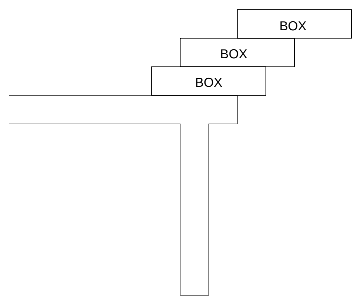
Lecture 17 - Harmonic Numbers and Approximating Sums

6.042 - April 17, 2003

1 Box Stacking

How far can a stack of identical boxes overhang the edge of a table without falling over? Can the top box in the stack lie *beyond* the edge of the table?



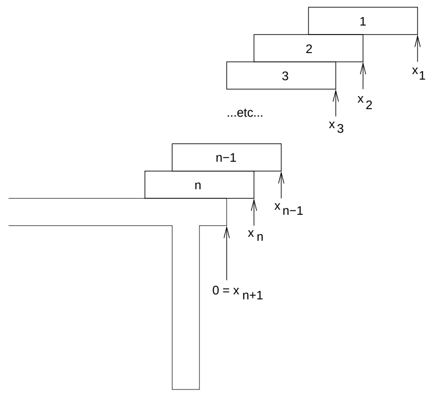
Here is the key intuition needed to answer this question: the center-of-mass of the top k boxes should be directly above the rightmost edge of the $(k+1)$-st box. If the center-of-mass were to the left of that point, we could build a stack that extended farther. If the center-of-mass were to the right, the top k boxes would topple. To develop this simple idea into a solution to the problem, we need a fact from physics.

Fact 1 *If one group of objects has mass m_1 and center-of-mass z_1 and another group has mass m_2 and center-of-mass z_2, then the center-of-mass for the two groups together is:*

$$z = \frac{z_1 m_1 + z_2 m_2}{m_1 + m_2}$$

We also need a coordinate system to describe the positions of boxes. As it turns out, only the horizontal dimension is relevant in our analysis. Let the rightmost edge of the table be the origin, and let box-lengths be the unit of measure. Thus, for example, position 2 lies two box-lengths past the edge of the table.

Now suppose that we want to build a stack of n boxes with maximum overhang. Number the boxes $1, \ldots, n$, starting with the topmost one. Let x_k be the position of the rightmost edge of the k-th box, and let z_k be the center-of-mass of the top k boxes. Also, define $x_{n+1} = 0$; in effect, we are regarding the table as an $(n+1)$-st book.



Putting together these ideas and notations, we can solve the box-stacking problem as follows:

$$\begin{aligned} x_{k+1} &= z_k \\ &= \frac{z_{k-1}\cdot(k-1)+\left(x_k-\frac{1}{2}\right)\cdot 1}{(k-1)+1} \\ &= \frac{x_k k - \frac{1}{2}}{k} \\ &= x_k - \frac{1}{2k} \end{aligned}$$

The first equation formalizes the key intuition described above: the center-of-mass of the top k boxes (at position z_k) should lie directly above the rightmost edge of the $(k+1)$-st box (at position x_{k+1}). The second step uses the physics fact. Specifically, we are expressing the center-of-mass of the top k boxes (z_k) in terms of the center-of-mass and total mass of the top $k-1$ boxes (z_{k-1} and $k-1$, respectively) and the center-of-mass and mass of the k-th box ($x_k - \frac{1}{2}$ and 1, respectively). The third step again uses the key intuition; that is, we should have $x_k = z_{k-1}$. The final step uses only simplification.

This calculation shows that the k-th box should overhang the $(k+1)$-st box by distance $1/(2k)$. Thus, the rightmost edge of the top box ends up at position:

$$\begin{aligned} x_1 &= \frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{2n} \\ &= \frac{1}{2}\sum_{i=1}^{n}\frac{1}{i} \\ &= \frac{1}{2}H_n \end{aligned}$$

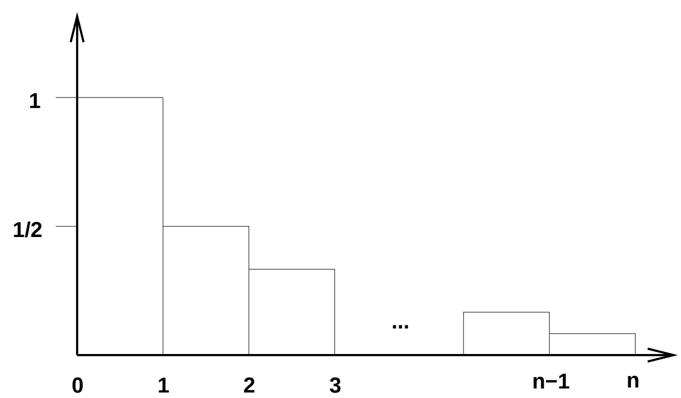
In the last line above, we introduce the notation H_n to stand for the sum $\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots\frac{1}{n}$.

