Kaan Aksoy — Feb 24, 2020

Collecting Coupons

Problem

Coupons in cereal boxes are numbered 1 to 5, and a set of one of each is required for a prize. With one coupon per box, how many boxes on the average are required to make a complete set?

Solution

Let X_1, X_2, $\ldots$, X_5 be random variables representing the number of boxes opened until receiving the 1^{st}, 2^{nd}, $\ldots$, and 5^{th} unique coupon, respectively, where the number of boxes is counted since the last box containing a unique coupon was opened. For example, X_3 represents the number of boxes opened after the 2^{nd} unique coupon was found up until the 3^{3rd} unique coupon was found.

Each X_i is a geometric random variable associated with a different probability of success. For example, X_1 is associated with $p_1 = 1$ because the 1^{st} box will obviously contain a unique coupon. X_2 is associated with $p_2 = \frac{4}{5}$ because only 4 of the 5 coupons are unique at that point. Probabilities for the remaining X_i can be calculated in a similar fashion. For geometric random variables, we also know that $E[X_i] = \frac{1}{p_i}$.

To solve the problem, we can apply the *Linearity of Expectation* property as follows:

$$
\begin{aligned}
E\left[\sum_{i=1}^{5} X_i\right] &= \sum_{i=1}^{5} E[X_i] \\
&= \sum_{i=1}^{5} \frac{1}{p_i} \\
&= 5\left(\frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= \frac{137}{12} \approx 11.42
\end{aligned}
$$

Notes

When generalized to n coupons, the solution to this problem can be written as:

$$C(n) = n \sum_{i=1}^{n} \frac{1}{i} = nH_n$$

In the above, H_n represents the n^{th} *Harmonic Number*, which is the sum of the reciprocals of integers from 1 to n.