{m-1}}. \end{aligned}$$

□

Now, we'll prove a few other useful theorems.

Theorem 3.2. *The discrete derivative of the sum of two functions is the sum of the discrete derivatives of the functions themselves.*

$$\Delta(f(x) + g(x)) = \Delta f(x) + \Delta g(x)$$

³Note that other authors use the notation $(x)_m$ to denote the falling power [5].

Proof. The proof is straight from the definition.

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

Rearranging terms, we get

$$\Delta(f(x) + g(x)) = \Delta f(x) + \Delta g(x). \quad \square$$

Theorem 3.3. *The discrete derivative of a constant times a function is the constant times the discrete derivative of the function.*

$$\Delta(cf(x)) = c\Delta f(x).$$

Proof. Simply factor out the constant from the application of the definition of the discrete derivative. \square

3.2 The Indefinite Sum and the Discrete Anti-Derivative

With some handle on what our finite derivative does, let's move on to discrete integration. First, we'll define some notation.

Definition (Discrete Anti-Derivative). *A function $f(x)$ with the property that $\Delta f(x) = g(x)$ is called the discrete anti-derivative of g . We denote the class of functions satisfying this property as the indefinite sum of $g(x)$,*

$$\sum g(x) \delta x = f(x) + C,$$

where C is an arbitrary constant. $\sum g(x) \delta x$ is also called the indefinite sum of $g(x)$.

The discrete anti-derivative corresponds to the anti-derivative or indefinite integral from calculus.

$$\int g(x) dx = f(x) + c \qquad \sum g(x) \delta x = f(x) + C$$

So far, all we've done is establish some notation. We have not shown any anti-derivatives yet. However, we can easily do just that. Recall that

$$\Delta x^m = mx^{m-1}.$$

Using this fact, along with Theorems 3.2, 3.3, allows us to state that

$$\sum x^m \delta x = \frac{x^{m+1}}{m+1} + C.$$

Now, let's work toward a fundamental theorem of finite calculus. First, we need to define the notion of a discrete definite integral or definite sum.

Definition (Discrete Definite Integral). *Let $\Delta f(x) = g(x)$. Then*

$$\sum_a^b g(x) \delta x = f(b) - f(a).$$

With the discrete definite integral, the theorem we'd like to show (by analogy with infinite calculus) is

$$\sum_a^b g(x) \delta x \stackrel{?}{=} \sum_{x=a}^b g(x).$$

Unfortunately, this isn't true as a quick check shows.

$$\sum_1^5 x \delta x = \left. \frac{x^2}{2} \right|_1^5 = \frac{(5)(4)}{2} = 10$$

But, $\sum_{x=1}^5 x = 15$, so the theorem isn't true. However, we were very close. Note that

$$\sum_1^5 x \delta x = 10 = \sum_{x=1}^4 x.$$

This revised form gives us the correct fundamental theorem of finite calculus.

Theorem 3.4. *The fundamental theorem of finite calculus is*

$$\sum_a^b g(x) \delta x = \sum_{x=a}^{b-1} g(x).$$

Proof. Here we have more algebra for the proof. Let $\Delta f(x) = f(x+1) - f(x) = g(x)$.

$$\begin{aligned} \sum_{x=a}^{b-1} g(x) &= \sum_{x=a}^{b-1} f(x+1) - f(x) \\ &= f(a+1) - f(a) + f(a+2) - f(a+1) + \dots \\ &\quad + f(b) - f(b-1) \\ &= f(b) - f(a), \end{aligned}$$

after the telescoping sum collapses.

Now $f(b) - f(a) = \sum_a^b g(x) \delta x$ by definition. □

3.3 Helpful Finite Calculus Theorems

Now that we have our fundamental theorem, this section is just a collection of theorems to make finite calculus useful. The uninterested reader can skip to Table 1 at the end.

