Sums, Products & Asymptotics

1 Closed Forms and Approximations

Sums and products arise regularly in the analysis of algorithms and in other technical areas such as finance and probabilistic systems. We've already seen that

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

Having a simple *closed form* expression such as $n(n+1)/2$ makes the sum a lot easier to understand and evaluate. We proved by induction that this formula is correct, but not where it came from. In Section 4, we'll discuss ways to find such closed forms. Even when there are no closed forms exactly equal to a sum, we may still be able to find a closed form that *approximates* a sum with useful accuracy.

The product we focus on in these notes is the familiar factorial:

$$n! ::= 1 \cdot 2 \cdots (n-1) \cdot n = \prod_{i=1}^{n} i.$$

We'll describe a closed form approximation for it called *Stirling's Formula*.

Finally, when there isn't a good closed form approximation for some expression, there may still be a closed form that characterizes its growth rate. We'll introduce *asymptotic notation*, such as "big Oh", to describe growth rates.

2 The Value of an Annuity

Would you prefer a million dollars today or \$50,000 a year for the rest of your life? On the one hand, instant gratification is nice. On the other hand, the total dollars received at \$50K per year is much larger if you live long enough.

Formally, this is a question about the value of an annuity. An *annuity* is a financial instrument that pays out a fixed amount of money at the beginning of every year for some specified number of years. In particular, an n-year, m-payment annuity pays m dollars at the start of each year for n years. In some cases, n is finite, but not always. Examples include lottery payouts, student loans, and home mortgages. There are even Wall Street people who specialize in trading annuities.

A key question is what an annuity is worth. For example, lotteries often pay out jackpots over many years. Intuitively, $\$50,000$ a year for 20 years ought to be worth less than a million dollars right now. If you had all the cash right away, you could invest it and begin collecting interest. But what if the choice were between $\$50,000$ a year for 20 years and a *half* million dollars today? Now it is not clear which option is better.

In order to answer such questions, we need to know what a dollar paid out in the future is worth today. To model this, let's assume that money can be invested at a fixed annual interest rate p. We'll assume an 8% rate[1] for the rest of the discussion.

Here is why the interest rate p matters. Ten dollars invested today at interest rate p will become $(1+p) \cdot 10 = 10.80$ dollars in a year, $(1+p)^2 \cdot 10 \approx 11.66$ dollars in two years, and so forth. Looked at another way, ten dollars paid out a year from now are only really worth $1/(1+p) \cdot 10 \approx 9.26$ dollars today. The reason is that if we had the $\$9.26$ today, we could invest it and would have $\$10.00$ in a year anyway. Therefore, p determines the value of money paid out in the future.

2.1 The Future Value of Money

Our goal is to determine the value of an n-year, m-payment annuity. The first payment of m dollars is truly worth m dollars. But the second payment a year later is worth only $m/(1+p)$ dollars. Similarly, the third payment is worth $m/(1+p)^2$, and the n-th payment is worth only $m/(1+p)^{n-1}$. The total value V of the annuity is equal to the sum of the payment values. This gives:

$$V = \sum_{i=1}^{n} \frac{m}{(1+p)^{i-1}}.$$

To compute the real value of the annuity, we need to evaluate this sum. One way is to plug in m, n, and p, compute each term explicitly, and then add them up. However, this sum has a special closed form that makes the job easier. (The phrase "closed form" refers to a mathematical expression without any summation or product notation.) First, lets make the summation prettier with some substitutions.

$$\begin{aligned} V &= \sum_{i=1}^{n} \frac{m}{(1+p)^{i-1}} \\ &= \sum_{j=0}^{n-1} \frac{m}{(1+p)^{j}} \quad \text{(substitute } j = i-1\text{)} \\ &= m \sum_{j=0}^{n-1} x^j \quad \text{(substitute } x = \frac{1}{1+p}\text{).} \end{aligned}$$

The goal of these substitutions is to put the summation into a special form so that we can bash it with a theorem given in the next section.

[1] U.S. interest rates have dropped steadily for several years, and ordinary bank deposits now earn around 3%. But just a few years ago the rate was 8%; this rate makes some of our examples a little more dramatic. The rate has been as high as 17% in the past twenty years.

In Japan, the standard interest rate is near zero%, and on a few ocasions in the past few years has even been slightly negative. It's a mystery to U.S. economists why the Japanese populace keeps any money in their banks.

2.2 Geometric Sums

Theorem 2.1. *For all $n \geq 1$ and all $x \neq 1$,*

$$\sum_{i=0}^{n-1} x^i = \frac{1-x^n}{1-x}.$$

The summation in this theorem is a *geometric sum*. The distinguishing feature of a geometric sum is that each of the terms

$$1,\ x,\ x^2,\ x^3,\ \ldots,\ x^{n-1}.$$

in the sum is a constant times the one before; in this case, the constant is x. The theorem gives a closed form for a geometric sum that starts with 1.

We already saw one proof of this theorem in our lectures on induction. As is often the case, the proof by induction gives no hint about how the formula was found in the first place. Here is a more insightful derivation. The trick is to let S be the value of the sum and then observe what $-xS$ is:

$$\begin{array}{rcllllll} S &=& 1 & +x & +x^2 & +x^3 & + \quad \cdots & +x^{n-1} \\ -xS &=& & -x & -x^2 & -x^3 & - \quad \cdots & -x^{n-1} - x^n. \end{array}$$

Adding these two equations gives:

$$S - xS = 1 - x^n,$$

so

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

We'll say more about finding (as opposed to just proving) summation formulas later.