This is called the *n-th harmonic number*. Thus, if we stack n boxes properly, the top one overhangs the table edge by $\frac{1}{2}H_n$ box lengths.

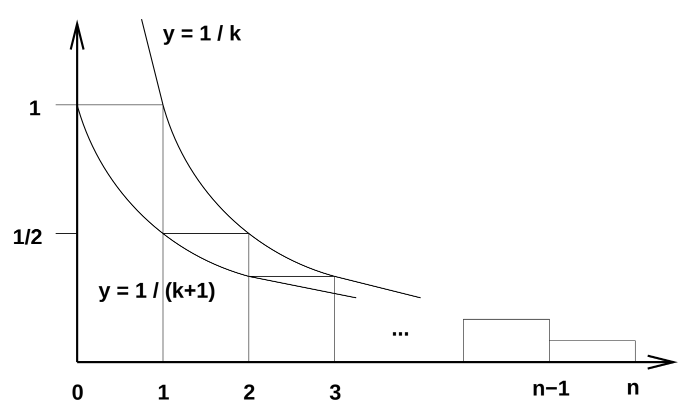
1.1 Harmonic Numbers

We've now expressed the solution to the box-stacking problem in terms of harmonic numbers. This naturally raises a question: how big is H_n, the n-th harmonic number? Harmonic numbers come up in many problems, so the answer is worth knowing in general. Furthermore, the trick we use to answer this question has much wider applicability.

Here's the trick. We create a bar graph where the k-th bar has height $1/k$ and width 1. Then the n-th harmonic number is simply the area under the first n bars:



Our goal now is to estimate the area under the bar graph. To that end, we construct two curves, one just above the bars ($y = 1/k$) and one just below ($y = 1/(k+1)$). The areas under these curves, which we can find by integration, are upper and lower bounds on the area under the bar graph, and thus are upper and lower bounds on H_n.



Thus, for the upper bound, we have:

$$\begin{aligned}\sum_{k=1}^{n} \frac{1}{k} &< 1 + \int_1^n \frac{1}{k}\, dk \\ &= 1 + \ln n\end{aligned}$$

Here, we used the fact that the first bar has area 1 and began integrating at $k = 1$. If we had begun integrating at $k = 0$, then our upper bound would be infinite. (Valid, but not very useful!) For the lower bound, we have:

$$\begin{aligned}\sum_{k=1}^{n} \frac{1}{k} &> \int_0^n \frac{1}{k+1}\, dk \\ &= \ln(n+1)\end{aligned}$$

Putting these inequalities together, we find that H_n is very close to $\ln n$. Therefore, H_n slowly tends to infinity as n gets large.

In terms of box-stacking, this means that we can achieve *arbitrarily large* overhang, provided we have sufficiently many boxes. Only four boxes are required to suspend one completely off the table, since $\frac{1}{2}H_4 = 1.0416\ldots$. Thirty-one boxes just barely suffice to build a stack that extends two box-lengths past the end of the table, since $\frac{1}{2}H_{31} = 2.0136\ldots$. A stack of CD cases extending all the way to the moon could overhang by about a dozen case-lengths. Of course, by then, many other physical assumptions would break down!

2 Arranged Marriages

The time has come for Princess Harmony to choose a husband. She has n suitors who visit her one at a time in a random order. She is able to rank each suitor relative those she has already seen, but she has no way to rank suitors on an absolute scale; the next candidate could always be better— or worse!— than all who came before. After appraising a suitor, Princess Harmony either accepts him as her future prince or else caustically spurns him as a worthless whelp. A suitor so cruelly spurned then departs, never to return, however much she might later regret her unkind words. (Princess Harmony was named after the harmonic numbers, not because she has a particularly pleasant disposition.) The princess is allowed to meet the second suitor only after deciding on the first and, in general, she meets the $(i+1)$-st suitor only after disposing of the i-th. She must marry the last suitor, if she rejects all the others.

