Lecture 2 – Sums

Sources: Chap. 2 and 9, Concrete Maths.

We often want to estimate sums of the form: $S_n := \sum_{k=0}^n a_k$; for instance, estimating $n!$ is the same as estimating the sum $\sum_{i=1}^n \log k$. Often it happens that we cannot get a closed form for the summation, or that the closed form does not give us insight into the magnitude of the quantity. In such cases, and even as a first policy, one should try to get an asymptotic estimate to the summation, and further on try to make it as tight (in the sense of Θ) as one can. E.g., we do not know a closed form for the sum of the first n primes, however, we do know good asymptotic estimates to the sum. In this lecture, we will first see few techniques that help us to get closed forms for S_n, and in the instances where we fail to do so, we will see some basic techniques that help us get good estimates.

Let's look at the familiar sum of the numbers from 1 to n, namely $A_1 := \sum_{k=1}^n k$. We all know, or remember, that $A_1 = n(n+1)/2$. The usual proofs for this statement are based upon induction. But inductive proofs need a good guess in the first place. So, how do we derive this formula from first principles? We can use a trick by Gauß [1]. By reversing the order of summation, we know that $A_1 = n + (n-1) + \cdots + 1$. Thus $2A_1 = n(n+1)$, which gives us the desired formula. In general, if we have want to sum the terms of an arithmetic progression $S_n = \sum_{k=1}^n (a + bk)$, then applying Gauß's trick, i.e. reversing the order of summation and adding the two sums, we get $S_n = n(2a + bn)/2$. Another way to interpret this sum is that the kth and $(n-k)$th term add to twice the mean $(2a + bn)/2$; since there are $n/2$ such pairs we get the desired formula for S_n. Thus Gauß's trick works well for such "linear" sums. What if we are interested in the sum $A_2 := \sum_{k=1}^n k^2$? Applying the the trick we get

$$2A_2 = \sum_{k=1}^n (k^2 + (n+1-k)^2) = 2A_2 + n(n+1)^2 - 2(n+1)A_1.$$

Though we don't get an expression for A_2, we do get an expression for A_1. So perhaps we should apply the trick to A_3:

$$2A_3 = \sum_{k=1}^n (k^3 + (n+1-k)^3) = n(n+1)^3 - 3(n+1)^2 A_1 + 3(n+1)A_2.$$

Since A_3 does not cancel out, we are stuck.

Let's try the **perturbation method**. Suppose we want to find closed form for S_n, then we express S_{n+1} in two different ways. Thus

$$S_{n+1} = S_n + a_{n+1} = a_0 + \sum_{k=1}^{n+1} a_k = a_0 + \sum_{k=0}^n a_{k+1}.$$

Now if we can express $\sum_{k=0}^n a_{k+1}$ in terms of S_n then we can hope to have an equation in terms of S_n and solve for it. Let's try this method for summing the geometric series $S_n := \sum_{k=0}^n x^k$. Now from the equation above we obtain

$$S_n + x^{n+1} = 1 + \sum_{k=0}^n x^{k+1} = 1 + x\sum_{k=0}^n x^k = 1 + xS_n,$$

which gives us the standard closed form formula

$$S_n = \frac{x^{n+1} - 1}{x - 1}.$$

[1] mention the story...

Let's see if the perturbation method works for A_2. Applying the method, we get

$$A_2 + (n+1)^2 = \sum_{k=0}^{n} (k+1)^2 = \sum_k (k^2 + 2k + 1) = A_2 + 2\sum_k k + n + 1.$$

But this does not lead us to a solution as A_2 cancels on both sides. However, all is not lost, since after canceling A_2, we get a closed form for A_1. This means that to get to A_2, we should apply the method to A_3:

$$A_3 + (n+1)^3 = \sum_{k=0}^{n} (k^3 + 3k^2 + 3k + 1) = A_3 + 3A_2 + 3A_1 + n + 1;$$

canceling A_3 on both sides, we get the desired formula for A_2.

A more general approach that uses the elements of calculus but in a discrete setting is based upon *finite calculus*, or calculus of finite differences.

1 Finite Calculus

As we had seen above, the continuous analogue of A_2 is the integral $\int_0^n x^2 dx$, which we know from the Fundamental Theorem of Calculus has a nice closed form $n^3/3$. Can we obtain something analogous for summation? Since the concept of derivative is necessary to get to the integral, let's explore that first.