2.3 Return of the Annuity Problem

Now we can solve the annuity pricing problem. The value of an annuity that pays m dollars at the start of each year for n years is computed as follows:

$$\begin{aligned} V &= m \sum_{j=0}^{n-1} x^j \\ &= m \frac{1-x^n}{1-x} \\ &= m \frac{1 - (\frac{1}{1+p})^n}{1 - \frac{1}{1+p}} \\ &= m \frac{1 + p - (\frac{1}{1+p})^{n-1}}{p}. \end{aligned}$$

The first line is a restatement of the summation we obtained earlier for the value of an annuity. The second line uses the closed form formula for a geometric sum. In the third line, we undo the earlier substitution $x = 1/(1+p)$. In the final step, both the numerator and denominator are multiplied by $1+p$ to simplify the expression.

The resulting formula is much easier to use than a summation with dozens of terms. For example, what is the real value of a winning lottery ticket that pays $\$50,000$ per year for 20 years? Plugging in $m = \$50,000$, $n = 20$, and $p = 0.08$ gives $V \approx \$530,180$. Because payments are deferred, the million dollar lottery is really only worth about a half million dollars! This is a good trick for the lottery advertisers!

2.4 Infinite Geometric Series

The question at the beginning of this section was whether you would prefer a million dollars today or $\$50,000$ a year for the rest of your life. Of course, this depends on how long you live, so optimistically assume that the second option is to receive $\$50,000$ a year *forever*. This sounds like infinite money!

We can compute the value of an annuity with an infinite number of payments by taking the limit of our geometric sum in Theorem 2.1 as n tends to infinity. This one is worth remembering!

Theorem 2.2. *If* $|x| < 1$, *then*

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}.$$

Proof.

$$\begin{aligned} \sum_{i=0}^{\infty} x^i &= \lim_{n\to\infty} \sum_{i=0}^{n-1} x^i \\ &= \lim_{n\to\infty} \frac{1-x^n}{1-x} \\ &= \frac{1}{1-x}. \end{aligned}$$

The first equality follows from the definition of an infinite summation. In the second line, we apply the formula for the sum of an n-term geometric sum given in Theorem 2.1. The final line follows by evaluating the limit; the x^n term vanishes since we assumed that $|x| < 1$. □

In our annuity problem, $x = 1/(1+p) < 1$, so the theorem applies. Substituting for x, we get an annuity value of

$$\begin{aligned} V &= m \cdot \frac{1}{1-x} \\ &= m \cdot \frac{1}{1 - 1/(1+p)} \\ &= m \cdot \frac{1+p}{(1+p) - 1} \\ &= m \cdot \frac{1+p}{p}. \end{aligned}$$

Plugging in $m = \$50,000$ and $p = 0.08$ gives only $\$675,000$. Amazingly, a million dollars today is worth much more than $\$50,000$ paid every year forever! Then again, if we had a million dollars today in the bank earning 8% interest, we could take out and spend $\$80,000$ a year forever. So the answer makes some sense.

2.5 Examples

We now have closed form formulas for geometric sums and series. Some examples are given below. In each case, the solution follows immediately from either Theorem 2.1 (for finite sums) or Theorem 2.2 (for infinite series).

$$1 + 1/2 + 1/4 + 1/8 + \cdots = \sum_{i=0}^{\infty} (1/2)^i = \frac{1}{1-(1/2)} = 2 \tag{1}$$

$$0.999999999\ldots = 0.9 \sum_{i=0}^{\infty} (1/10)^i = 0.9 \frac{1}{1-1/10} = 0.9 \frac{10}{9} = 1 \tag{2}$$

$$1 - 1/2 + 1/4 - 1/8 + \cdots = \sum_{i=0}^{\infty} (-1/2)^i = \frac{1}{1-(-1/2)} = 2/3 \tag{3}$$

$$1 + 2 + 4 + 8 + \cdots + 2^{n-1} = \sum_{i=0}^{n-1} 2^i = \frac{1-2^n}{1-2} = 2^n - 1 \tag{4}$$

$$1 + 3 + 9 + 27 + \cdots + 3^{n-1} = \sum_{i=0}^{n-1} 3^i = \frac{1-3^n}{1-3} = \frac{3^n - 1}{2} \tag{5}$$

If the terms in a geometric sum or series grow smaller, as in equation (1), then the sum is said to be *geometrically decreasing*. If the terms in a geometric sum grow progressively larger, as in (4) and (5), then the sum is said to be *geometrically increasing*.

Here is a good rule of thumb: *a geometric sum or series is approximately equal to the term with greatest absolute value*. In equations (1) and (3), the largest term is equal to 1 and the sums are 2 and 2/3, both relatively close to 1. In equation (4), the sum is about twice the largest term. In the final equation (5), the largest term is 3^{n-1} and the sum is $(3^n - 1)/2$, which is only about a factor of 1.5 greater.

2.6 Related Sums

We now know all about geometric sums. But in practice one often encounters sums that cannot be transformed by simple variable substitutions to the form $\sum x^i$.