The Princess wants to maximize the probability that she picks the *best* suitor. Anything else is defeat; to her refined sensibilities, settling for second best is as bad as picking the worst. So what should she to do with a promising suitor met early on? Accepting him as her prince might be a mistake; there will *probably* be a better suitor later. On the other hand, if the princess screeches, "Don't let the door smack yer butt on the way out!", she *might* make the awkward discovery that all subsequent suitors are worse.

As the number of suitors n grows very large, a reasonable guess is that the Princess has a vanishingly small chance of picking the very best one. Surprisingly, this is not true! Suppose that the princess employs the following strategy:

1. Reject the first r suitors.
2. Accept the next suitor who is better than all who came before.
3. If there is no such suitor, marry the last one.

The key question is how to choose the magic number r. For example, suppose that Princess Harmony chooses $r = n/2$. Then she wins if the second-best candidate is among the first $n/2$ suitors and the very best candidate is among the last $n/2$ suitors. When n is large, these events each occur with probabilty $\frac{1}{2}$ and are nearly independent. Therefore, her probability of picking the best suitor is at least $\frac{1}{4}$, no matter how many suitors there are!

Let's work out the best value for r, the value that maximizes the probability that the princess gets her man. Number the suitors from 1 to n in the order that they arrive. Let $F(k)$ be the number of the best suitor among the first k. Note that F is a nondecreasing function. In particular, $F(k + h) = F(k)$ only if the best of the first $k + h$ suitors is also among the first k. (This happens with probability $k/(k + h)$, a fact we'll use momentarily.) Otherwise, $F(k + h) > F(k)$.

Now we need to consider various cases to determine the fate of our sharp-tongued princess:

- With probability r/n, the best suitor is among the first r. If this happens, Princess Harmony rejects him and loses.
- With probability $1/n$, the best suitor is number $r + 1$. If this happens, the princess picks him and thus wins with probability 1.
- With probability $1/n$, the best suitor is number $r + 2$. Now the princess picks him only if $F(r + 1) = F(r)$, which happens with probability $r/(r + 1)$. Otherwise, if $F(r + 1) > F(r)$, then suitor $r + 1$ is better than everyone who came before, and so Princess Harmony erroneously picks him instead and loses.
- With probability $1/n$, the best suitor is number $r+3$. Now the princess picks him only if $F(r + 2) = F(r)$, which happens with probability $r/(r + 2)$. Otherwise, she loses.

Continuing to reason in this way, we find that the best suitor is in position $r + k$ with probability $1/n$, and the princess picks him with probability $r/(r + k - 1)$. Since these are disjoint events, her probability of success is the sum of the event probabilities:

$$
\begin{aligned}
\Pr\{\text{picks best}\} &= \frac{1}{n}\left(1 + \frac{r}{r+1} + \frac{r}{r+2} + \cdots + \frac{r}{n-1}\right) \\
&= \frac{r}{n}\left(\frac{1}{r} + \frac{1}{r+1} + \frac{1}{r+2} + \cdots + \frac{1}{n-1}\right) \\
&= \frac{r}{n} \cdot (H_{n-1} - H_{r-1}) \\
&\approx \frac{r}{n} \cdot (\ln n - \ln r)
\end{aligned}
$$

In the first step, we are implicitly using the fact (which we shall not prove) that that best suitor being in position k is independent of the event that $F(r+k-1) = F(r)$. Otherwise, the first steps use only simplification. The final step uses the approximation $H_n \approx \ln n$, which we derived in the preceding section by integration.

Now we have a formula for the probability that the Princess wins as a function of r and n. To maximize this probability, we set the derivative with respect to r equal to zero:

$$
\begin{aligned}
\frac{d}{dr}\left(\frac{r}{n} \cdot (\ln n - \ln r)\right) &= 0 \\
\frac{r}{n} \cdot \left(-\frac{1}{r}\right) + \frac{1}{n} \cdot (\ln n - \ln r) &= 0
\end{aligned}
$$

Simplifying this equation, we find $r = n/e$. Thus, the princess should reject a $1/e$ fraction of suitors and then pick the next one who is better than all who came before. If she follows this strategy, she wins with probability:

$$
\begin{aligned}
\frac{n/e}{n} \cdot (\ln n - \ln(n/e)) &= \frac{1}{e} \\
&= 0.36788\ldots
\end{aligned}
$$

Amazingly, the princess picks the best possible prince with probability greater than $\frac{1}{3}$, regardless of the number of suitors!