These notes will attempt to collect a lot of information about polynomials and Stirling numbers and the calculus of finite differences. Let $\mathcal{P}_n$ denote the set of all polynomials of degree $\leq n$. It is easy to see that this is a vector space over the reals (or the complex numbers if you like) of dimension $n + 1$. The most familiar basis for $\mathcal{P}_n$ is:

$$\{1, x, x^2, \ldots, x^n\}.$$ 

Any $p \in \mathcal{P}_n$ is almost instantaneously expressed in terms of this basis:

$$p(x) = \sum_{k=0}^{n} a_k x^k.$$ 

And calculus is easy in terms of this basis, which implies that

$$p'(x) = \sum_{k=0}^{n} a_k k x^{k-1} = \sum_{k=0}^{n-1} (k+1) a_{k+1} x^k.$$ 

Although antidifferentiation is defined for any polynomial, it’s not defined within $\mathcal{P}_n$, except for polynomials of degree at most $n - 1$, and it is defined only up to an “arbitrary constant”:

$$p(x) = \sum_{k=0}^{n-1} a_k x^k \implies \int p(x) = C + \sum_{k=0}^{n-1} \frac{a_k}{k+1} x^{k+1} = C + \sum_{k=1}^{n} \frac{a_{k-1}}{k} x^k.$$ 

Fans of linear algebra might want to make the matrices associated to these operators!

Another important basis is the “descending product”, defined recursively by

$$\begin{align*}
(x)_0 &= 1; \\
(x)_1 &= x \\
(x)_2 &= x(x-1) = x^2 - x \\
(x)_3 &= x(x-1)(x-2) = x^3 - 3x^2 + 2x \\
(x)_4 &= x(x-1)(x-2)(x-3) = x^4 - 6x^3 + 11x^2 - 6x.
\end{align*}$$

It is plain that $\{(x)_k : 0 \leq k \leq n\}$ also forms a basis for $\mathcal{P}_n$. This is because it is a subset of $\mathcal{P}_n$ which is linearly independent: the degree of $\sum_{k=0}^{n} c_k (x)_k$ will be the largest $k$ for which $c_k \neq 0$. This is also clear because we can write matrices for
one basis with respect to another, and these will be lower triangular with 1’s on the diagonal. We will have considerable interest in considering the entries of this matrix! For reference,

\[
\begin{align*}
1 &= (x)_0 \\
x &= (x)_1 \\
x^2 &= (x)_2 + (x)_1 \\
x^3 &= (x)_3 + 3(x)_2 + (x)_1 \\
x^4 &= (x)_4 + 6(x)_3 + 7(x)_2 + (x)_1
\end{align*}
\]

To represent these in terms of the standard order of the bases, we get the following \(5 \times 5\) matrices, which are inverses of each other:

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 \\
0 & 2 & -3 & 1 & 0 \\
0 & -6 & 11 & -6 & 1 \\
\end{pmatrix} \quad ; \quad
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 \\
0 & 1 & 3 & 1 & 0 \\
0 & 1 & 7 & 6 & 1 \\
\end{pmatrix}
\]

It is natural to try to find general formulas for these coefficients, and to do so, we will separate out the first basis element 1, which is the same for both. Since \(x^k\) and \((x)_k\) are both divisible by \(x\) for \(k \geq 1\), this first basis element will never appear, and we can ignore it. Taking into account the signs in the first equation, let’s write

\[
(x)_m = \sum_{k=1}^{m} (-1)^{m-k} s(m, k) x^k; \quad x^m = \sum_{k=1}^{m} S(m, k)(x)_k
\]

with the understanding that \(s(m, k) = S(m, k) = 0\) if \(k > m\) or if \(k = 0\). Traditionally, the \(s(m, k)’s\) are called the Stirling numbers of the first kind and the \(S(m, k)’s\) are called the Stirling numbers of the second kind. Clearly, \(s(1, 1) = S(1, 1) = 1\). We can use the definitions to find the recurrences satisfied by the Stirling numbers when \(m \geq 2\). Indeed, we have

\[
\begin{align*}
(x)_m &= \sum_{k=1}^{m} (-1)^{m-k} s(m, k) x^k = (x - (m - 1)) \sum_{k=1}^{m-1} (-1)^{m-1-k} s(m - 1, k) x^k \\
&= \sum_{k=1}^{m-1} (-1)^{m-1-k} s(m - 1, k) x^{k+1} - (m - 1) \sum_{k=1}^{m-1} (-1)^{m-1-k} s(m - 1, k) x^k \\
&= \sum_{j=2}^{m} (-1)^{m-j} s(m - 1, j - 1) x^j + \sum_{k=1}^{m-1} (-1)^{m-k} (m - 1) s(m - 1, k) x^k
\end{align*}
\]