A non-obvious, but useful way to obtain new summation formulas from old is by differentiating or integrating with respect to x. As an example, consider the following sum:

$$\sum_{i=1}^{n} i x^i = x + 2x^2 + 3x^3 + \cdots + nx^n$$

This is not a geometric sum, since the ratio between successive terms is not constant. Our formula for the sum of a geometric sum cannot be directly applied. But suppose that we differentiate that formula:

$$\begin{aligned}
\frac{d}{dx}\sum_{i=0}^{n} x^i &= \frac{d}{dx}\frac{1-x^{n+1}}{1-x} \\
\sum_{i=1}^{n} ix^{i-1} &= \frac{-(n+1)x^n(1-x)-(-1)(1-x^{n+1})}{(1-x)^2} \\
&= \frac{-(n+1)x^n+(n+1)x^{n+1}+1-x^{n+1}}{(1-x)^2} \\
&= \frac{1-(n+1)x^n+nx^{n+1}}{(1-x)^2}.
\end{aligned}$$

Often differentiating or integrating messes up the exponent of x in every term. In this case, we now have a formula for a sum of the form $\sum ix^{i-1}$, but we want a formula for the series $\sum ix^i$. The solution is simple: multiply by x. This gives:

$$\sum_{i=1}^{n} ix^i = \frac{x-(n+1)x^{n+1}+nx^{n+2}}{(1-x)^2}$$

Since we could easily have made a mistake, it is a good idea to go back and validate a formula obtained this way with a proof by induction.

Notice that if $|x| < 1$, then this series converges to a finite value even if there are infinitely many terms. Taking the limit as n tends infinity gives the following theorem:

Theorem 2.3. *If* $|x| < 1$, *then*

$$\sum_{i=1}^{\infty} ix^i = \frac{x}{(1-x)^2}.$$

As a consequence, suppose there is an annuity that pays im dollars at the *end* of each year i forever. For example, if $m = \$50,000$, then the payouts are $\$50,000$ and then $\$100,000$ and then $\$150,000$ and so on. It is hard to believe that the value of this annuity is finite! But we can use the preceding theorem to compute the value:

$$\begin{aligned}
V &= \sum_{i=1}^{\infty} \frac{im}{(1+p)^i} \\
&= m\frac{\frac{1}{1+p}}{(1-\frac{1}{1+p})^2} \\
&= m\frac{1+p}{p^2}.
\end{aligned}$$

The second line follows by an application of Theorem 2.3. The third line is obtained by multiplying the numerator and denominator by $(1+p)^2$.

For example, if $m = \$50,000$, and $p = 0.08$ as usual, then the value of the annuity is $V = \$8,437,500$. Even though payments increase every year, the increase is only additive with time; by contrast, dollars paid out in the future decrease in value exponentially with time. The geometric decrease swamps out the additive increase. Payments in the distant future are almost worthless, so the value of the annuity is finite.

The important thing to remember is the trick of taking the derivative (or integral) of a summation formula. Of course, this technique requires one to compute nasty derivatives correctly, but this is at least theoretically possible!

3 Book Stacking

Suppose you have a pile of books and you want to stack them on a table in some off-center way so the top book sticks out past books below it. How far past the edge of the table do you think you could get the top book to go without having the stack fall over? Could the top book stick out completely beyond the edge of table?

Most people's first response to this question—sometimes also their second and third responses—is "No, the top book will never get completely past the edge of the table." But in fact, you can get the top book to stick out as far as you want: one booklength, two booklengths, any number of booklengths!

3.1 Formalizing the Problem

We'll approach this problem recursively. How far past the end of the table can we get one book to stick out? It won't tip as long as its center of mass is over the table, so we can get it to stick out half its length, as shown in Figure 1.

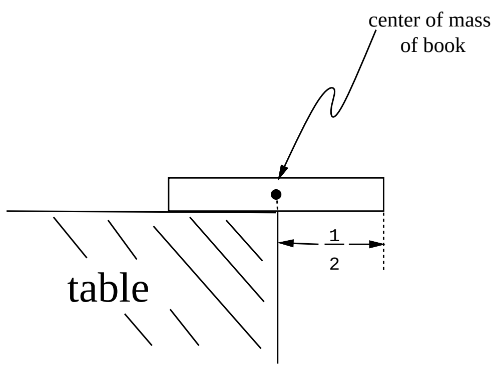

Figure 1: One book can overhang half a book length.

Now suppose we have a stack of books that will stick out past the table edge without tipping over—call that a *stable* stack. Let's define the *overhang* of a stable stack to be the largest horizontal distance from the center of mass of the stack to the furthest edge of a book. If we place the center of mass of the stable stack at the edge of the table as in Figure 2, that's how far we can get a book in the stack to stick out past the edge.

So we want a formula for the maximum possible overhang, B_n, achievable with a stack of n books.

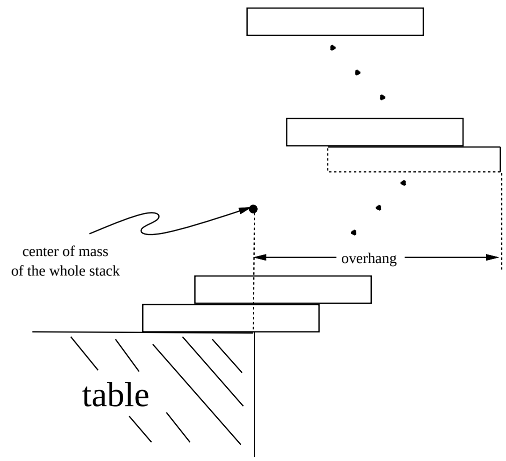

Figure 2: Overhanging the edge of the table.

We've already observed that the overhang of one book is 1/2 a book length. That is,

$$B_1 = \frac{1}{2}.$$

Now suppose we have a stable stack of $n+1$ books with maximum overhang. If the overhang of the n books on top of the bottom book was not maximum, we could get a book to stick out further by replacing the top stack with a stack of n books with larger overhang. So the maximum overhang, B_{n+1}, of a stack of $n+1$ books is obtained by placing a maximum overhang stable stack of n books on top of the bottom book. And we get the biggest overhang for the stack of $n+1$ books by placing the center of mass of the n books right over the edge of the bottom book as in Figure 3.

