Prereq Review #5: Geometric and Harmonic Series

Ian Tullis, CS161 Summer 2022

A Ubiquitous Series

The following series has a habit of turning up in CS161:

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} ...$$

Written more formally as an infinite summation, it is:

$$\sum_{i=0}^{\infty} \frac{1}{2^i}$$

We have been using the fact that no matter how many terms this series goes on for, the total is bounded above by 2. How do we know this?

Here is one way of justifying it. Let X be the sum, whatever it ends up evaluating to:

$$X = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} ...$$

Now subtract 1 from both sides:

$$X - 1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} ...$$

and multiply both sides by 2:

$$2(X - 1) = 1 + \frac{1}{2} + \frac{1}{4} + ...$$

But now the right side of the expression is the same as the original right side of the expression. So the two left sides must be equal as well:

$$\begin{aligned} X &= 2(X - 1) \\ X &= 2X - 2 \\ X &= 2 \end{aligned}$$

There is one important subtlety in this explanation: we implicitly assumed that the series actually did converge to a value, rather than diverging to ∞ – that is, continuing to grow so much that any bound that we specified would eventually be broken through. But the fact that it converges is not super hard to believe – notice that each new term gets us exactly half of the remaining distance toward a total of 2. It seems clear that this can never cause us to *exceed* 2.

This is a *geometric series*. The idea above works for any such series of the form $1 + \frac{1}{k} + \frac{1}{k^2} + ...$, which – as long as $k > 1$ – converges to $\frac{k}{k-1}$. (We had $k = 2$, so we got $\frac{2}{1} = 2$.)

Geometric Distributions

One place the series arises is when answering a question like: Suppose we flip a fair coin until we see our first "heads". What is the expected[1] number of flips we will make?

We can notice that we will always make at least one flip. Half of the time, that first flip will come up heads and we will be done. The other half of the time, we are essentially right back where we started – this process has no memory of how many times we have already tried and failed. So we can write the following expression for the expectation:

$$\mathbf{E}[X] = 1 + \frac{1}{2}\mathbf{E}[X]$$

The way to interpret this is as follows. What is the expected number of flips from the outset? We have to make one flip to begin with (hence the 1). Then, half the time, we are done and don't need to make any more flips, so there is an implicit $+\frac{1}{2}(0)$ term that we left out. The other half of the time, we are back at exactly an instance of the original problem, with the same expected number of flips needed to finish.

If we solve the above expression, we find that $\mathbf{E}[X] = 2$. (As in the example on the previous page, this kind of recursive setup only works if $\mathbf{E}[X]$ is actually finite.)

What if we had approached the problem a different way? We see that we take 1 flip half of the time, and 2 flips $\frac{1}{4}$ of the time, and 3 flips $\frac{1}{8}$ of the time, and so on. Then, using the definition of expectation, we can write:

$$\mathbf{E}[X] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + ...$$

or, more formally,

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} i \cdot \frac{1}{2^i}$$

We can type `sum from i=1 to infinity of i * 1/(2 ∧ i)` into Wolfram Alpha and see that it indeed converges to 2. But it would have been much trickier to actually solve the summation.[2]

Still another approach involves indicator random variables. Let $I_1, I_2, I_3 ...$ be indicators for the events that we need a first flip, a second flip, a third flip, etc., respectively. Then the total number of flips needed is

$$I_1 + I_2 + I_3 + ...$$

[1] If the concept of "expectation" feels rusty, you may wish to check out Prereq Review #3.

[2] There are various ways to do this, one of which even involves calculus. See `https://math.stackexchange.com/questions/129302/finite-sum-sum-i-1n-frac-i-2i` for some examples.

and the expected number of flips is therefore

$$\mathbf{E}[I_1 + I_2 + I_3 + ...]$$

By linearity of expectation, we can turn this expectation of a sum into a sum of expectations. Linearity of expectation works even though the variables $I_1, ...$ are not independent.[3] Then we have

$$\mathbf{E}[I_1] + \mathbf{E}[I_2] + \mathbf{E}[I_3] + ...$$

The expectation of an indicator random variable is the probability of the underlying event. We established that the probabilities of making it to the first, second, third, etc. flips are $1, \frac{1}{2}, \frac{1}{4}, ...$, so this sum becomes the familiar geometric series $1 + \frac{1}{2} + \frac{1}{4} + ... = 2$.