In the continuos domain the derivative operator D for a function f is

$$Df(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

In finite calculus, we have the corresponding **difference** operator Δ:

$$\Delta f(x) := f(x+1) - f(x), \tag{1}$$

which is the same as D but the limit tends to one, as that is the closest natural numbers can get to each other. We will call Δf as the **discrete derivative** of f. From the definition, we have two following algebraic properties of the discrete derivative:

$$\begin{aligned} \text{Distributivity: } & \Delta(f+g) = \Delta f + \Delta g \\ \text{Scalar Multiplication: } & \Delta c f = c \Delta f. \end{aligned} \tag{2}$$

We know that $D(x^m) = mx^{m-1}$. Can we derive something analogous for Δ? Does the same formula work? Let's start with $m = 1$: $\Delta x = (x+1) - x = 1$. So that's correct; for $m = 2$, $\Delta x^2 = 2x + 1$ doesn't work; but that is not hopeless since it implies $\Delta x^2 - \Delta x = 2x$, or equivalently $\Delta(x^2 - x) = 2x$ (where we use the distributivity of Δ). Guessing further, let's define $x^{\underline{3}} := x(x-1)(x-2)$, then it's easy to see that $\Delta x^{\underline{3}} = 3x^{\underline{2}}$. And so with an educated guess we define the **falling factorial** for a positive integer m

$$x^{\underline{m}} := x(x-1)\ldots(x-m+1). \tag{3}$$

When $m = 0$, let $x^{\underline{0}} = 1$. It is not hard to verify that

$$\Delta x^{\underline{m}} = m x^{\underline{m-1}}. \tag{4}$$

The above development works for positive powers, i.e., $k^{\underline{m}}$, where m is a positive integer. What about negative powers? How should we define $x^{\underline{-1}}$? In going from $x^{\underline{3}}$ to $x^{\underline{2}}$ we divide by $x - 2$, from $x^{\underline{2}}$ to $x^{\underline{1}}$ by $x - 1$, from $x^{\underline{1}}$ to $x^{\underline{0}}$ by x, so it seems natural to divide by $x + 1$ to go from $x^{\underline{0}}$ to $x^{\underline{-1}}$; continuing in this manner we define for $m > 0$

$$x^{\underline{-m}} := \frac{1}{(x+1)(x+2)\ldots(x+m)}.$$

And what is $\Delta x^{\underline{-m}}$? Well it is what we expect it to be, $-mx^{\underline{-m-1}}$.

But what happens when $m = -1$? In the continuous case we know that $D \ln x = 1/x$. So what is the discrete analogue of the ln function? What function f has the property that $\Delta f(x) = x^{\underline{-1}} = 1/(x+1)$? We know from the definition of summation that the function $H_x := \sum_{k<x+1} 1/k$ will work. This is our familiar Harmonic function, and it makes sense that H_x is the discrete analogue of $\ln x$ since their values differs by at most one (more precisely, the Euler-Mascheroni constant 0.57721...) as x increases.

We can further ask, what function corresponds to the exponential? That is $\Delta f(x) = f(x)$. The equality gives us the recurrence $f(x+1) = 2f(x)$, so we can take $f(x) = 2^x$. In general, $\Delta c^x = c^x/(c-1)$.

Now that we have the analogue of D, we want the analogue for $\int$. Naturally, it is the summation operator $\sum$. Define $\sum g(x)\delta x$ as the class of functions f s.t. $\Delta f = g(x)$; similar to the big-O notation we will write $\sum g(x)\delta x = f(x)$ rather than the set-membership operation. The δx in this definition can be replaced with one; however, we will soon see that the "indefinite summation" has a slightly different upper limit than the usual summation; moreover, we also keep it to continue our analogy with standard calculus where a similar "dx" term appears in the integral. What functions equal to$\sum g(x)\delta x$? One such function is $\sum_{k<x} g(k)$, since the difference operator yields $\sum_{k<x+1} g(k) - \sum_{k<x} g(k) = g(x)$. In fact, let's define

$$\sum g(x)\delta x := \sum_{k<x} g(k);$$

note that the summation is strictly smaller than x, which was one of the justifications for introducing δx in our summation notation above to distinguish it from the summation $\sum^x$. We then have the following theorem:

Theorem 1 (Fundamental Theorem of Finite Calculus) *Given two functions g, f*

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

where C is a function such that $C(x+1) = C(x)$.