One of the more useful functions from calculus is $f(x) = e^x$. This special function has the property that

$$D(e^x) = e^x \quad \text{and} \quad \int e^x dx = e^x + C.$$

Our first theorem is that there is an analogous function in the finite calculus – a function whose derivative is itself. To find it, let's "round" e . If we do this right, the analog of e should either be 2 or 3, right? Let's see which one works.

$$\Delta(2^x) = 2^{x+1} - 2^x = 2 \cdot 2^x - 2^x = (2-1)2^x = 2^x.$$

$$\Delta(3^x) = 3^{x+1} - 3^x = 3 \cdot 3^x - 3^x = (3-1)3^x = 2 \cdot 3^x.$$

Thus, e corresponds to the integer 2.

Theorem 3.5. *The function 2^x satisfies*

$$\Delta(2^x) = 2^x \quad \text{and} \quad \sum 2^x \delta x = 2^x + C.$$

Let's compute the general derivative of an exponent.

$$\Delta(c^x) = c^{x+1} - c^x = (c-1)c^x.$$

Because c is a constant in this expression, we can then immediately compute the anti-derivative as well.

$$\sum c^x \delta x = \frac{c^x}{c-1} + C.$$

One of the important theorems in infinite calculus is the formula for $\frac{d}{dx}(u(x)v(x))$. Let's find the corresponding formula for finite calculus.

$$\begin{aligned} \Delta(u(x)v(x)) &= u(x+1)v(x+1) - u(x)v(x) \\ &= u(x+1)v(x+1) - u(x)v(x+1) + u(x)v(x+1) - u(x)v(x) \\ &= v(x+1)\Delta u(x) + u(x)\Delta v(x). \end{aligned}$$

Now, we can use this derivative to write a discrete integration by parts formula.

Theorem 3.6.

$$\sum u(x)\Delta v(x) \delta x = u(x)v(x) - \sum v(x+1)\Delta u(x) \delta x.$$

Proof. If we take the anti-derivative on both sides of

$$\Delta(u(x)v(x)) = u(x)\Delta v(x) + v(x+1)\Delta u(x)$$

we get

$$u(x)v(x) = \sum u(x)\Delta v(x) \delta x + \sum v(x+1)\Delta u(x) \delta x.$$

Rearranging this equation yields the theorem. □

There are three facts I'd like to finish up with. First, let's look at the derivative and anti-derivative of some combinatorial functions. Second, does our formula for falling powers and their derivative also work for negative powers? Finally, what about x^{-1} ?

The binomial coefficients often arise in Combinatorics. Let's compute $\Delta\binom{x}{k}$. If we use the well-known formula, $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \implies \binom{n}{k} - \binom{n-1}{k} = \binom{n-1}{k-1}$, we can easily show that

$$\Delta\binom{x}{k} = \binom{x+1}{k} - \binom{x}{k} = \binom{x}{k-1}.$$

We can rewrite this as

$$\sum \binom{x}{k} \delta x = \binom{x}{k+1} + C.$$

Since the falling power is only defined for positive exponent, you may have been wondering about negative falling powers. Let's handle them now. We can derive a definition by looking at a pattern in the falling powers.

$$x^{\underline{3}} = x(x-1)(x-2),$$

$$x^{\underline{2}} = x(x-1),$$

$$x^{\underline{1}} = x,$$

$$x^{\underline{0}} = 1$$

To move from x^3 to x^2 we divide by $(x - 2)$. Then we divide by $(x - 1)$ to go from x^2 to x^1 . When we continue this pattern, we get

$$x^{-1} = x^0/(x + 1) = \frac{1}{x + 1}$$

$$x^{-2} = x^{-1}/(x + 2) = \frac{1}{(x + 1)(x + 2)}$$

Definition (Negative Falling Power).

$$x^{-m} = \frac{1}{(x + 1)(x + 2) \cdots (x + m)}$$

Let's check our derivative rule for negative falling powers.

$$\begin{aligned} \Delta(x^{-m}) &= \frac{1}{(x + 2)(x + 3) \cdots (x + m + 1)} - \frac{1}{(x + 1)(x + 2) \cdots (x + m)} \\ &= \frac{(x + 1) - (x + m + 1)}{(x + 1)(x + 2) \cdots (x + m + 1)} \\ &= mx^{-m-1}. \end{aligned}$$

Amazingly, I haven't led you astray here. It does! This fact implies that our formula for the discrete anti-derivative for falling powers holds for negative falling powers as well.

