Panini Sticker Books

Dr. Maria Keet
University of Cape Town, South Africa

ICPC Training Session, Aug 2, 2014

The 2104 World Cup has just finished, but you're still enjoying the afterglow—and aftermath. Your grannies have given you a Brazil World Cup 2014 Panini sticker book to complete, just like they did for the 2010 World Cup in South Africa. With a sticker for each player in a team, there are 638 stickers to collect in total. Unlike in 2010, however, when many of your friends were trying to complete the sticker book so you could do swapsies with your friends to trade duplicates for ones you didn't already have, now you have to complete it all by yourself. Each packet has 5 stickers, costing you R10 per packet.

Being the cash-strapped student you are, you start wondering: how many packets will I have to buy to complete the sticker book, without doing swapsies with anybody, and how much is that going to cost me? And, more generally: for any world cup—after all, cricket and rugby are important sports too!—how many packets would one have to buy at what total cost to complete a world cup sticker book?

Your task is to compute the total cost C to complete a World Cup sticker book that requires N stickers, and each pack of five stickers cost P (e.g., some stickers are glossier than others, and you may want to reuse the solution for other countries with different currencies). For the Soccer World Cup and its 638 stickers, this amounts to having to buy about 4488 stickers, and with each packet at R10 each makes a total sum of R 8977 (rounded up) to complete the sticker book.

You may assume that there is an equal number of each sticker and that they are randomly distributed among the packets. For the cost, this is simplified to the total amount.

$5 \leq N \leq 1000$ and $0 < P \leq 200$, with N and P integers, and round up the cost to the nearest integer.

The input will be a file with two numbers on each line denoting N and P, respectively, and your output should give a single number on each line denoting the total cost C. The input ends with a `-1`.

Sample input

```
638 10
50 25
400 35
-1
```

Sample output

```
8977
1120
18389
```

WARNING: the outline of the solution is on the next page, so do NOT scroll down further if you're still working on the solution!

Solution

The problem itself is essentially a variation on what is called the *Coupon Collectors Problem*[1]. The storyline is based on Simon Whitehouse's blog post[2], though modified here a bit to make it more interesting as a computational problem and coding (ok, it still can be done in spreadsheet software, but that's not the point, and actually, the computation is much more interesting—and, in a way, 'more correct'...).

First, an alternative way to formulate the problem is "Given N stickers, how many stickers do you expect you need to get (technically: with replacement) before having gotten each sticker at least once?", and then add the bit on packets and cost to get the requested answer, i.e.,

$$\frac{total\_no\_of\_stickers}{5} * P \tag{1}$$

To arrive at the $total\_no\_of\_stickers$, you need to know some basics of probability first. The first sticker is guaranteed to be a new one, i.e., has a chance of 1 of being new. The second sticker has a chance of $N-1$ of being a new one, and the third one a chance of being $N-2$ of being one you didn't already have, and so on until the last one.

Looking at this from a probabilities viewpoint, we have a probability p of an event happening, and thus the expected number of times to get the desired outcome is $\frac{1}{p}$. The probability of getting a new sticker is the desired event. Then it amounts to adding up the number of stickers you need to have (N), and that sequence of decreasing probabilities. This makes for a summation:

$$\frac{N}{N} + \frac{N}{N-1} + \frac{N}{N-2} + \ldots + \frac{N}{N-(N-1)} \tag{2}$$

This is what is called a 'harmonic series'. From there, and with a bunch of maths, one can end up with, *roughly*:

$$total\_no\_of\_stickers = N * (ln(N) + \gamma) \tag{3}$$

with ln being the natural log and γ the Euler number (roughly: 0.57)—according to the blog post and similarly with the wikipedia entry, mentioned above. Good for you if you recognised that; if you didn't but want to know, then have a look at a more detailed description of the coupon collectors problem.

Filling in the formula, then for $N = 638$, we'd have $total\_no\_of\_stickers = 638 * (ln(638) + 0.57)$, and with $P = 10$ and Eq. 3, then $\frac{638*(ln(638)+0.57)}{5} * 10$ is 8969 rounded up. BUT. The Euler number has more significant digits, and programming Eq. 1 with Eq. 2 instead, along the line of (quick 'n dirty python code cf. the C/C++ or Java required for the regionals)

```
def panini(n,p):
    tns = 0
    for i in range(1,n):
        tns = tns + n/i
    return math.ceil(tns/5*p)
```

gives 8977. Likewise, there are small differences in the other values: the smaller N is, the smaller the difference, for the impreciseness is essentially due to how imprecise you have the Euler number (if you had it included at all, which the original blog post of Whitehouse does not).

In the original version of this file for the Aug 2 training session, I had a "Whether you make the jump to the ln (Eq. 3) or decide to work with Eq. 2 when implementing the solution, is for you to sort out.". This now has been changed to the latter, also for the sample output file numbers. From a computational viewpoint, that's a lot easier problem solving (and more precise, if you will).

[1] informal description: `http://en.wikipedia.org/wiki/Coupon_collector%27s_problem`

[2] `http://siwhitehouse.co.uk/blog/2010/04/25/panini-football-stickers-and-the-coupon-collector-problem/`