Proof. Since

$$\sum g(x)\delta x = \sum_{k<x} g(k) = \sum_{k<x} (f(k+1) - f(k)) = f(x) + C(x).$$

Q.E.D.

As an immediate consequence we have the definite integral: for $b \geq a$,

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

Once we have this, it immediately follows that

$$\sum_a^b g(x)\delta x + \sum_b^c g(x)\delta x = \sum_a^c g(x)\delta x$$

and

$$\sum_a^b g(x)\delta x = -\sum_b^a g(x)\delta x;$$

the order of a and b is irrelevant here.

Well, now that we have developed this machinery, where do we apply it? Since we started with wanting an analogy for $\int x^2 dx$, given the above what can we say about the sum $\sum_{k=0}^{<n} k^{\underline{m}}$? From definition we know that this sum is equal to $\sum_0^n x^{\underline{m}}\delta x$, and since $x^{\underline{m}} = \Delta x^{\underline{m+1}}/(m+1)$, we get

$$\sum_{k=0}^{<n} k^{\underline{m}} = \left.\frac{x^{\underline{m+1}}}{(m+1)}\right|_0^n = \frac{n^{\underline{m+1}}}{m+1}.$$

So for $m = 1$ we have $\sum_{k=0}^{<n} k = n^{\underline{2}}/2 = n(n-1)/2$. But what about our sum A_2? Well, if we can express ordinary powers in terms of falling powers then our approach will work. Since $k^2 = k^{\underline{2}} + k^{\underline{1}}$, we have

$$\sum_{k=0}^{<n} k^2 = \frac{n^{\underline{3}}}{3} + \frac{n^{\underline{2}}}{2} = \frac{1}{3}n(n-\frac{1}{2})(n-1);$$

replacing n by $(n+1)$ gives us the result for A_2. We can similarly work with k^3. In general we can express ordinary powers as linear combinations of falling powers, but we won't see this relation until later.

Finite Calculus, even though it has astounding similarities to infinite calculus, has its limitations. For instance the chain rule of differentiation of compositions of functions $(Df(g) = f'(g)g')$ does not carry over to the finite setting. The product rule, however, does:

$$\Delta uv = u\Delta v + Sv\Delta u,$$

where S is the shift operator, $Sv = v(x+1)$. Summing both sides and rearranging terms we get

$$\sum u\Delta v = uv - \sum Sv\Delta u,$$

which is useful when computing the RHS is easier than the LHS. Let's apply it to the sum $\sum_{k=0}^{<n} kH_k$. What should u and Δv be? Taking the latter to be H_k seems to complicate matters, so let's choose $u := H_x$ and $\Delta v := x = x^{\underline{1}}$; thus $v = x^{\underline{2}}/2$. So we have

$$\begin{aligned}\sum H_x x &= H_x \frac{x^{\underline{2}}}{2} - \sum \frac{(x+1)^{\underline{2}}}{2} x^{\underline{-1}} \\ &= H_x \frac{x^{\underline{2}}}{2} - \frac{1}{2}\sum x^{\underline{1}} \\ &= H_x \frac{x^{\underline{2}}}{2} - \frac{x^{\underline{2}}}{4}.\end{aligned}$$

Attaching limits we get that $\sum_{k=0}^{<n} kH_k = \frac{n^{\underline{2}}}{2}(H_n - \frac{1}{2})$.

Just as calculus can be carried out formally, once we have the appropriate setup, the same ease carries over to computing sums (though not as powerfully as calculus). Nevertheless, we must remember that for any function g if we can find an f such that $\Delta f = g$ then we can sum g conveniently.

We have seen that $\sum g(x)$ is closely related to $\int g(x)dx$. We next see a very general approach that gives a precise meaning to this relation.

2 Euler-Maclaurin Formula

Again consider the sum $A_2 = \sum_{i=0}^{n} i^2$. The continuous version of A_2 is the integral $\int_0^n x^2 dx$. What is the relation between the integral and A_2? Clearly, the area underneath the parabola in the range $0, \ldots, n$ falls short of covering the area corresponding to A_2. So if we can form a closed expression for the error term $A_2 - \int_0^n x^2 dx$ then we have an expression for A_2. Let's try

$$\begin{aligned}\sum_{k=0}^{n} k^2 - \int_0^n x^2 dx &= \sum_{k=1}^{n} k^2 - \sum_{k=1}^{n}\int_{k-1}^{k} x^2 dx \\ &= \sum_{k=1}^{n}\left(k^2 - \frac{k^3-(k-1)^3}{3}\right) \\ &= \sum_{k=1}^{n}(k - \frac{1}{3}).\end{aligned}$$