(We’ve fiddled with the \(-1’s, but it all works out.) Extracting the coefficient of \(x^k\) from both sides above, and keeping in mind that \(s(m - 1, 0) = s(m - 1, m) = 0\), we
get the recurrence:

\[ s(m, k) = s(m - 1, k - 1) + (m - 1)s(m - 1, k). \]

The other recurrence is a bit subtler. Observe that

\[ x(x)_k = (x - k)(x)_k + k(x)_k = (x)_{k+1} + k(x)_k. \]

Thus, assuming this holds for \( m \), we have

\[ x^m = \sum_{k=1}^{m} \frac{s(m, k)}{x} (x)^{m-1} \sum_{k=1}^{m-1} x \cdot s(m, k) x(x)_k \]

\[ = \sum_{k=1}^{m-1} s(m - 1, k)(x)_{k+1} + k \sum_{k=1}^{m-1} s(m - 1, k)(x)_k \]

\[ = \sum_{j=2}^{m} s(m - 1, j - 1)(x)_j + \sum_{k=1}^{m-1} kS(m - 1, k)(x)_k \]

Again, extracting the coefficient of \((x)_k\) from both sides, and recalling that \( S(m - 1, 0) = S(m - 1, m) = 0 \), we get

\[ S(m, k) = S(m - 1, k - 1) + kS(m - 1, k). \]

Both sets of Stirling numbers have combinatorial interpretations: Consider the set \([1, m] = \{1, 2, \ldots, m\}\) for integers \( m \geq 1 \). How many ways are there to write \([1, m]\) as a union of exactly \( k \) distinct subsets? Call this number \( a(m, k) \). Clearly, \( a(1, 1) = 1 \) and \( a(m, k) = 0 \) unless \( 1 \leq k \leq m \). We can derive a recurrence for \( a(m, k) \) in the following way: consider the ways to write \([1, m - 1]\) as a union of subsets. Any decomposition of \([1, m - 1]\) into \( k \) subsets can be thought of either as a decomposition of \([1, m - 1]\) into \( k - 1 \) subsets with \( \{m\} \) as a new subset (there are \( a(m - 1, k - 1) \) of these) or as a decomposition of \([1, m - 1]\) into \( k \) subsets, with “\( m \)” added to one of the pre-existing subsets. Since this can be done in \( k \) different ways, for each of the \( a(m - 1, k) \) subsets, we have that

\[ a(m, k) = a(m - 1, k - 1) + ka(m - 1, k). \]

Comparing this with the earlier formula, we see that \( a(m, k) = S(m, k) \), and this is the combinatorial interpretation of the Stirling numbers of the second kind. We can add to this by interpreting the equation

\[ x^m = \sum_{k=1}^{m} a(m, k)(x)_k \]

for integer \( x \). Consider all strings of \( m \) characters, each of which can be one of \( x \) different letters. There are \( x \) choices for each position, and so \( x^m \) strings. Now suppose we want to count the number of different letters in a string. Suppose there
are \( k \) different letters in a string. Then the \( m \) positions are divided into \( k \) different subsets according to their value. For example

\[
2770230 \iff \{1, 5\} \cup \{2, 3\} \cup \{4, 7\} \cup \{6\}
\]
because the 1st and 5th position have the same letter, as do the 2nd and 3rd position with another letter and the 4th and 7th with a third letter, and the 6th with a fourth letter. For each such choice of \( k \) subsets, there are \( x \) choices of assignment for the first subset, and then \( x - 1 \) for the second subset, and so on, leaving \( (x)_k \) overall. Since there are \( a(m, k) \) ways of finding these \( k \) subsets, we see that there are \( a(m, k)(x)_k \) strings with exactly \( k \) different characters, so adding them up, we get \( x^m \) as desired.