So we know where to place the $n+1$st book to get maximum overhang, and all we have to do is calculate what it is. The simplest way to do that is to let the center of mass of the top n books be the origin. That way the horizontal coordinate of the center of mass of the whole stack of $n+1$ books will equal the increase in the overhang. But now the center of mass of the bottom book has horizontal coordinate $1/2$, so the horizontal coordinate of center of mass of the whole stack of $n+1$ books is

$$\frac{0 \cdot n + (1/2) \cdot 1}{n+1} = \frac{1}{2(n+1)}.$$

In other words,

$$B_{n+1} = B_n + \frac{1}{2(n+1)}, \tag{6}$$

as shown in Figure 3.

Expanding equation (6), we have

$$\begin{aligned} B_{n+1} &= B_{n-1} + \frac{1}{2n} + \frac{1}{2(n+1)} \\ &= B_1 + \frac{1}{2 \cdot 2} + \cdots + \frac{1}{2n} + \frac{1}{2(n+1)} \\ &= \frac{1}{2} \sum_{i=1}^{n+1} \frac{1}{i}. \end{aligned}$$

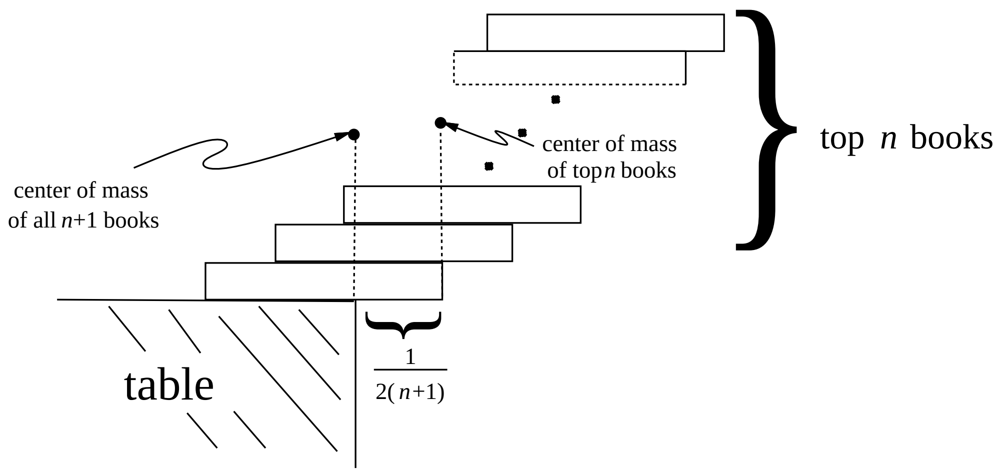

Figure 3: Additional overhang with $n + 1$ books.

Define

$$H_n ::= \sum_{i=1}^{n} \frac{1}{i}.$$

H_n is called the nth *Harmonic number*, and we have just shown that

$$B_n = \frac{H_n}{2}.$$

The first few Harmonic numbers are easy to compute. For example, $H_1 = 1$, $H_2 = 1 + \frac{1}{2} = \frac{3}{2}$, $H_3 = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}$, $H_4 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{25}{12}$. The fact that H_4 is greater than 2 has special significance; it implies that the total extension of a 4-book stack is greater than one full book! This is the situation shown in Figure 4.

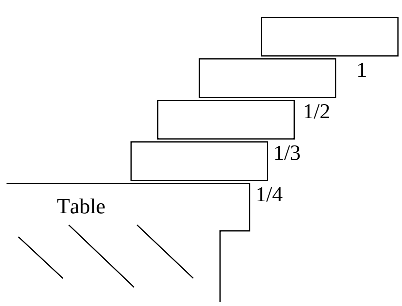

Figure 4: Stack of four books with maximum overhang.

In the next section we will prove that H_n grows slowly, but *unboundedly* with n. That means we can get books to overhang *any distance* past the edge of the table by piling them high enough!

3.2 Evaluating the Sum—The Integral Method

It would be nice to answer questions like, "How many books are needed to build a stack extending 100 book lengths beyond the table?" One approach to this question would be to keep computing Harmonic numbers until we found one exceeding 200. However, as we will see, this is not such a keen idea.

Such questions would be settled if we could express H_n in a closed form. Unfortunately, no closed form is known, and probably none exists. As a second best, however, we can find closed forms for very good approximations to H_n using the Integral Method. The idea of the Integral Method is to bound terms of the sum above and below by simple functions as suggested in Figure 5. The integrals of these functions then bound the value of the sum above and below.

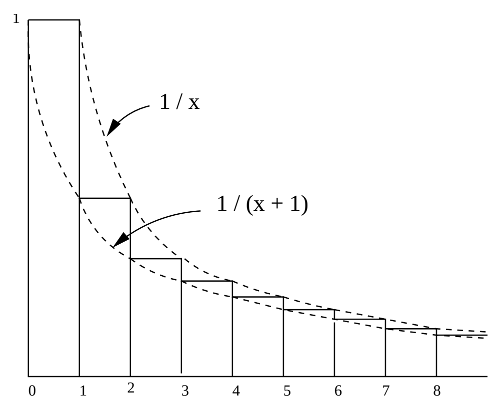

Figure 5: *This figure illustrates the Integral Method for bounding a sum. The area under the "stairstep" curve over the interval* $[0, n]$ *is equal to* $H_n = \sum_{i=1}^{n} 1/i$. *The function* $1/x$ *is everywhere greater than or equal to the stairstep and so the integral of* $1/x$ *over this interval is an upper bound on the sum. Similarly,* $1/(x+1)$ *is everywhere less than or equal to the stairstep and so the integral of* $1/(x+1)$ *is a lower bound on the sum.*

The Integral Method gives the following upper and lower bounds on the harmonic number H_n:

$$\begin{aligned} H_n &\leq 1 + \int_1^n \frac{1}{x}\,dx = 1 + \ln n \qquad (7) \\ H_n &\geq \int_0^n \frac{1}{x+1}\,dx = \int_1^{n+1} \frac{1}{x}\,dx = \ln(n+1). \end{aligned}$$

These bounds imply that the harmonic number H_n is around $\ln n$. Since $\ln n$ grows without bound, albeit slowly, we can make a stack of books that extends arbitrarily far.