Thus

$$A_2 = \frac{n^3}{3} + \frac{n(n+1)}{2} - \frac{n}{3}$$

as desired. The benefit of working with the integral, which is the "continuous sum", as compared to the discrete sum is that we know how to manipulate integrals; reducing the discrete sum to the integral helps us bear down the powerful machinery of calculus on our sums. In the example above, we were able to give a closed form to the error term. This may not be possible always. But if we can bound the error term, or derive a form for it which can be bounded, then we are in good shape. The Euler-Maclaurin formula developed independently by Euler and Maclaurin helps us achieve this goal.

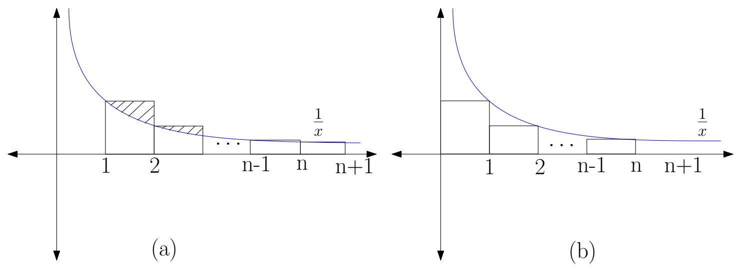

Figure 1: The sum and the integral

How is the sum $H_{n-1} = \sum_{i=1}^{n-1} 1/n$ and the integral $\int_1^n 1/x dx$ related? Let $f(x) := 1/x$; it will be convenient to use this notation, since later on we want to talk of a general function; then we want the relation between $H_{n-1} = \sum_{k=1}^{n-1} f(k)$ and $\int_1^n f(x)dx$. Pictorially, the relation between the two is as shown in Figure 1(a): H_n is the sum under the rectangles with base $[k, k+1]$ and height $f(k)$, whereas the integral is the area under the graph of the function $f(x)$. Clearly, the area covered by the rectangles is greater than that covered by the integral. However, if we shift the rectangles one unit to the left, then the area covered by the rectangles is smaller than the integral $\int_0^n f(x)dx$. Thus we have shown that

$$\int_1^n f(x)dx < \sum_{k=1}^{n-1} f(k) \text{ and } \sum_{k=1}^{n} f(k) \leq \int_0^n f(x)dx.$$

Can we bound the error term

$$E_n := \sum_{k=1}^{n-1} f(k) - \int_1^n f(x)dx? \tag{5}$$

Geometrically, E_n is the sum of the portions of the rectangale that is above the graph of $1/x$ over each interval $[k, k+1]$ (shaded in Figure **??**); the portion of the graph above the interval $[k, k+1]$ is bounded by $f(k) - f(k+1)$. Thus $\lim_{n\to\infty} E_n < 1$. Since E_n is increasing and it is bounded from above by one, it follows that it must converge to a constant $C(f) < 1$. Moreover,

$$0 < C(f) - E_n < f(n).$$

The lower bound follows from definition of $C(f)$ and the upper bound from the telescoping sum $\sum_{k\geq n}(f(k) - f(k+1))$. Let $\varepsilon_f(n) := f(n) - (C(f) - E_n)$ then it follows from (5) that

$$\sum_{k=1}^{n-1} f(k) = \int_1^n f(x)dx + C(f) + \varepsilon_f(n) - f(n),$$

or taking $f(n)$ to the left we obtain that

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx + C(f) + \varepsilon_f(n). \tag{6}$$

The constant $C(f)$ for $f = 1/x$ is called **Euler's constant**.[2] Honestly speaking, however, in the equation above we have not obtained anything new – we have introduced some new terms and are merely saying that

[2]The constant e is called *Euler's number* and is named in honour of Euler; it is sometimes also called Napier's constant.

with the help of these terms the summation can be expressed in terms of the integral. The key question is, Can we say anything about these terms? In particular, can we derive an explicit form for $\varepsilon_f(n)$ in terms of f? The goal is to express the difference between the discrete sum $\sum_1^n f$ and the integral $\int_1^n f$ in the following form:

$$c_0(f(1) \pm f(n)) + c_1(f'(1) \pm f'(n)) + c_2(f''(1) \pm f''(n)) + \cdots + c_m(f^{(m)}(1) \pm f^{(m)}(n)) + R_m, \tag{7}$$

i.e., a weighted combination of the value of the function and its derivatives at the *endpoints of the summation boundaries* plus an error term R_m. The requirement of the evaluations at the endpoints is what makes this sum interesting.