The combinatorial interpretation of \( s(m, k) \) is only slightly more complicated. We consider permutations of \([1, m]\); that is, bijective functions \( \pi : [1, m] \to [1, m] \). As you may know, there are \( m! \) permutations and, furthermore, each permutation can be decomposed into a product of disjoint cycles. For example, if \( \pi(1) = 3, \pi(2) = 5, \pi(3) = 4, \pi(4) = 1 \) and \( \pi(5) = 2 \) is a permutation of \([1, 5]\), then we can write \( \pi = (134)(25) \) (or \( \pi = (3, 4, 1)(52) \), for that matter). Let \( b(m, k) \) be the number of permutations of \([1, m]\) for which there are exactly \( k \) cycles. Again, clearly, \( b(m, k) = 0 \) unless \( 1 \leq k \leq m \) and \( b(1, 1) = 1 \). To find a recurrence, we again look at the element “\( m \)” separately. If \( \pi \) is a permutation of \([1, m]\) and \( \pi(m) = m \) and \( \pi \) has \( k \) cycles, then we can make \( \pi \) by taking a permutation of \([1, m-1]\) with \( k-1 \) cycles and letting \( (m) \) be a \( k \)-th cycle. This can be done in \( b(m-1, k-1) \) ways. If “\( m \)” is part of a cycle, then we can simply delete it, and look at the permutations of \([1, m-1]\) with \( k \) cycles. There are \( b(m-1, k) \) of these, and \( m \) can follow any of the \( m-1 \) elements of \([1, m-1]\), hence this can be done in \((m-1)b(m-1, k)\) ways. Thus,

\[
b(m, k) = b(m-1, k-1) + (m-1)b(m-1, k),
\]
and, comparing with above, we see that \( b(m, k) = s(n, k) \), the Stirling numbers of the first kind. The combinatorial interpretation of

\[
(x)_m = \sum_{k=1}^{m} (-1)^{m-k} b(m, k) x^k
\]
is most easily seen by setting \( x = -1 \) above. Note that

\[
(-1)_m = (-1)(-2) \cdots (-m) = (-1)^m m!
\]
so we have

\[
(-1)^m m! = \sum_{k=1}^{m} (-1)^{m-k} b(m, k) (-1)^m \implies m! = \sum_{k=1}^{m} b(m, k).
\]
This is of course unsurprising, because every permutation of \([1, m]\) has to have a certain number of cycles.

Recall that \( \mathbb{P}_n \) is the vector space of all polynomials of degree \( \leq n \). It is easy to see that this is a vector space over the reals (or the complex numbers if you like) of
dimension $n + 1$. We now give names to the elements in the most familiar basis for $\mathcal{P}_n$: 
\[ \{ f_0(x), f_1(x), f_2(x), \ldots, f_n(x) \} := \{ 1, x, x^2, \ldots, x^n \}. \]
As noted before, $p \in \mathcal{P}_n$ is almost instantaneously expressed in terms of this basis:
\[ p(x) = \sum_{k=0}^{n} a_k f_k(x) = \sum_{k=0}^{n} a_k x^k. \]
But how can we find $a_k$? A theoretically sophisticated approach to this almost trivial question is to consider differentiation abstractly. Let $D$ denote differentiation; we want to iterate $D$, so we write:
\[ Dp = p'; \quad D^0p = p, \quad D^j p := D(D^{j-1}p), \text{ for } j \geq 1. \]
This definition allows us to consider the 0-th derivative, as well as arbitrary higher derivatives. You already know that $Df_0 = D(1) = 0$, and, if $k \geq 1$,
\[ Df_k = k f_{k-1}. \]
So $D^2 f_k = k(k-1) f_{k-2}$ if $k \geq 2$, etc. Recall that 
\[ (n)_m = n(n-1) \cdots (n-(m-1)) = \begin{cases} \frac{n!}{(n-m)!} & \text{if } n \geq m; \\ 0 & \text{if } n < m. \end{cases} \]
It is not hard to prove by induction that, for any integer $j$,
\[ D^j f_k = \begin{cases} (k)_j f_{k-j}, & \text{if } k \geq j; \\ 0, & \text{if } k < j \end{cases} \]
Since differentiation is linear, (1) implies that (with $k = j + \ell$),
\[ D^j p = \sum_{k=0}^{n} a_k D^j f_k = \sum_{k=j}^{n} a_k \cdot (k)_j f_{k-j} = \sum_{\ell=0}^{n-j} a_{j+\ell} \cdot \frac{(j+\ell)!}{\ell!} f_{\ell}. \]
The next trivial thing to observe is that
\[ f_k(0) = \Delta_{0,k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0. \end{cases} \]
If we stick this information into (2), evaluate at 0, and note that only the first term of the last sum appears, we get
\[ D^j p(0) = j! a_j \]
\[ \implies a_j = \frac{p^{(j)}(0)}{j!} \implies p(x) = \sum_{k=0}^{n} \frac{p^{(k)}(0)}{k!} x^k. \]
In other words, we have re-proved Taylor’s Theorem for polynomials! (A minor variation is to fix a constant $c$ and take as the basis $f_{k,c}(x) = (x - c)^k$; evaluation at $c$ then gives $p$ as a polynomial in powers of $x - c$.)
We now make a significant variation. Recall the basis 
\[ \{g_0(x), g_1(x), g_2(x), \ldots, g_n(x)\} := \{1, (x)_1, (x)_2, \ldots, (x)_n\}, \]
where \( (x)_k = x(x-1) \cdots (x-(k-1)) = x(x-1)_{k-1} = (x-(k-1))_{k-1} \), and the fact that we can also write every element \( p \) in \( \mathcal{P}_n \) as 
\[ p(x) = \sum_{k=0}^{n} b_k g_k(x) = \sum_{k=0}^{n} b_k(x)_k. \]