For example, to build a stack extending three book lengths beyond the table, we need a number of books n so that $H_n \geq 6$. Exponentiating the above inequalities gives

$$e^{H_n - 1} \leq n \leq e^{H_n} - 1.$$

This implies that we will need somewhere between 149 and 402 books. Actual calculation of H_n shows that 227 books will be the minimum number to overhang three book lengths.

3.3 More about Harmonic Numbers

In the preceding section, we showed that H_n is about $\ln n$. A even better approximation is known:

$$H_n = \ln n + \gamma + \frac{1}{2n} + \frac{1}{12n^2} + \frac{\epsilon(n)}{120n^4}$$

Here γ is a value $0.577215664\ldots$ called Euler's constant, and $\epsilon(n)$ is between 0 and 1 for all n. We will not prove this formula.

The shorthand $H_n \sim \ln n$ is used to indicate that the leading term of H_n is $\ln n$. More precisely:

Definition 3.1. For functions $f, g : \mathbb{R} \to \mathbb{R}$, we say f is *asymptotically equal* to g, in symbols,

$$f(x) \sim g(x)$$

iff

$$\lim_{x\to\infty} f(x)/g(x) = 1.$$

We also might write $H_n \sim \ln n + \gamma$ to indicate two leading terms. While this notation is widely used, it is not really right. Referring to the definition of $\sim$, we see that while $H_n \sim \ln n + \gamma$ is a true statement, so is $H_n \sim \ln n + c$ where c is any constant. The correct way to indicate that γ is the second-largest term is $H_n - \ln n \sim \gamma$.

The reason that the $\sim$ notation is useful is that often we do not care about lower order terms. For example, if $n = 100$, then we can compute $H(n)$ to great precision using only the two leading terms:

$$|H_n - \ln n - \gamma| \leq \left|\frac{1}{200} - \frac{1}{120000} + \frac{1}{120 \cdot 100^4}\right| < \frac{1}{200}.$$

4 Finding Summation Formulas

The source of the simple formula $\sum_{i=1}^{n} i = n(n+1)/2$ is still a mystery! Sure, we can prove this statement true by induction, but where did the expression on the right come from? Even more inexplicable is the summation formula for consecutive squares:

$$\begin{aligned}\sum_{i=1}^{n} i^2 &= \frac{(2n+1)(n+1)n}{6} \\ &= \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \\ &\sim \frac{n^3}{3}.\end{aligned}$$

Here is how we might find the sum-of-squares formula if we forgot it or had never seen it. First, the Integral Method gives a quick estimate of the sum:

$$\int_0^n x^2\,dx \leq \sum_{i=1}^{n} i^2 \leq \int_0^n (x+1)^2\,dx$$

$$\frac{n^3}{3} \leq \sum_{i=1}^{n} i^2 \leq \frac{(n+1)^3}{3} - \frac{1}{3}.$$

These upper and lower bounds obtained by the Integral Method show that $\sum_{i=1}^{n} i^2 \sim n^3/3$. To get an exact formula, we then guess the general form of the solution. Where we are uncertain, we can add parameters $a, b, c, \ldots$. For example, we might make the guess:

$$\sum_{i=1}^{n} i^2 = an^3 + bn^2 + cn + d.$$

If the guess is correct, then we can determine the parameters a, b, c, and d by plugging in a few values for n. Each such value gives a linear equation in a, b, c, and d. If we plug in enough values, we may get a linear system with a unique solution. Applying this method to our example gives:

$$\begin{aligned} n = 0 \quad &\rightarrow \quad 0 = d \\ n = 1 \quad &\rightarrow \quad 1 = a + b + c + d \\ n = 2 \quad &\rightarrow \quad 5 = 8a + 4b + 2c + d \\ n = 3 \quad &\rightarrow \quad 14 = 27a + 9b + 3c + d. \end{aligned}$$

Solving this system gives the solution $a = 1/3$, $b = 1/2$, $c = 1/6$, $d = 0$. Therefore, if our initial guess at the form of the solution was correct, then the summation is equal to $n^3/3 + n^2/2 + n/6$. In fact, our initial guess *was* correct, this is the right formula for the sum of squares!

Be careful! After obtaining a formula by this method, always go back and prove it using induction or some other method. This is not merely a check for algebra blunders; if the initial guess at the solution was not of the right form, then the resulting formula will be completely wrong!

5 Double Sums

Sometimes we have to evaluate sums of sums, otherwise known as *double summations*. Sometimes it is easy: we can evaluate the inner sum, replace it with a closed form, and then evaluate the outer sum which no longer has a summation inside it.

