



© 1997–2009, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

December 2005

Regulars



## Outer space: A collector's piece

by John D. Barrow



Last weekend, hidden between books in the back of my bookcase, I came across two sets of cards that I collected as a young child. Each of them contained fifty high quality colour pictures of classic motor cars with a rather detailed description of their origins, design and mechanical specification on the reverse. Collecting sets of cards was all the rage. There were collections of wartime aircraft, animals, flowers, ships, and sportsmen since these collections all seemed to be aimed at boys to be amassed from buying lots of packets of bubble gum, breakfast cereals, or packets of tea. Of the sports cards, just like today's Panini "stickers", the favoured game was football (in the US it was baseball) and I always had my suspicions about the assumption that all the player's cards were produced in equal numbers. Somehow everyone seemed to be trying to get the

## Outer space: A collector's piece

final "Bobby Charlton" card that was needed to complete the set. All the other cards could be acquired by swapping duplicates with your friends but everyone lacked this essential one.

It is a relief to discover that even my own children engaged in similar acquisitive practices. The things collected might change but the basic idea was the same.

So what has mathematics got to do with it? The interesting question is to ask how many cards we should expect to have to buy in order to complete the set, if we assume that each of them is produced in equal numbers and so has an equal chance of being found in the next packet that you open. The motor car sets I came across each contained 50 cards. The first card I get will always be a new one, but what about the second? There is a 49/50 chance that I haven't already got it. Next time it will be a 48/50 chance and so on. After you have acquired 40 different cards there will be a 10/50 chance of the next one being one you haven't already got. So, on the average, you will have to buy another 50/10, or 5 more cards to have a better than even chance of getting another new one that you need for the set. Therefore, the total number of cards you will need to buy on average to get the whole set of 50 will be the sum of 50 terms:

$$50/50 + 50/49 + 50/48 + \dots + 50/3 + 50/2 + 50/1,$$

where the first term is the certain case of the first card you get and each successive term tells you how many extra cards you need to buy to get the 2nd, 3rd, and so on, members of the set of 50 cards.

As there can be collections with all sorts of different numbers of cards in them, let's consider acquiring a set with any number of cards in it, and let's call this number  $N$ . Then the same logic tells us that on the average we will have to buy a total of

$$N/N + N/(N - 1) + N/(N - 2) + \dots + N/3 + N/2 + N/1.$$

Taking out the common factor  $N$  in the numerators of each term, this is just

$$N(1 + 1/2 + 1/3 + \dots + 1/N).$$

The sum of terms in the brackets is just the famous harmonic series that we have met in this column before (see [outer space](#) in issue 32). When  $N$  becomes large it is well approximated by  $0.577 + \ln(N)$  where  $\ln(N)$  is the natural logarithm of  $N$ . So as  $N$  gets realistically large we see that the number of cards that we need to buy on the average to complete our set is about

$$N \times [0.577 + \ln(N)].$$

For my sets of 50 Player's Motor Car cards the answer is 224.5, and I should expect to have had to buy 225 cards to make up my set of 50.

Incidentally, our calculation shows how much harder it gets to complete the second half of the collection than the first half. The number of cards that you need to buy in order to collect  $N/2$  cards for half a set is

$$N/N + N/(N - 1) + N/(N - 2) + \dots + N/(N/2 + 1),$$

which is the difference between  $N$  times the harmonic series summed to  $N$  and summed to  $N/2$  terms, so the expected number of cards needed for half a set is

$$N \times [\ln(N) + 0.577 - \ln(N/2) - 0.577] = N \times \ln(2) = 0.7N.$$

I wonder if the original manufacturers performed such calculations. They should have, because they enable you to work out the maximum profit you could expect to gain in the long run from marketing a set of cards of a particular size. But it is only likely to be a maximum because collectors will trade cards and be able to acquire new cards by swapping rather than buying new ones. So, what impact can friends make?

## Outer space: A collector's piece

Suppose that you have  $F$  friends and you all pool cards in order to build up  $F + 1$  sets so that you have one each. How many cards would you need to do this? On average, when the number of cards  $N$  is large, and you share cards, the answer approaches

$$N \times [\ln(N) + F \times \ln(\ln(N)) + 0.577].$$

On the other hand, if you had each collected a set without swapping you would have needed about

$$(F + 1) \times N \times [\ln(N) + 0.577]$$

cards to complete  $F + 1$  separate sets. For  $N = 50$  the number of card purchases saved would be  $156F$ . Even with  $F = 1$  this is a considerable economy.

If you know a little statistics, you might like to show that the standard deviation on the

$$N \times [0.577 + \ln(N)]$$

result is close to  $1.3N$ . This is quite significant in practice because it means that you have a 66% chance of needing to collect  $1.3N$  more or less than the average. For our 50-card set this uncertainty in the expected number of purchases is 65.

There was a story a few years ago that a consortium were targeting suitable national lotteries by calculating the average number of tickets that needed to be bought in order to have a good chance of collecting all the possible numbers and hence including the winning one. They neglected to include the likely variance away from the average result and were very lucky to find that they did have a winning ticket among the millions they had bought.

If the probability of each card appearing is not the same then the problem becomes harder but can still be solved. In that case it is more like the problem of coin collecting where you try to collect a coin with each available year date on. You don't know that equal numbers were minted in each year (almost certainly they weren't) or how many may have been withdrawn later, so you can't rely on their being an equal chance of collecting an 1840 penny or an 1890 one. But if you do find a 1933 English penny (of which only 7 were made and 6 are accounted for) then be sure to let me know.

---

Hint: to work out the standard deviation, you'll need to know that

$$1 + 1/4 + 1/9 + \dots = \sum_{N=1}^{\infty} 1/N^2 = \pi^2/6$$

.

---

Did you manage to answer the puzzle posed in [Outer space: series](#)? If not, you can [find the answer here](#)!

---



*Plus* is part of the family of activities in the Millennium Mathematics Project, which also includes the [NRICH](#) and [MOTIVATE](#) sites.