The role of differentiation here is played by the difference operator \( \Delta \), defined by 
\[ \Delta(p)(x) = p(x+1) - p(x), \quad \Delta^0 p = p, \quad \Delta^j p := \Delta(\Delta^{j-1} p), \text{ for } j \geq 1. \]
Note that \( \Delta^2 p(x) = p(x+2) - 2p(x+1) + p(x) \), \( \Delta^3 p(x) = p(x+3) - 3p(x+2) + 3p(x+1) - p(x) \), and it is easy to show by induction that 
\[ (\Delta^j p)(x) = \sum_{\ell=0}^{j} (-1)^\ell \binom{j}{\ell} p(x+j-\ell). \]
Observe that \( \Delta g_0 = 0 \) and, for \( k \geq 1, \)
\[ (\Delta g_k)(x) = g_k(x+1) - g_k(x) = (x+1)(x)_k-1 - (x-(k-1))(x)_k-1 = k g_{k-1}(x). \]
(For the initiates: the action of \( \Delta \) on the \( g \)-basis is the same as the action of \( D \) on the \( f \)-basis – they have the same matrix.) In any case, we get the analogous formula:
\[ \Delta^j g_k = \begin{cases} (k)_j g_{k-j}, & \text{if } k \geq j; \\ 0, & \text{if } k < j. \end{cases} \]
Thus, since \( \Delta \) is linear, (3) implies that 
\[ \Delta^j p = \sum_{k=0}^{n} b_k \cdot \Delta^j g_k = \sum_{k=j}^{n} b_k \cdot (k)_j g_{k-j} = \sum_{\ell=0}^{n-j} b_{j+\ell} \cdot \frac{(j+\ell)!}{\ell!} g_\ell. \]
And, as before, we note that 
\[ g_k(0) = \delta_{0,k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0. \end{cases} \]
and so 
\[ \Delta^j p(0) = j! b_j \]
\[ \implies b_j = \frac{\Delta^j(p)(0)}{j!} \implies p(x) = \sum_{k=0}^{n} \frac{(\Delta^k p)(0)}{k!} (x)_k. \]
This fact is less well-known than Taylor’s Theorem; it was first proved by Isaac Newton.
A few other things to mention. In addition to the closed form for $\Delta^j p(x)$, there is a simple “practical” way to compute the successive differences, using a table of differences. For example, in
\[
\begin{array}{ccc}
p(x) & p(x + 1) & p(x + 2) \\
p(x + 1) - p(x) & p(x + 2) - p(x + 1) & p(x + 2) - 2p(x + 1) + p(x)
\end{array}
\]
the successive rows are $p(x)$, $(\Delta p)(x)$, $(\Delta^2 p)(x)$, etc.

