Coupon Collector's Problem: An Application to Lidl's Vegesaurios

Jorge Guerra Rodríguez, José Ignacio Díez Ruiz

1 Introduction

The coupon collector's problem is a classic issue in probability theory that most students encounter in advanced probability courses, much like the birthday paradox or the Monty Hall problem. Coupon collector's problem explores how many coupons one should expect to draw with replacement before having obtained each coupon at least once and complete the collection in a game such as a promotional event involving n different coupons.

A mini-Vegesaurio is a small plastic figure shaped like a hybrid between a dinosaur and a vegetable. Lidl's mini-Vegesaurios is a promotional campaign that started in Spain on September 25, 2023. In this campaign, for every 25 euros spent on purchasing groceries at Lidl's supermarkets, customers receive a surprise envelope containing one of the 16 different mini-Vegesaurios.

Mini-Vegesaurios are primarily designed for children to collect, but they also hold appeal for older grown individuals like us. The campaign's primary goal is to promote vegetable consumption. As an example, Ginger is a central character in this collection, and she is a combination of a carrot and a triceratops. On the mini-Vegesaurios website, you can find descriptions of each character along with the associated benefits of consuming vegetables such as carrots, and more for each of the mini-Vegesaurios.

The purpose of this document is to apply the well-known coupon collector's problem to determine the expected number of 25 euro purchases one would need to make at Lidl's stores in order to complete the collection and acquire all the mini-Vegesaurios. We expect the reader to have a basic understanding of probability and random variables. We are working on the assumption that each mini-Vegesaurio has an equal likelihood of being obtained in the surprise envelope. However, our experience suggests that this assumption might not be accurate. For instance, we have received multiple Dormilón mini-Vegesaurio figures but have not yet obtained Ginger, who is the principal character and logically might be more challenging to acquire. Despite these observations, for the sake of simplicity, we continue to maintain the assumption of equal probability for each mini-Vegesaurio.

2 Coupon Collector's Problem

Let T be the random variable of the number of draws needed to collect the $n = 16$ mini-Vegesaurios. Let t_i represent the random variable od the number of trials required to collect the i-th mini-Vegesaurio after obtaining $i-1$ other different mini-Vegesaurios. Hence, one shold note that $T = t_1 + \cdots + t_n$.

Observe that the probability of getting a new mini-Vegesaurio after obtaining $i-1$ other different mini-Vegesaurios is $p_i = [n-(i-1)]/n$. Hence, t_i is modelled as a geometric distribution of expectation $1/p_i$ and the expected number of surprise envelopes that one should open for completing the collection is computed as

$$
\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[t_1 + \cdots t_n] \\
&= \mathbb{E}[t_1] + \cdots + \mathbb{E}[t_n] \\
&= \frac{1}{p_1} + \cdots + \frac{1}{p_n} \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} \\
&= n \cdot \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\
&= n \cdot H_n,
\end{aligned}
$$

where $H_n = \sum_{i=1}^{n} 1/i$ is the n-th harmonic number.

The expectation for the mini-Vegesaurios problem is 54.09, and as using the asymptotics of the harmonic numbers one gets that

$$\mathbb{E}[T] = n\gamma + n\log(n) + 1/2 + O(1/n),$$

where γ is the Euler–Mascheroni constant [1].

Given that it is anticipated to require 54 purchases of 25 euros each to obtain the complete collection, one can expect to spend over 1350 euros in total to acquire all the mini-Vegesaurios. Considering the promotion's duration of 41 days, approximately 6 weeks, consumers aiming to collect them all would need to spend around 225 euros per week.

3 Simulations

Prior to the mathematical development, we performed simulations to find the solution to the problem. In this repository you can find the function to perform the simulation in R (`coupon_collector`), as well as a much faster version written in Rcpp (`coupon_collectorC`). In this repository, you can also find a function that calculates the analytical solution (`analytical_solution`).

Also, we wanted to see how the number of expected coupons to draw to get them all grows as a function of the total number of different coupons. In the following graph we can see this for between 10 and 40 different coupons.