But there's a special trick that is often extremely useful for sums, which is *exchanging the order of summation.* It's best demonstrated by example. Suppose we want to compute the sum of the harmonic numbers

$$\sum_{k=1}^{n} H_k = \sum_{k=1}^{n} \sum_{j=1}^{k} 1/j$$

For intuition about this sum, we can try the integral method:

$$\sum_{k=1}^{n} H_k \approx \int_1^n \ln x \, dx \approx n \ln n - n.$$

Now let's look for an exact answer. If we think about the pairs (k, j) over which we are summing, they form a triangle:

		j: 1	2	3	4	5	...	n
k	1	1						
	2	1	1/2					
	3	1	1/2	1/3				
	4	1	1/2	1/3	1/4			
		...						
	n	1	1/2		...			1/n

The summation above is summing each row and then adding the row sums. Instead, we can sum the columns and then add the column sums. Inspecting the table we see that this double sum can be written as

$$\begin{aligned}
\sum_{k=1}^{n} H_k &= \sum_{k=1}^{n} \sum_{j=1}^{k} 1/j \\
&= \sum_{j=1}^{n} \sum_{k=j}^{n} 1/j \\
&= \sum_{j=1}^{n} 1/j \sum_{k=j}^{n} 1 \\
&= \sum_{j=1}^{n} \frac{1}{j}(n-j+1) \\
&= \sum_{j=1}^{n} \frac{n-j+1}{j} \\
&= \sum_{j=1}^{n} \frac{n+1}{j} - \sum_{j=1}^{n} \frac{j}{j} \\
&= (n+1)\sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{n} 1 \\
&= (n+1)H_n - n. \qquad (8)
\end{aligned}$$

6 Stirling's Approximation

The familiar factorial notation, $n!$, is an abbreviation for the product

$$\prod_{i=1}^{n} i.$$

This is by far the most common product in Discrete Mathematics. In this section we describe a good closed-form estimate of $n!$ called *Stirling's Approximation*. Unfortunately, all we can do is estimate: there is no closed form for $n!$ — though proving so would take us beyond the scope of 6.042.

6.1 Products to Sums

A good way to handle a product is often to convert it into a sum by taking the logarithm. In the case of factorial, this gives

$$\begin{aligned}\ln(n!) &= \ln(1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n) \\ &= \ln 1 + \ln 2 + \ln 3 + \cdots + \ln(n-1) + \ln n \\ &= \sum_{i=1}^{n} \ln i.\end{aligned}$$

We've not seen a summation containing a logarithm before! Fortunately, one tool that we used in evaluating sums is still applicable: the Integral Method. We can bound the terms of this sum with $\ln x$ and $\ln(x+1)$ as shown in Figure 6. This gives bounds on $\ln(n!)$ as follows:

$$\begin{aligned}\int_1^n \ln x \, dx \leq & \quad \textstyle\sum_{i=1}^n \ln i & \leq \int_0^n \ln(x+1)\, dx \\ n \ln(\frac{n}{e}) + 1 \leq & \quad \textstyle\sum_{i=1}^n \ln i & \leq (n+1)\ln\left(\frac{n+1}{e}\right) + 1 \\ \left(\frac{n}{e}\right)^n e \leq & \quad n! & \leq \left(\frac{n+1}{e}\right)^{n+1} e.\end{aligned}$$

The second line follows from the first by completing the integrations. The third line is obtained by exponentiating.

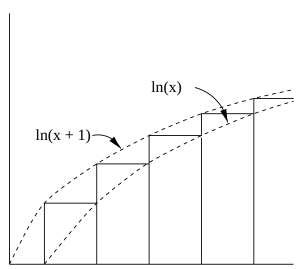

Figure 6: *This figure illustrates the Integral Method for bounding the sum* $\sum_{i=1}^n \ln i$.

So $n!$ behaves something like the closed form formula $(\frac{n}{e})^n$. A more careful analysis yields an unexpected closed form formula that is asymptotically exact:

Lemma (Stirling's Formula).

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

Stirling's Formula describes how $n!$ behaves in the limit, but to use it effectively, we need to know how close it is to the limit for different values of n. That information is given by the bounding formulas:

Fact (Stirling's Approximation).

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)} \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}.$$

The Approximation implies the asymptotic Formula, since since $e^{1/(12n+1)}$ and $e^{1/12n}$ both approach 1 as n grows large. These inequalities can be verified by induction, but the details are nasty.

The bounds in Stirling's formula are very tight. For example, if $n = 100$, then Stirling's bounds are:

$$\begin{aligned} 100! &\geq \sqrt{200\pi}\left(\frac{100}{e}\right)^{100} e^{1/1201} \\ 100! &\leq \sqrt{200\pi}\left(\frac{100}{e}\right)^{100} e^{1/1200} \end{aligned}$$

The only difference between the upper bound and the lower bound is in the final term. In particular $e^{1/1201} \approx 1.00083299$ and $e^{1/1200} \approx 1.00083368$. As a result, the upper bound is no more than $1 + 10^{-6}$ times the lower bound. This is amazingly tight! Remember Stirling's formula; we will use it often.

6.2 Bounds by Double Summing

Another way to derive Stirling's approximation is to remember that $\ln n$ is roughly the same as H_n. This lets us use the result we derived before for $\sum H_k$ via double summation. Our approximation for H_k told us that $\ln(k+1) \leq H_k \leq 1 + \ln k$. Rewriting, we find that $H_k - 1 \leq \ln k \leq H_{k-1}$. It follows that (leaving out the $i = 1$ term in the sum, which contributes 0),

$$\begin{aligned} \sum_{i=2}^{n} \ln i &\leq \sum_{i=2}^{n} H_{i-1} \\ &= \sum_{i=1}^{n-1} H_i \\ &= nH_{n-1} - (n-1) && \text{by (8)} \\ &\leq n(1 + \ln(n-1)) - (n-1) && \text{by (7)} \\ &= n\ln(n-1) + 1, \end{aligned}$$

roughly the same bound as we proved before via the integral method. We can derive a similar lower bound.