Well, our theorem holds, unless $m = -1$. In that case, we'd get $\sum x^{-1} = \frac{x^0}{0}$, which is a problem. Recall from infinite calculus, we had to appeal to the $\log(x)$ function to integrate $\frac{1}{x}$. There is no easy way of getting the correct function intuitively.

Theorem 3.7. *Let*

$$H_x = \underbrace{\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{x}}_x,$$

for integer x . The function H_x is the antiderivative of $\frac{1}{x}$.

Proof.

$$\Delta(H_x) = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{x+1} - \frac{1}{1} - \frac{1}{2} - \cdots - \frac{1}{x} = \frac{1}{x+1} \quad \square$$

4 Making Finite Calculus Useful: Stirling and His Numbers

If you've gotten this far, you're probably thinking. "Great David, we've shown all these really neat facts about finite calculus. But, you still haven't shown me how to solve

$$\sum_{x=1}^n x^2$$

as you promised way back in the introduction." Well, let's do that now!

$f(x)$	$\Delta f(x)$	$\sum f(x) \delta x$
x^m	mx^{m-1}	$\frac{x^{m+1}}{m+1}$
x^{-1}	$-x^{-2}$	H_x
2^x	2^x	2^x
c^x	$(c-1)c^x$	$\frac{c^x}{c-1}$
$\binom{x}{m}$	$\binom{x}{m-1}$	$\binom{x}{m+1}$
$u(x) + v(x)$	$\Delta u(x) + \Delta v(x)$	$\sum u(x) \delta x + \sum v(x) \delta x$
$u(x)v(x)$	$u(x)\Delta v(x) + v(x+1)\Delta u(x)$	$u(x)v(x) - \sum v(x+1)\Delta u(x) \delta x$

Table 1: Here we give a set a useful finite calculus theorems. These should make solving even nasty sums easy (or at least tractable). All of the discrete anti-derivatives have the constant omitted for brevity.

In order to solve regular powers, we need to find some way to convert between x^m and $x^{\underline{m}}$ so we can use our integration theorems.

Let's see all the conversions we can come up with ourselves.

$$\begin{aligned}x^0 &= x^{\underline{0}} \\x^1 &= x^{\underline{1}} \\x^2 &= x^{\underline{2}} + x^{\underline{1}} \\x^3 &= ???\end{aligned}$$

So we find that the first few powers were easy to convert, but then the higher powers are not so obvious. Let's see if we can work with

$$x^3 = ax^{\underline{3}} + bx^{\underline{2}} + cx^{\underline{1}}$$

to come up with a formula for a, b , and c .

$$\begin{aligned}ax^{\underline{3}} + bx^{\underline{2}} + cx^{\underline{1}} &= [a(x)(x-1)(x-2)] + [b(x)(x-1)] + [c(x)] \\&= [ax^3 - 3ax^2 + 2ax] + [bx^2 - bx] + [cx] \\&= ax^3 + (b-3a)x^2 + (2a-b+c)x.\end{aligned}$$

If we want

$$ax^{\underline{3}} + bx^{\underline{2}} + cx^{\underline{1}} = x^3,$$

then we need the following coefficients

$$a = 1 \quad (b-3a) = 0 \quad (2a-b+c) = 0,$$

or $a = 1, b = 3$, and $c = 1$. Hence, we have

$$x^3 = x^{\underline{3}} + 3x^{\underline{2}} + x^{\underline{1}}.$$

That was a lot of work just for x^3 . I'm not going to attempt x^4 , but feel free if you are so inclined. Instead, I'll devote the rest of the section to proving the following theorem.

Theorem 4.1. *We can convert between powers and falling powers using the following formula*

$$x^m = \sum_{k=0}^m \left\{ \begin{matrix} m \\ k \end{matrix} \right\} x^{\underline{k}},$$

where $\left\{ \begin{matrix} m \\ k \end{matrix} \right\}$ is a Stirling number of the second kind.