If one fixes $t$ and defines
\[
(\Delta_t p)(x) = \frac{p(x + t) - p(x)}{t}, \quad h_{k,t}(x) = x(x - t) \cdots (x - (k - 1)t),
\]
then it is worth checking that $\Delta_t h_{k,t} = k \cdot h_{k-1,t}$, so the same construction goes through as before. As $t \to 0$, $h_{k,t}(x) \to x^k$, and we obtain a formula really familiar only for $k = 1$:
\[
p^{(k)}(0) = \lim_{t \to 0} \frac{\sum_{j=0}^{k} (-1)^j \binom{k}{j} p((k - j)t)}{t^k}.
\]
For example,
\[
p'(0) = \lim_{t \to 0} \frac{p(t) - p(0)}{t}, \quad p''(0) = \lim_{t \to 0} \frac{p(2t) - 2p(t) + p(0)}{t^2}, \quad \text{etc.}
\]
We have previously established the coefficients relating the two bases in terms of the Stirling numbers of both kinds:
\[
f_0 = g_0, \quad f_k = \sum_{j=1}^{k} S(k, j) g_j, \quad g_k = \sum_{j=1}^{k} (-1)^{k-j} s(k, j) g_j.
\]
It is worth working out the details to see the connections between the successive differences of a polynomial at 0 and its successive derivatives. For example, for $p \in P_n$,
\[
p(x) = \sum_{k=0}^{n} \frac{(\Delta^k p)(0)}{k!} (x)_k = \sum_{k=0}^{n} \frac{p^{(k)}(0)}{k!} x^k = \sum_{k=0}^{n} \frac{p^{(k)}(0)}{k!} \sum_{j=0}^{k} S(k, j)(x)_j.
\]
If we pull out the coefficient of say $(x)_1$ from the above (and we can, since the $g_k$’s are a basis!), we see that
\[
p(1) - p(0) = \Delta p(0) = \sum_{k=0}^{n} \frac{p^{(k)}(0)}{k!} S(k, 1) = \sum_{k=0}^{n} \frac{p^{(k)}(0)}{k!}.
\]
this is of course not surprising, since it’s a direct consequence of Taylor’s formula. (The formula $S(k, 1) = 1$ used above is an immediate consequence of its combinatorial definition in the last notes.) It does suggest a symbolic interpretation, taken for
functions in general, and not just polynomials. If we assume that the Taylor series for \( f \) converge everywhere, then symbolically
\[
f(x_0 + 1) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} 1^k \implies f(x_0 + 1) - f(x_0) = \sum_{k=1}^{\infty} \frac{D^k}{k!} f(x_0)
\]
\[
\implies (\Delta f)(x_0) = \sum_{k=1}^{\infty} \frac{D^k}{k!} f(x_0) \implies \Delta = e^D - 1.
\]

One interesting consequence of Newton’s formula is that it gives an explicit formula for the Stirling numbers of the second kind. We have:
\[
x^n = \sum_{k=0}^{n} S(n, k)(x)_k = \sum_{k=0}^{n} \frac{(\Delta^k f_n)(0)}{k!} (x)_k
\]
\[
\implies S(n, k) = \frac{(\Delta^k f_n)(0)}{k!} \implies k!S(n, k) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n
\]
\[
\implies k!S(n, k) = k^n - \binom{k}{1} (k-1)^n + \binom{k}{2} (k-2)^n \cdots .
\]

This identity for \( k!S(n, k) \) can be given a combinatorial interpretation: consider an alphabet of \( k \) letters and words of length \( n \). The number of such words which use each of the letters at least once can be counted using the Principle of Inclusion and Exclusion, giving the value on the right hand side. This number can also be found by counting the number of ways to divide the \( n \) positions of the word into exactly \( k \) subsets (\( S(n, k) \)) and then making \( k! \) different words using each such pattern.

One final point here is that we can give an interesting generating function for the Stirling numbers, and then for \( B_m := \sum_k S(m, k) \), the Bell numbers. It is easiest to start with the answer and work backwards. If you want to see how this was discovered, start at the bottom and work your way up! As in all derivations, we do not worry about convergence, reversing the orders of summation (and such) until the end, if at all.

\[
e^{x(e^y-1)} = \sum_{k=0}^{\infty} \frac{x^k (e^y-1)^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k}{k!} \sum_{j=0}^{k} \binom{k}{j} (-1)^j (e^y)^{k-j}
\]
\[
= \sum_{k=0}^{\infty} \sum_{j=0}^{k} \binom{k}{j} (-1)^j \sum_{m=0}^{\infty} \frac{(k-j)^m}{m!} y^m
\]
\[
= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \sum_{j=0}^{k} \binom{k}{j} (-1)^j (k-j)^m \frac{x^k y^m}{k! m!}
\]
\[
= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} k! S(m, k) \frac{x^k y^m}{k! m!} = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} S(m, k) \frac{x^k y^m}{m!} .
\]
In particular, if we set \( x = 1 \) above, we see that
\[
e^{e^y-1} = \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} S(m, k) \frac{y^m}{m!} = \sum_{m=0}^{\infty} B_m \frac{y^m}{m!},
\]
If you are picky, you may notice that we assumed
\[
k!S(m, k) = \sum_{j=0}^{k} \binom{k}{j} (-1)^j (k - j)^m
\]
for all \( k, m \geq 0 \), and we’ve actually only proved it for \( k, m \geq 1 \). In fact, the definition of \( S(m, k) \) in terms of the change of basis shows that \( S(0, 0) = 1 \), \( S(k, 0) = S(0, m) = 0 \) if \( k, m > 0 \), and a routine check of the sum shows the same thing, if we interpret \( 0^0 = 1 \).