Let us again look at E_n. We had seen that E_n is the sum of the portions, $I(k)$, of the rectangle that is above the graph of $f(x)$ over each interval $[k, k+1]$. These portions can be expressed as

$$I(k) = \int_k^{k+1} (f(k) - f(x))dx. \tag{8}$$

What is interesting is that the relation above holds for any function f and not just $1/x$. Henceforth, f will be an arbitrary function (not necessarily positive); we will add more assumptions later on, but to begin with we only assume that $\int_1^n f(x)dx$ exists all $n \geq 2$. Integrating by parts, we can express

$$I(k) = [(f(k) - f(x))(x + c)]|_k^{k+1} - \int_k^{k+1} (x + c)f'(x)dx$$

for some constant c. If we choose $c := -(k+1)$ then the first term vanishes completely, and we obtain

$$I(k) = \int_k^{k+1} (x - k - 1)f'(x)dx.$$

Note, however, that $k = [x]$, that is, the largest integer smaller than x. Thus

$$\begin{aligned} E(n) &= \sum_{k=1}^{n} I(k) \\ &= \sum_{k=1}^{n} \int_k^{k+1} (x - [x] - 1)f'(x)dx \\ &= \int_1^n (x - [x])f'(x)dx - \int_1^n f'(x)dx \\ &= \int_1^n (x - [x])f'(x)dx + f(1) - f(n). \end{aligned}$$

Substituting this in (5) we obtain

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx + \int_1^n (x - [x])f'(x)dx + f(1). \tag{9}$$

Note that $P_1(x) := x - [x]$ is a periodic (discontinuous) function with non-negative sign. The equation above gives us more insight into the relation between the sum and the integral than (6). The form of the equation above, however, is not in the form we want: one reason is that the error term in the second integral is large if f' doesn't change sign; also, from an aesthetic viewpoint, the lone $f(1)$ on the RHS looks ugly. Our second step naturally is to apply integration by parts to the second integral in (10), just as we had done for $I(k)$ in (8). Let us again consider the interval $[k, k+1]$:

$$\int_k^{k+1} P_1(x)f'(x)dx = \int_0^1 xf'(x+k)dx = P_2(x)f'(x+k)|_0^1 - \int_0^1 P_2(x)f''(x+k)dx,$$

where P_2 is a function such that $P_2'(x) = P_1(x)$, that is $P_2(x) = \int_0^x P_1(t)dt + c$ for some contant c. If P_2 was periodic with period one, just as P_1 was, then the equation above will be

$$\int_k^{k+1} P_1(x)f'(x)dx = \int_0^1 xf'(x+k)dx = P_2(0)(f'(k+1) - f'(k)) - \int_0^1 P_2(x)f''(x+k)dx$$

and when we sum over k, the difference $f'(k+1) - f'(k)$ telescope out as desired. Now P_2 is periodic iff

$$P_2(x+1) - P_2(x) = \int_x^{x+1} P_1(t)dt = 0.$$

But this is not the case since $\int_x^{x+1} P_1(t)dt = 1$. Thus we have to rectify our choice of P_1 such that it has period one, i.e, its integral on an interval of unit length is zero. How can we do that? If we shift the range of P_1 by $-1/2$ then we still get a polynomial with periodicity one with the desired property that the integral on any interval of unit length is zero. Thus the desired polynomial is $B_1(x) := x - 1/2$, for $x \in [0,1)$. To obtain B_1 in (10) we add and subtract $\int_1^n f'(x)dx/2$ on the RHS of (10) to obtain

$$\begin{aligned}\sum_{k=1}^{<n} f(k) &= \int_1^n f(x)dx + \int_1^n (x - [x] - \frac{1}{2})f'(x)dx + \int_1^n \frac{1}{2}f'(x)dx + f(1) \\ &= \int_1^n f(x)dx + \int_1^n (x - [x] - \frac{1}{2})f'(x)dx + \frac{1}{2}(f(1) + f(n)) \\ &= \int_1^n f(x)dx + \int_1^n B_1(x)f'(x)dx + \frac{1}{2}(f(1) + f(n)).\end{aligned} \tag{10}$$