Okay, that theorem was a lot to digest at once. Before we even start to prove it, let's pick it apart. What the theorem is saying is that we can convert any power into a sum of falling powers as long as we can compute these special Stirling numbers (I'll get to these next). So, if we wanted x^4 in terms of falling powers, this theorem tells us the answer:

$$x^4 = \left\{ \begin{matrix} 4 \\ 0 \end{matrix} \right\} x^{\underline{0}} + \left\{ \begin{matrix} 4 \\ 1 \end{matrix} \right\} x^{\underline{1}} + \left\{ \begin{matrix} 4 \\ 2 \end{matrix} \right\} x^{\underline{2}} + \left\{ \begin{matrix} 4 \\ 3 \end{matrix} \right\} x^{\underline{3}} + \left\{ \begin{matrix} 4 \\ 4 \end{matrix} \right\} x^{\underline{4}}.$$

Let's address the Stirling numbers now.

4.1 Stirling Numbers (of the Second Kind)

If you haven't seen them before, you are now probably wondering about these Stirling numbers. To start, I'll just define them.

Definition (Stirling Numbers). *The Stirling numbers of the second kind,⁴ denoted*

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\},$$

count the number of ways of partitioning n distinct objects into k non-empty sets.

I'm guessing that didn't help much and you are still scratching your head. Hopefully working with some of these numbers will help. First, let's try and compute $\left\{ \begin{matrix} 1 \\ 1 \end{matrix} \right\}$. This is the number of ways of partitioning 1 item into 1 non-empty set. I think it's pretty clear that there is only one way to do this. So $\left\{ \begin{matrix} 1 \\ 1 \end{matrix} \right\} = 1$. What about $\left\{ \begin{matrix} 1 \\ 0 \end{matrix} \right\}$, or the number of ways to partition 1 item into 0 non-empty sets? Since each of the sets has to be non-empty, there are 0 ways of doing this, $\left\{ \begin{matrix} 1 \\ 0 \end{matrix} \right\} = 0$. We can generalize this argument and show that

$$\left\{ \begin{matrix} n \\ 0 \end{matrix} \right\} = 0$$

if $n > 0$. And what about $\left\{ \begin{matrix} 0 \\ 0 \end{matrix} \right\}$? We want to divide 0 things into 0 non-empty sets. Shouldn't this be 0 too? Strangely, no, $\left\{ \begin{matrix} 0 \\ 0 \end{matrix} \right\} = 1$. If you have trouble accepting this fact logically, just consider it a definition.

Another set of simple Stirling numbers are $\left\{ \begin{matrix} n \\ n \end{matrix} \right\}$. These are the number of ways of dividing n objects into n non-empty sets. Since the number of objects equals the number of sets, there can only be one object in each set, so $\left\{ \begin{matrix} n \\ n \end{matrix} \right\} = 1$.

Now, let's look at a more complicated Stirling number, $\left\{ \begin{matrix} 3 \\ 2 \end{matrix} \right\}$. Suppose our three items are ♣, ♦, and ♥. When we divide these three card suits into 2 non-empty sets

⁴There is another set of numbers called *Stirling numbers of the first kind* which are slightly different. See [5] for more information.

we get the following sets.

$$\begin{aligned} &\{\clubsuit, \diamond\}, \{\heartsuit\} \\ &\{\clubsuit, \heartsuit\}, \{\diamond\} \\ &\{\diamond, \heartsuit\}, \{\clubsuit\} \end{aligned}$$

So $\left\{ \begin{smallmatrix} 3 \\ 2 \end{smallmatrix} \right\} = 3$, since there are three ways to put 3 items into 2 non-empty sets.

We still don't really know much about Stirling numbers though. Let's prove a theorem to help us general new Stirling numbers from previous ones.

Theorem 4.2.

$$\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix} \right\} + k \left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}.$$

Proof. We'll give a combinatorial proof of this identity. In a combinatorial proof, we choose a particular quantity and give two ways of counting this quantity. Since these two different ways count the same thing, they must be equal. In this proof, we'll count the number of ways to partition n distinct objects into k non-empty sets.

The obvious way to count this quantity is by using our previous definition of Stirling numbers. By this definition, $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ counts the number of ways to partition n distinct objects into k non-empty sets.