Let’s turn now to finding an inverse operation to \( \Delta \). What we want is an operator, call it \( S \) so that for all real \( x \).
\[
\Delta(Sp) = p \implies (Sp)(x + 1) - (SP)(x) = p(x).
\]
Just as constants are killed by differentiation, they are killed by differencing, so the best we can hope for is to define \( Sp \) uniquely up to an additive constant. If we sum the second equation in (5) for \( x = 0, 1, \ldots, n - 1 \), then the left hand side is a telescoping sum, and, if we set \( (SP)(0) = 0 \) rather arbitrarily, we get
\[
(Sp)(n) = p(0) + p(1) + \cdots + p(n - 1) = \sum_{k=0}^{n-1} p(k)
\]
Of course, this is only true for integers \( n \), and (6) makes no sense if \( n \) is not an integer, but if we can show that, if \( p \) is a polynomial, then there is a polynomial \( Sp \) satisfying (6), then we can extend the definition of \( Sp \) to a single real variable. We’ll be in fine shape, because then, indeed,
\[
(\Delta(Sp))(n) = \sum_{k=0}^{n} p(k) - \sum_{k=0}^{n-1} p(k) = p(n).
\]
Notice also that, as a consequence of this definition
\[
(S(\Delta(p)))(n) = \sum_{k=0}^{n-1} (\Delta p)(k) = \sum_{k=0}^{n-1} p(k + 1) - p(k) = p(n) - p(0),
\]
an analogue to \( \int_0^x (Dp) t \, dt = p(x) - p(0) \), the Fundamental Theorem of Calculus!.

We are in fine shape! It is easy to see directly from the definition that
\[
(Sg_0)(n) = \sum_{k=0}^{n-1} g_0(k) = \sum_{k=0}^{n-1} 1 = n = (n)_1
\]
and, since $\Delta g_k = k \cdot g_{k-1}$ for $k \geq 1$,

$$S(g_k) = S\left(\frac{1}{k+1}(g_{k+1})\right) = \frac{1}{k+1}(g_{k+1}(x) - g_{k+1}(0)) = \frac{g_{k+1}}{k+1}.$$

The analogy with differentiation and integration should be clear. Using this, we now make the definition of $S$ for a polynomial $p$. (Important mathematical point: the above definition makes no sense as a definition for the evaluation of $Sp$ at non-integral values of $x$!)

$$p = \sum_{k=0}^{n} b_k \cdot g_k \implies S(p) = \sum_{k=0}^{n} \frac{b_k}{k+1} g_{k+1}.$$

Finally, we can get a closed form for $\sum_{k=1}^{n} k^r$. In fact, we have

$$x^r = \sum_{j=1}^{r} S(r, j)(j)_{k} \implies \sum_{k=1}^{n-1} k^r = \sum_{k=1}^{r} S(r, k) \cdot S((x)_{k})|_{x=n} = \sum_{k=1}^{r} \frac{S(r, k)}{k+1} (n)_{k+1}$$

$$\implies \sum_{k=1}^{n} k^r = \sum_{k=1}^{r} \frac{S(r, k)}{k+1} (n+1)_{k+1}.$$

For example, $x^4 = (x)_4 + 6(x)_3 + 7(x)_2 + (x)_1$, hence

$$\sum_{k=1}^{n} k^4 = \frac{1}{5}(n+1)_5 + \frac{6}{4}(n+1)_4 + \frac{7}{3}(n+1)_3 + \frac{1}{2}(n+1)_2$$

$$= \left(\frac{n(n+1)}{30}\right) (6(n-1)(n-2)(n-3) + 45(n-1)(n-2) + 70(n-1) + 15)$$

$$= \frac{(n+1)n(6n^3 + 9n^2 + n - 1)}{30} = \frac{(n+1)n(2n+1)(3n^2 + 3n - 1)}{30}.$$