Code Jam 2009 - Round 1A

Analysis: Collecting Cards

This problems requires some basic knowledge of probability and combinatorics. We want to calculate the expected number of packs we need to buy to obtain all C different cards.

Let's denote by E(*x*) the expected number of packs we would need to buy if we started with *x* different cards (it doesn't matter what those cards are). The answer to the problem is the value for E(0). We also know that E(C) = 0, because if we already have C different cards we don't need to buy any additional packs.

We can derive useful equations for other values of E(*x*) by thinking about all the possible outcomes after buying one additional pack. Let's call T(*x*,*y*) the probability of ending up with *y* different cards after opening a new pack. Then we have the following equation for E(*x*):

$$E(x) = 1 + \sum_{y=x}^{\min(C, x+N)} T(x,y) \cdot E(y)$$

We need to buy at least one new pack, so that's where the 1 comes from. The expected number of packs we need to buy after that depends on how many new cards we get. If we end up with *y* different cards we need to add the expected number of packs to reach C starting from *y*, which we called E(*y*), multiplied by the probability of this particular alternative given by T(*x*, *y*).

All these equations put together form a system of linear equations with an upper triangular matrix which can be solved using the standard back substitution method.

But we still haven't said how to calculate the entries of the matrix T (that is, the values of T(*x*, *y*) for all different *x* and *y*). We'll calculate this with the help of binomial coefficients: the number of different possible packs is $\binom{C}{N}$. To end up with *y* different cards, we need to choose y-x out of C-x possible new cards, and the remaining N-(y-x) have to be chosen from the x cards we already have. The answer then is:

$$T(x,y) = \frac{\binom{C-x}{y-x} \cdot \binom{x}{N-(y-x)}}{\binom{C}{N}}$$

(For those with some knowledge of probability, this is called the hypergeometric distribution).

The special case, where N = 1, is the well known coupon collector's problem.