
March 2006
Regulars

A collector's piece: solution

In last issue's *outer space* we worked out how many pruchases you would expect to have to make in order to obtain every one of a set of N distinct cards. The puzzle was to show that the standard deviation of this result is close to $1.3N$ for large N. To work out the expectation, we considered N random variables X_0, X_1, etc, up to X_{N-1}. Each random variable X_i counts how many purchases you have to make to get a new card if you already have i distinct cards. The overall expectation is then just the sum of the individual expectations of the N variables.

You can use the same technique to work out the standard deviation (which is of course the square root of the variance): since the N random variables are independent of each other, the overall variance is simply the sum of the individual values.

For each individual variable X_i, the probability that you have to make j purchases to get a new card is

$$\frac{N-i}{N} \times \left(1 - \frac{N-i}{N}\right)^{j-1}.$$

So each individual X_i has the well−known *geometric probability distribution* with parameter

$$p = \frac{N-i}{N}.$$

This distribution has variance

$$\frac{1-p}{p^2} = \frac{iN}{(N-i)^2}.$$

Now if X is the sum of the N random variables X_0, X_1, ... , X_{N-1}, then the variance of X is

$$variance(X) = \sum_{i=0}^{N-1} \frac{iN}{(N-i)^2} = N \sum_{i=0}^{N-1} \frac{i}{(N-i)^2}.$$

A little thought shows that this sum can be expressed as

$$N\sum_{i=1}^{N}\frac{N-i}{i^2}=N\sum_{i=1}^{N}\frac{N}{i^2}-N\sum_{i=1}^{N}\frac{i}{i^2},$$

which is equal to

$$N^2\sum_{i=1}^{N}\frac{1}{i^2}-N\sum_{i=1}^{N}\frac{1}{i}.$$

Dividing through by N^2 gives

$$\frac{variance(X)}{N^2}=\sum_{i=1}^{N}\frac{1}{i^2}-1/N\sum_{i=1}^{N}\frac{1}{i}.$$

As N gets large, the first term tends to $\pi^2/6$, as was stated in the hint, while the second term tends to zero. This proves that the standard deviation for large N is close to $\sqrt{\pi^2 N^2/6}$, which is approximately *1.3 N* QED.

Back to Outer space

Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.