Now, let's count this quantity using a different way. I like to call this the "weirdo" method. Let's pick out one of our distinct objects and call it x , the "weirdo." We can now count the number of ways to partition n distinct objects into k non-empty sets by keeping track of where the "weirdo" goes. The first possibility is that the "weirdo" is in a set all by itself. In this case, there are $\left\{ \begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix} \right\}$ ways for the other $n-1$ objects to be partitioned into $k-1$ non-empty sets.

If the "weirdo" is not in a set by itself, then it must be inside another non-empty set. So the others must have been divided into k non-empty sets. There are $\left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}$ ways of partitioning them. For each of these partitions, we can insert the "weirdo" into any of the k sets and get another partition. So there are $k \left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}$ total ways putting the "weirdo" into these existing sets. Since we have completely dealt with the "weirdo", there are $\left\{ \begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix} \right\} + k \left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}$ ways to partition n distinct objects into k non-empty sets.

Thus,

$$\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix} \right\} + k \left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}. \quad \square$$

This proof is a little challenging. Let me just go through it with our 3 items from before. Let \clubsuit be the "weirdo." (It does look a little strange, doesn't it?). If \clubsuit is in a set by itself, then we need to partition \diamond, \heartsuit between the 1 remaining set. There is no way to divide them between one set, so there is just one way of putting them in a set together. We get the following partition

$$\{\diamond, \heartsuit\}, \{\clubsuit\}.$$

The other case was that \clubsuit was not by itself. So we partition \diamond, \heartsuit into 2 sets. There's still just one way of doing this $\{\diamond\}, \{\heartsuit\}$. We can insert \clubsuit into either of the sets in this partition, so we get 2 more ways of partitioning the items into two sets.

$$\begin{aligned} &\{\clubsuit, \diamond\}, \{\heartsuit\} \\ &\{\clubsuit, \heartsuit\}, \{\diamond\} \end{aligned}$$

Using this theorem, we can build the Stirling number triangle (like Pascal's triangle). See Figure 1.

				1			
			0	1			
		0	1	1			
	0	1	3	1			
	0	1	7	6	1		
	0	1	15	25	10	1	
0	1	31	90	65	15	1	

Figure 1: A table of Stirling Numbers of the second kind $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ which induces the Stirling triangle. The value of $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ is the $k + 1$ th number on the $n + 1$ th row, e.g. $\left\{ \begin{smallmatrix} 4 \\ 2 \end{smallmatrix} \right\} = 7$. To build the triangle, use Theorem 4.2. This theorem says that the value in a particular position is k times the number up and to the right, plus the number up and to the left, e.g. $7 = 2 * 3 + 1$.

4.2 Proving the Theorem

In order to prove Theorem 4.1, we'll first prove a helpful lemma, then do a lot of algebra.

Lemma. *A helpful lemma*

$$x \cdot x^k = x^{k+1} + kx^k.$$

Proof.

$$\begin{aligned} x \cdot x^k &= x \cdot x^k - k \cdot x^k + k \cdot x^k \\ &= (x - k) \cdot x^k + k \cdot x^k \\ &= x^{k+1} + k \cdot x^k. \end{aligned}$$

□

Now, recall Theorem 4.1.

$$x^m = \sum_{k=0}^m \left\{ \begin{smallmatrix} m \\ k \end{smallmatrix} \right\} x^k,$$

Proof. We'll prove this by induction on n .

The base case, $n = 1$ is trivial. $x^1 = \left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\} x^0 + \left\{ \begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right\} x^1 = x$.

Our induction hypothesis is that $x^n = \sum_{k=0}^n \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} x^k$ holds for all $n \leq r$.