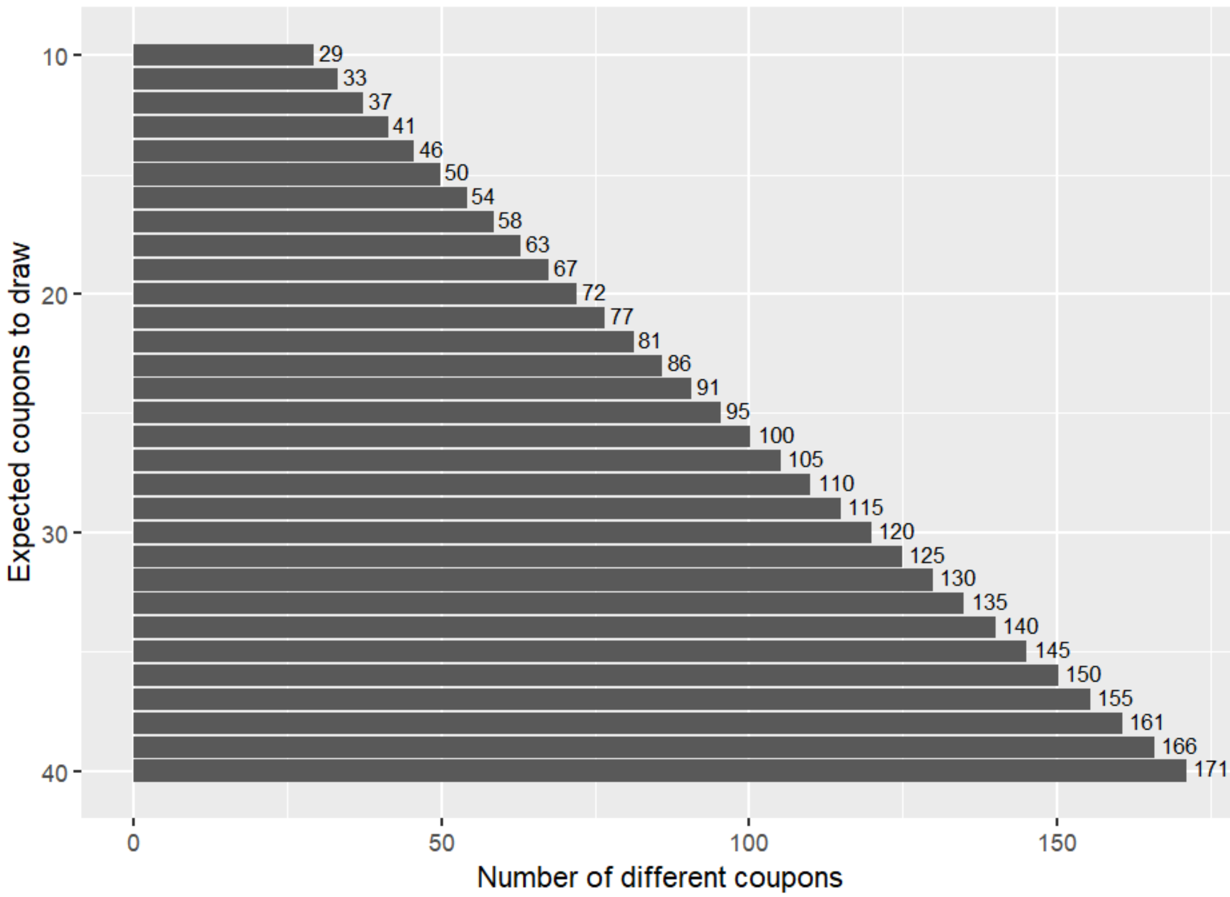

Figure 1: Plot of the number of expected coupons to draw in order to obtain all the coupons vs the total number of different coupons.

Looking at the plot we can say that, if, for example, the total number of mini-Vegesaurios were slightly more, say 20, we would have to open an average of 72 mini-Vegesaurios packs to get them all (compared to the 54 we have to open with 16). Increasing the total number of mini-Vegesaurios by 4 would increase the money to spend at Lidl to get them all from, on average, 1350 euros to 1800 euros, 450 euros more. This would make it even more difficult to get them all in the short period of the promotion.

4 Conclusion

Behind Lidl's mini vegesaurios there is a very interesting mathematical problem that makes us see that, for a normal family, getting the 16 in the time interval that the promotion lasts is almost impossible for a normal family. We think that this may be designed to encourage children (and not so children) to exchange mini-Vegesaruios with friends and family in order to complete the collection.

This can be extended to any type of collection, as we have seen with the Coupon collector's problem and makes us see the enormous difficulty of achieving complete collections in which obtaining their elements depends purely on chance. We think that, perhaps this difficulty is the reason why for many years and until today, this type of collections have become so popular.

References

[1] Philippe Flajolet, Danièle Gardy, and Loÿs Thimonier. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, 39(3):207–229, November 1992.