The number of trials that we need up to and including the first success is a random variable with a *geometric distribution*, and the **expected** number of trials is always $\frac{1}{p}$, where p is the probability of success. (In the above cases, we had $p = \frac{1}{2}$, but the idea generalizes to any $0 < p \leq 1$.)

Harmonic Series

What about this similar-looking series that also comes up sometimes in the course?

$$\sum_{i=1}^{\infty} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + ...$$

This is called the *harmonic series*, and it doesn't seem to grow very fast, but surprisingly, it can't be bounded! Here is one way to see that. Try grouping up the terms as follows:

$$\begin{aligned} &1 \\ &+\frac{1}{2} \\ &+\frac{1}{3}+\frac{1}{4} \\ &+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}... \end{aligned}$$

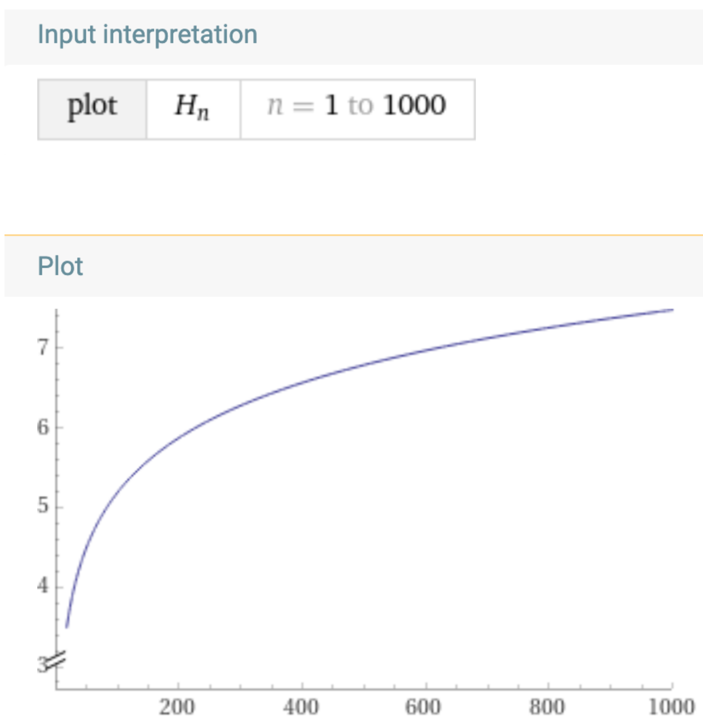
[3] For example, knowing that $I_2 = 0$ tells us for sure that $I_3 = 0$, since if we didn't need a second flip, we must have been done on our first flip and we don't need any flips whatsoever.

and then compare with the terms from our earlier geometric series:

$$\begin{aligned}
1 &\geq \frac{1}{2} \\
\frac{1}{2} &\geq \frac{1}{2} \\
\frac{1}{3}+\frac{1}{4} &\geq \frac{1}{4}+\frac{1}{4}=\frac{1}{2} \\
\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8} &\geq \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}\ldots
\end{aligned}$$

We can keep grouping terms like this to get as many $\frac{1}{2}$s as we want, so the sum of the series becomes arbitrarily big as we add more terms.

The notation H_n represents the sum of the first n values of the harmonic series. The above analysis informally implies that the series grows logarithmically – it takes about twice as many values each time to get each new boost of $\frac{1}{2}$. H_n is in fact very slow-growing, as this plot suggests:



And indeed, $H_n = O(\log n)$. This can be proven using a similar term-grouping argument:

$$
\begin{aligned}
1 &\leq 1 \\
+\frac{1}{2} + \frac{1}{3} \leq \frac{1}{2} + \frac{1}{2} &= 1 \\
+\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \leq \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} &= 1 \\
&\ldots
\end{aligned}
$$

There are $\lfloor \log_2 n \rfloor + 1$ such groups, each bounded above by 1, so $H_n \leq \lfloor \log_2 n \rfloor + 1$. The $O(\log n)$ bound follows naturally from this.