(See the end for a discussion of the steps in the following work.)

$$\begin{aligned}
x^{r+1} &= xx^r \\
&= \sum_{k=0}^r \binom{r}{k} x \cdot x^k \\
&= \sum_{k=0}^r \binom{r}{k} [x^{k+1} + kx^k] \\
&= \sum_{k=1}^{r+1} \binom{r}{k-1} x^k + \sum_{k=0}^r \binom{r}{k} kx^k \\
&= \binom{r}{r} x^{r+1} + \sum_{k=1}^r \left[\binom{r}{k-1} x^k + \binom{r}{k} kx^k \right] + \binom{r}{0} x^0 \\
&= x^{r+1} + \sum_{k=0}^r \binom{r+1}{k} x^k \\
&= \sum_{k=0}^{r+1} \binom{r+1}{k} x^k.
\end{aligned}$$

In the first step, we applied our induction hypothesis, then we moved the x inside the sum. Following that, we applied the lemma we just proved. We then split the sum into one sum based on $k+1$ and the other based on k . Then, we pulled out the $\binom{r}{r} = 1$ and $\binom{r}{0} = 0$ terms from the sum. When we put things back together in the sum, we use Theorem 4.2. \square

Excellent! Now we have a means to convert between x^m and falling powers!

4.3 Computing Stirling Numbers

We can derive a formula to compute Stirling numbers without the recurrence relation derived in Theorem 4.2. This derivation is somewhat complex, and I'd encourage readers to skip it unless they are really interested. The important part of this section is the statement of the theorem.

Theorem 4.3.

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^n.$$

Proof. We'll prove this using a combinatorial argument. The objects we are counting are the number of surjections between an n -set X and a k -set Y . Let $Y = \{y_1, y_2, \dots, y_k\}$. Recall that a surjective mapping is a mapping onto the set Y , so some element in X must map to each element in Y .

First, we can count this using $k! \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$. To see this fact, consider that after the partitioning of n into k non-empty subsets, we have a partition such as the following one

$$\{x_1, x_5, x_9\}, \{x_2, x_4\}, \dots$$

If we apply an ordering to the sets comprising these partitions, i.e. $\pi_1, \pi_2, \dots, \pi_k$, then we can construct a surjective map from X to Y using the map

$$f(x) = y_j \mid x \in \pi_j,$$

that is, the set π_j is the inverse image of y_j . Since there are $k!$ orderings of the numbers between 1 and k , we get that there are $k! \binom{n}{k}$ surjections.

We can also count this using an inclusion-exclusion argument. If we take the total number of mappings from X to Y , we want to subtract all the mappings that do not map to at least one element in Y . We can do this by look at all maps that exclude a particular element y_i . To do this, we take the total number of maps between X and Y and subtract the number of maps that exclude an element y_i . However, now we've subtracted some maps twice (maps that exclude y_i and y_j), and we then need to add them back. So we then add the number of maps that exclude two elements, then subtract the maps that exclude at least three (for a similar reason), etc. This strange sum actually counts all of our elements. For more information about a powerful generalization of this proof technique, see [5].

There are $\binom{k}{i}$ ways to pick the i elements from k to exclude and there are $(k-i)^n$ maps that exclude one set of i elements from Y . Thus, we get the number of surjective maps from X to Y is

$$\underbrace{\binom{k}{0} k^n}_{\text{all maps}} - \underbrace{\binom{k}{1} (k-1)^n}_{\text{exclude 1}} + \underbrace{\binom{k}{2} (k-2)^n}_{\text{exclude 2}} - \dots = \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^n.$$

We divide both sides by $k!$ to complete the proof. □

5 Examples Galore...

In this section, we just provide a few in-depth examples to show the utility of finite calculus and how it really takes the thought out of complicated sums.

5.1 $\sum_{x=1}^n x^2$

Since I still haven't gotten around to showing how to derive the closed form solution to this sum, let's do that now. It's a nice quick warm-up to the following examples.

$$\sum_{x=1}^n x^2 = \sum_1^{n+1} x^2 \delta x = \sum_1^{n+1} x^2 + x^1 \delta x = x^{\frac{3}{2}}/3 + x^{\frac{2}{2}}/2 \Big|_1^{n+1} = (n+1)^{\frac{3}{2}}/3 + (n+1)^{\frac{2}{2}}/2.$$

5.2 A Double Sum

As promised in the introduction, let's compute $\sum_{x=1}^n \sum_{y=1}^n (x+y)^2$ using finite calculus. First, we'll convert to falling powers and take the anti-derivative, then evaluate the inner sum and repeat the process.

