How Many Singles, Doubles, Triples, Etc., Should The Coupon Collector Expect?

Doron ZEILBERGER [1]

There are m equi-probable baseball cards placed at random in chewing gums. It is well known and easy to see that a collector should expect to buy $m(1+1/2+1/3+\ldots+1/m)$ gums before acquiring all the kinds of cards. At the end, he would have some singles, some doubles, some triples, etc. Let $A(m,i)$ be the expected number of kinds of cards of which he has exactly i copies of. Here I give a short proof of:

Formula (Foata-Han-Lass[1]): $\sum_{i=1}^{\infty} A(m,i)t^i = t-1+m!/\prod_{j=2}^{m}(j-t)$.

Proof: Let's number the (kinds of) cards, in order of first arrival by $1,2,\ldots m$. The purchased cards define a word given by the regular expression $11^*2\{1,2\}^*3\{1,2,3\}^*4\ldots\{1,2,\ldots,m-1\}^*m$, whose (probability) generating function is

$$f(x_1,\ldots,x_m) = \frac{x_m}{m}\prod_{j=1}^{m-1}\frac{x_j}{m-(x_1+\ldots+x_j)} \quad .$$

Let a *marked word* be a pair $[w,i]$ where w is a word that is an instance of the above regular expression, and $1 \le i \le m$. By the familiar trick of computing expectations by changing the order of summation, it follows that the left side of the Foata-Han-Lass formula is the sum of the weights of all eligible marked words, where $weight([w,i]) := (1/m)^{|w|}$ times t raised to the power [the number of times the letter i occurs in w]. For example, if $m=3$ and w=1111211213, then $weight([w,1]) = (1/3)^{10}t^7$, $weight([w,2]) = (1/3)^{10}t^2$, $weight([w,3]) = (1/3)^{10}t$. Hence $\sum_{i=1}^{\infty} A(m,i)t^i = m!\sum_{i=1}^{m} f_i$ (the factor of $m!$ is to account for all possible orderings), where f_i is f with all the x's replaced by 1, except for x_i that is replaced by t. But

$$f_{m-i} = \frac{1}{m}\prod_{j=1}^{m-i-1}\frac{1}{m-j}\cdot\frac{t}{i+1-t}\cdot\prod_{j=m-i+1}^{m-1}\frac{1}{m-j+1-t} = \frac{i!t}{m!(2-t)(3-t)\cdots(i+1-t)} \quad ,$$

when $i>0$ and $f_m = t/m!$. Hence

$$\sum_{i=1}^{\infty} A(m,i)t^i = t + t\sum_{i=1}^{m-1}\frac{i!}{(2-t)(3-t)\cdots(i+1-t)} = t + \frac{m!}{(2-t)(3-t)\ldots(m-t)} - 1 \quad \Box.$$

Reference

1. D. Foata, G.-N. Han, et B. Lass, *Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes*, preprint, available from Foata's website.

[1] Department of Mathematics, Temple University, Philadelphia, PA 19122, USA. `zeilberg@math.temple.edu` `http://www.math.temple.edu/~zeilberg/` . July 5, 2001 Supported in part by the NSF.