7 Asymptotic Notation

Asymptotic notation is a shorthand used to give a quick measure of the behavior of a function $f(n)$ as n grows large.

7.1 Little Oh

The asymptotic notation $\sim$ is an equivalence relation indicating that $\sim$-equivalent functions grow at exactly the same rate. There is a corresponding strict partial order on functions indicating that one function grows at a significantly slower rate. Namely,

Definition 7.1. For functions $f, g : \mathbb{R} \to \mathbb{R}$, we say f is *asymptotically smaller* than g, in symbols,

$$f(x) = o(g(x)),$$

iff

$$\lim_{x\to\infty} f(x)/g(x) = 0.$$

For example, $1000x^{1.9} = o(x^2)$, because $1000x^{1.9}/x^2 = 1000/x^{0.1}$ and since $x^{0.1}$ goes to infinity with x and 1000 is constant, we have $\lim_{x\to\infty} 1000x^{1.9}/x^2 = 0$. This argument generalizes directly to yield

Lemma 7.2. *$x^a = o(x^b)$ for all nonnegative constants $a < b$.*

Using the familiar fact that $\log x < x$ for all $x > 1$, we can prove

Lemma 7.3. *$\log x = o(x^\epsilon)$ for all $\epsilon > 0$ and $x > 1$.*

Proof. Choose $\epsilon > \delta > 0$ and let $x = z^\delta$ in the inequality $\log x < x$. This implies

$$\log z < z^\delta/\delta = o(z^\epsilon) \qquad \text{by Lemma 7.2.} \tag{9}$$

□

Corollary 7.4. *$x^b = o(a^x)$ for any $a, b \in \mathbb{R}$ with $a > 1$.*

Proof. From (9),

$$\log z < z^\delta/\delta$$

for all $z > 1$, $\delta > 0$. Hence

$$\begin{aligned}
(e^b)^{\log z} &< (e^b)^{z^\delta/\delta} \\
z^b &< \left(e^{\log a(b/\log a)}\right)^{z^\delta/\delta} \\
&= a^{(b/\delta\log a)z^\delta} \\
&< a^z
\end{aligned}$$

for all z such that $(b/\delta\log a)z^\delta < z$. But since $z^\delta = o(z)$, this last inequality holds for all large enough z. □

Lemma 7.3 and Corollary 7.4 can also be proved easily in several other ways, *e.g.*, using L'Hopital's Rule or the McLaurin Series for $\log x$ and e^x. Proofs can be found in most calculus texts.

Problem 1. Prove the initial claim that $\log x < x$ for all $x > 1$ (requires elementary calculus).

Problem 2. Prove that the relation, R, on functions such that fRg iff $f = o(g)$ is a strict partial order, namely, R is transitive and *asymmetric*: if fRg then $\neg gRf$.

Problem 3. Prove that $f \sim g$ iff $f = g + h$ for some function $h = o(g)$.

7.2 Big Oh

Big Oh is the most frequently used asymptotic notation. It is used to give an upper bound on the growth of a function, such as the running time of an algorithm.

Definition 7.5. Given functions $f, g : \mathbb{R} \to \mathbb{R}$, with g nonnegative, we say that

$$f = O(g)$$

iff

$$\limsup_{x\to\infty} |f(x)| / g(x) < \infty.$$

This definition[2] makes it clear that

Lemma 7.6. *If $f = o(g)$ or $f \sim g$, then $f = O(g)$.*

Proof. $\lim f/g = 0$ or $\lim f/g = 1$ implies $\lim f/g < \infty$. □

It is easy to see that the converse of Lemma 7.6 is not true. For example, $2x = O(x)$, but $2x \not\sim x$ and $2x \neq o(x)$.

The usual formulation of Big Oh spells out the definition of $\limsup$ without mentioning it. Namely, here is an equivalent definition:

Definition 7.7. Given functions $f, g : \mathbb{R} \to \mathbb{R}$, we say that

$$f = O(g)$$

iff there exists a constant $c \geq 0$ and an x_0 such that for all $x \geq x_0$, $|f(x)| \leq cg(x)$.

This definition is rather complicated, but the idea is simple: $f(x) = O(g(x))$ means $f(x)$ is less than or equal to $g(x)$, except that we're willing to ignore a constant factor, *viz.*, c, and to allow exceptions for small x, *viz.*, $x < x_0$.

We observe,

Lemma 7.8. *Assume that g is nonnegative. If $f = o(g)$, then it is* not *true that $g = O(f)$.*

Proof.

$$\lim_{x\to\infty} \frac{g(x)}{f(x)} = \frac{1}{\lim_{x\to\infty} \frac{f(x)}{g(x)}} = \frac{1}{0} = \infty,$$

so $g \neq O(f)$.

□

---

[2]

$$\limsup_{x\to\infty} h(x) ::= \lim_{x\to\infty} \text{lub}_{y \geq x} h(y).$$

We need the $\limsup$ in the definition of $O()$ because if $f(x)/g(x)$ oscillates between, say, 3 and 5 as x grows, then $f = O(g)$ because $f \leq 5g$, but $\lim_{x\to\infty} f(x)/g(x)$ does not exist. However, in this case we would have $\limsup_{x\to\infty} f(x)/g(x) = 5$.

Proposition 7.9. $100x^2 = O(x^2)$.

Proof. Choose $c = 100$ and $x_0 = 1$. Then the proposition holds, since for all $x \geq 1$, $\left|100x^2\right| \leq 100x^2$. □

Proposition 7.10. $x^2 + 100x + 10 = O(x^2)$.