$$\begin{aligned}
\sum_{x=1}^n \sum_{y=1}^n (x+y)^2 &= \sum_1^{n+1} \sum_1^{n+1} (x+y)^2 \delta y \delta x = \sum_1^{n+1} \sum_1^{n+1} (x+y)^2 + (x+y)^1 \delta y \delta x \\
&= \sum_1^{n+1} (x+n+1)^3/3 + (x+n+1)^2/2 - (x+1)^3/3 - (x+1)^2/2 \delta x \\
&= (2n+2)^4/12 + (2n+2)^3/6 - (n+2)^4/6 - (n+2)^3/3
\end{aligned}$$

Oops. While that was correct, we skipped one step. What is the $\sum (x+c)^m \delta x$? We assumed it was the same as $\sum x^m \delta x$ without proof. Looking at the proof of Theorem 3.1, it is true since we just add a constant to the falling power.

5.3 Average Codeword Length

(This example, without the story, comes courtesy of CS 141 at Harvey Mudd College [2]). Suppose you are the director of programming for one of the Mars rovers. These rovers accept command codewords of up to 10 bits. A brief summary of the commands are listed below.

0	stop
1	take photo
00	move forward
01	move backward
10	move left
11	move right
⋮	
001101	take picture of green alien
⋮	
1111111111	self-destruct

There are a few interesting aspects to this table. First, 0 and 00 are different commands! Second, all possible commands are used and there is no redundancy between the command codes.

After you have designed and implemented this codeword system, your boss calls you on the phone and wants to know the average codeword length. Since NASA pays interplanetary communications costs per bit, he really wants to know how much money your codeword system will cost.

While still talking to your boss, you quickly jot down a few notes.

First, for a codeword of exactly x bits, you note that there are 2^x possible codes. Hence, the total number of codes is

$$\sum_{x=1}^{10} 2^x = 2^{11} - 2.$$

Second, the total length of all codewords of exactly x bits is $x2^x$. Aha! Now you can easily compute the average codeword length.

$$\frac{\sum_{x=1}^{10} x2^x}{\sum_{x=1}^{10} 2^x}.$$

While you've already solved the denominator, the numerator is a little more challenging. However, since you have (obviously) read all portions of this tutorial you know about finite calculus and, unabashed, proceed ahead.

$$\sum_{x=1}^{10} x2^x = \sum_1^{11} x2^x \delta x.$$

Now, you remember this little theorem about “discrete integration by parts” (Theorem 3.6).

$$\sum x2^x \delta x = x2^x - \sum 2^{x+1} \delta x = x2^x - 2^{x+1} = 2^x(x-2).$$

Then,

$$\sum_{x=1}^{10} x2^x = 2^x(x-2)|_1^{11} = 9 \cdot 2^{11} + 2.$$

The average codeword length is then

$$\frac{9 \cdot 2^{11} + 2}{2^{11} - 2} = 9.00978 \approx 9.$$

After a brief pause in the conversation with your boss, you definitively state: “Sir, the average codeword length is 9 bits to two decimal places.”

6 Conclusion

I hope you've enjoyed learning about finite calculus as much as I've enjoyed writing about it. I think that finite calculus is a very nice set of mathematics that is quite easy to understand and apply. Hopefully, you now understand it and can apply finite calculus to your own work if the need arises. If you are still interested in the subject, please see the excellent book, *Concrete Mathematics* [1], by Ronald Graham, Don Knuth, and Oren Patashnik. In that book, the authors provide a similar derivation of finite calculus and show many other cases where it helps solve yet another nasty sum.

References

- [1] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete mathematics*. Addison-Wesley, Reading, MA, 1989.
- [2] Ran Libeskind-Hadas. Advanced algorithms. Course at Harvey Mudd College, 2004.
- [3] Jeff Miller. History of mathematics. <http://members.aol.com/jeff570/calculus.html>.
- [4] Lloyd N. Trefethen and David Bau. *Numerical Linear Algebra*. SIAM, 1997.
- [5] J.H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge University Press, 2nd edition, 2002.
- [6] Daniel Zwillinger, editor. *Standard Mathematical Tables and Formulae*. CRC Press, Boca Raton, FL, 1996.