The polynomial $B_1(x)$ is the first of the Bernoulli's polynomial. Based upon our discussion earlier, we want to define $B_2(x)$ as $\int_0^x B_1(t) + c$, however, to ensure that the leading term is monic we define

$$B_2(x) := 2\int_0^x B_1(t) + c = x^2 - x + c.$$

How do we fix the constant c? Notice that our argument is proceeding inductively: to ensure B_2 is periodic we wanted $\int_0^1 B_1(x)dx = 0$; similarly, to ensure $B_3(x)$ is periodic at the next inductive step, so that the telescoping of $f''(k+1) - f''(k)$ works out as desired, we want $\int_0^1 B_2(x)dx = 0$, i.e., $c = 1/6$. Thus

$$\int_1^n B_1(x)f'(x)dx = \frac{1}{2}B_2(0)(f'(n) - f'(1)) - \frac{1}{2}\int_1^n B_2(x)f''(x)dx.$$

Substituting this in (10) we obtain

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx - \frac{1}{2}\int_1^n B_2(x)f''(x)dx + \frac{B_2(0)}{2}(f'(n) - f'(1)) + \frac{1}{2}(f(1) + f(n)), \tag{11}$$

Again, choose $B_3(x)$ such that $B_3'(x) = 3B_2(x)$; the 3 is to ensure that the leading term after integrating $B_2(x)$ is monic. Thus $B_3(x) = x^3 - 3x^2/2 + x/2$. Thus we have

$$\int_1^n B_2(x)f''(x)dx = \frac{1}{3}B_3(0)(f''(n) - f''(1)) - \frac{1}{3}\int_1^n B_3(x)f'''(x)dx.$$

But note that $B_3(0) = 0$. Thus substituting the above in (11) we obtain

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx + \frac{1}{6}\int_1^n B_3(x)f'''(x)dx + \frac{B_2(0)}{2}(f'(n) - f'(1)) + \frac{1}{2}(f(1) + f(n)), \tag{12}$$

Proceeding in this manner, in general we obtain a form similar to what we desired in (7):

$$\sum_{k=1}^{<n} f(k) = \int_1^n f(x)dx + \sum_{r=1}^{m} \frac{B_r(0)}{r!}(f^{(r-1)}(n) - f^{(r-1)}(1)) + (-1)^m \int_1^n \frac{B_{m+1}}{(m+1)!}f^{(m+1)}(x)dx. \tag{13}$$

The polynomials $B_k(x)$ that satisfy the conditions: for $x \in [0, 1)$

$$B_0(x) := 1, \quad B_k'(x) = kB_{k-1}(x) \quad \text{and} \int_0^1 B_k(x)dx = 0 \tag{14}$$

are called **Bernoulli polynomials** and the numbers $b_k := B_k(0)$ are called the **Bernoulli numbers**. From the second condition it follows that $B_k^{(r)}(x) = r!\binom{k}{r}B_{k-r}(x)$; thus $B_k^{(r)}(0) = r!\binom{k}{r}b_{k-r}$. From Taylor series expansion of a polynomial we know that

$$B_k(x) = \sum_{r=0}^{k} \frac{B_k^{(r)}(0)}{r!}x^r = \sum \binom{k}{r} b_{k-r}x^r.$$

There are many other interesting properties of Bernoulli polynomials.

1. Symmetric: $B_n(1 - x) = (-1)^n B_n(x)$.
2. We can also show that for $x \in [0, 1]$

   $$|B_{2k}(x)| \leq |b_{2k}| \text{ and } |B_{2k+1}(x)| \leq (2k + 1)|b_{2k}|.$$

3. Since $\int_0^1 B_n(x) = 0$, when $n > 0$, it follows that

   $$\sum_{i=0}^{n} \binom{n}{i} \frac{b_i}{n - i + 1} = 0$$

   for all $n > 0$, and it is 1 for $n = 0$. Thinking in terms of convolution, the term on the LHS appears to be the coefficeint of x^n in the product of the egf for the Bernoulli numbers is the inverse of the egf for the series $(1/(i + 1))$; furthermore, the RHS says that the product is exactly 1, i.e., the two series are inverse of each other. It is easy to verify that the egf for the series $(1/(i+1))$ is $(e^x - 1)/x$. Therefore, the egf for the Bernoulli numbers is $x/(e^x - 1)$.