Proof. $(x^2+100x+10)/x^2 = 1+100/x+10/x^2$ and so its limit as x approaches infinity is $1+0+0 = 1$. So in fact, $x^2 + 100x + 10 \sim x^2$, and therefore $x^2 + 100x + 10 = O(x^2)$. Indeed, it's conversely true that $x^2 = O(x^2 + 100x + 10)$. □

Proposition 7.10 generalizes to an arbitrary polynomial by a similar proof, which we omit.

Proposition 7.11. *For $a_k \neq 0$, $a_k x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 = O(x^k)$.*

Big Oh notation is especially useful when describing the running time of an algorithm. For example, the usual algorithm for multiplying $n \times n$ matrices requires proportional to n^3 operations in the worst case. This fact can be expressed concisely by saying that the running time is $O(n^3)$. So this asymptotic notation allows the speed of the algorithm to be discussed without reference to constant factors or lower-order terms that might be machine specific. In this case there is another, ingenious matrix multiplication procedure that requires $O(n^{2.55})$ operations. This procedure will therefore be much more efficient on large enough matrices. Unfortunately, the $O(n^{2.55})$-operation multiplication procedure is almost never used because it happens to be less efficient than the usual $O(n^3)$ procedure on matrices of practical size. It is even conceivable that there is an $O(n^2)$ matrix multiplication procedure, but none is known.

7.3 Theta

Definition 7.12.

$$f = \Theta(g) \quad \text{iff} \quad f = O(g) \wedge g = O(f).$$

The statement $f = \Theta(g)$ can be paraphrased intuitively as "f and g are equal to within a constant factor."

The value of these notations is that they highlight growth rates and allow suppression of distracting factors and low-order terms. For example, if the running time of an algorithm is

$$T(n) = 10n^3 - 20n^2 + 1,$$

then

$$T(n) = \Theta(n^3).$$

In this case, we would say that T *is of order* n^3 or that $T(n)$ *grows cubically*.

Another such example is

$$\pi^2 3^{x-7} + \frac{(2.7x^{113} + x^9 - 86)^4}{\sqrt{x}} - 1.08^{3x} = \Theta(3^x).$$

Just knowing that the running time of an algorithm is $\Theta(n^3)$, for example, is useful, because if n doubles we can predict that the running time will *by and large*[3] increase by a factor of at most 8 for large n. In this way, Theta notation preserves information about the scalability of an algorithm or system. Scalability is, of course, a big issue in the design of algorithms and systems.

7.4 Pitfalls with Big Oh

There is a long list of ways to make mistakes with Big Oh notation. This section presents some of the ways that Big Oh notation can lead to ruin and despair.

7.4.1 The Exponential Fiasco

Sometimes relationships involving Big Oh are not so obvious. For example, one might guess that $4^x = O(2^x)$ since 4 is only a constant factor larger than 2. This reasoning is incorrect, however; actually 4^x grows much faster than 2^x.

Proposition 7.13. $4^x \neq O(2^x)$

Proof. $2^x/4^x = 2^x/(2^x 2^x) = 1/2^x$. Hence, $\lim_{x\to\infty} 2^x/4^x = 0$, so in fact $2^x = o(4^x)$. We observed earlier that this implies that $4^x \neq O(2^x)$. □

7.4.2 Constant Confusion

Every constant is $O(1)$. For example, $17 = O(1)$. This is true because if we let $f(x) = 17$ and $g(x) = 1$, then there exists a $c > 0$ and an x_0 such that $|f(x)| \leq cg(x)$. In particular, we could choose c = 17 and $x_0 = 1$, since $|17| \leq 17 \cdot 1$ for all $x \geq 1$. We can construct a false theorem that exploits this fact.

False Theorem 7.14.

$$\sum_{i=1}^{n} i = O(n)$$

False proof. Define $f(n) = \sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n$. Since we have shown that every constant i is $O(1)$, $f(n) = O(1) + O(1) + \cdots + O(1) = O(n)$. □

Of course in reality $\sum_{i=1}^{n} i = n(n+1)/2 \neq O(n)$.

The error stems from confusion over what is meant in the statement $i = O(1)$. For any *constant* $i \in \mathbb{N}$ it is true that $i = O(1)$. More precisely, if f is any constant function, then $f = O(1)$. But in this False Theorem, i is not constant but ranges over a set of values 0,1,…,n that depends on n.

And anyway, we should not be adding $O(1)$'s as though they were numbers. We never even defined what $O(g)$ means by itself; it should only be used in the context "$f = O(g)$" to describe a relation between functions f and g.

[3]Since $\Theta(n^3)$ only implies that the running time, $T(n)$, is between cn^3 and dn^3 for constants $0 < c < d$, the time $T(2n)$ could regularly exceed $T(n)$ by a factor as large as $8d/c$. The factor is sure to be close to 8 for all large n only if $T(n) \sim n^3$.

7.4.3 Lower Bound Blunder

Sometimes people incorrectly use Big Oh in the context of a lower bound. For example, they might say, "The running time, $T(n)$, is at least $O(n^2)$," when they probably mean something like "$O(T(n)) = n^2$," or more properly, "$n^2 = O(T(n))$."

7.4.4 Equality Blunder

The notation $f = O(g)$ is too firmly entrenched to avoid, but the use of "=" is really regrettable. For example, if $f = O(g)$, it seems quite reasonable to write $O(g) = f$. But doing so might tempt us to the following blunder: because $2n = O(n)$, we can say $O(n) = 2n$. But $n = O(n)$, so we conclude that $n = O(n) = 2n$, and therefore $n = 2n$. To avoid such nonsense, we will never write "$O(f) = g$."