MATHEMATICAL PROBLEM SOLVING
The coupon collector's problem

Group 4 Members :
Hawa Diop
Jean de Dieu NGIRINSHUTI
David SIBOMANA
Emmanuel DJIMMO TALLA

September 18, 2024

Problem

Each box of a certain breakfast cereal contains one of ten different coupons, each with the same probability. We win a prize if we manage to obtain a complete collection of all the different coupons. How long on average do we have to wait? (For example, suppose we draw the following coupons in order: 5, 2, 3, 7, 5, 1, 1, 4, 8, 4, 1, 9, 10, 2, 3, 3, 6. With the last coupon 6, we have completed our collection, and so we stop after 17 steps.)

Problem Statement

Let X denote the total number of cereal boxes needed to collect all 10 unique coupons. We can break this down into individual random variables X_i, where X_i represents the number of cereal boxes required to obtain the i-th coupon.

For example:

- X_1: Number of boxes to collect the 1st coupon.
- X_2: Number of boxes to collect the 2nd coupon (after the 1st).
- $\vdots$
- X_{10}: Number of boxes to collect the 10th coupon (after the 9th).

Therefore, the total number of boxes is:

$$X = X_1 + X_2 + \cdots + X_{10}$$

The figure below illustrates the steps involved in collecting all 10 coupons.

Coupon Collection Process

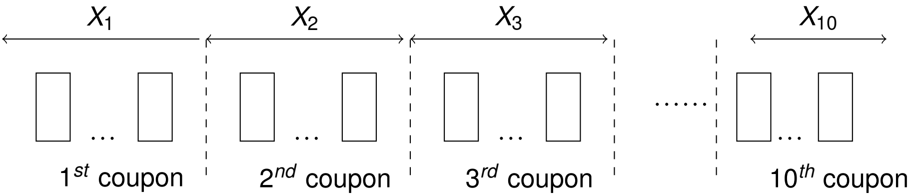

Figure 1: Coupon Collection Process.

Probability

- The first coupon is collected immediately, so the probability to collect the first coupon is

$$P(X_1 = 1) = 1$$

- Let's find the probability of getting the 2nd coupon:

Decision Tree

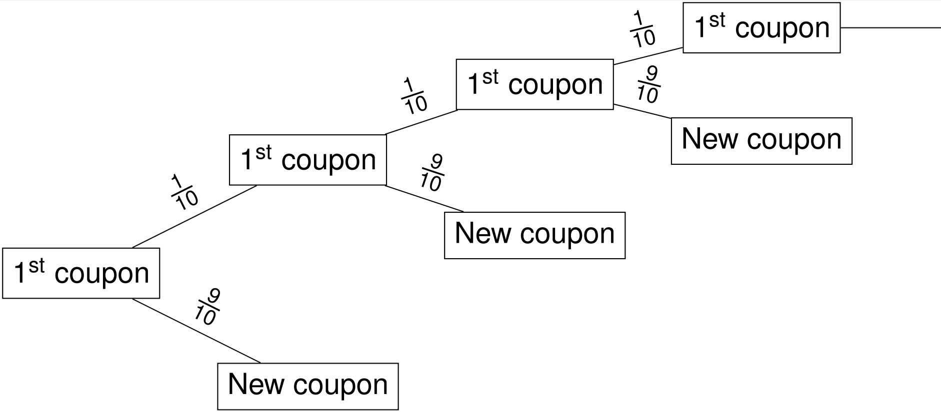

Figure 2: Decision Tree for the Drawing of the Second Coupon

Using the decision tree above, we have:

$$
\begin{aligned}
P(X_2 = 1) &= \frac{9}{10} \\
&= \left(\frac{10-2+1}{10}\right)\left(1 - \frac{10-2+1}{10}\right)^{1-1}, \\
P(X_2 = 2) &= \frac{1}{10} \times \frac{9}{10} \\
&= \left(\frac{10-2+1}{10}\right)\left(1 - \frac{10-2+1}{10}\right)^{2-1}.
\end{aligned}
$$

General Case

- In general, the probability that it takes i boxes to get the second coupon is

$$P(X_2 = i) = \left(\frac{10-2+1}{10}\right)\left(1-\frac{10-2+1}{10}\right)^{i-1}, \quad i \in \mathbb{N}.$$

- Let's now determine the probability of obtaining the 3rd coupon :

Decision Tree

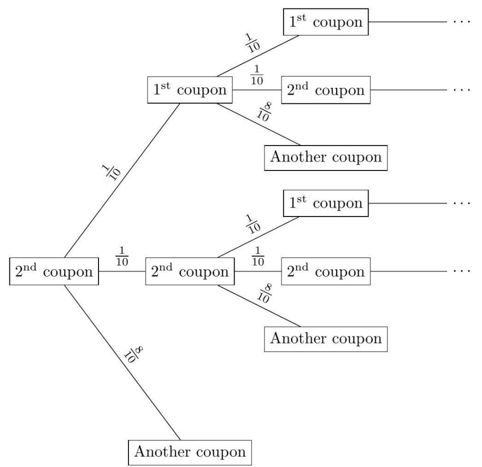

Figure 3: Decision Tree for the Drawing of the Second Coupon.

Probability

Cont'd

$$
\begin{aligned}
P(X_3 = 1) &= \frac{8}{10} \\
&= \left(\frac{10-3+1}{10}\right)\left(1-\frac{10-3+1}{10}\right)^{1-1} \\
P(X_3 = 2) &= \frac{1}{10}\times\frac{8}{10}+\frac{8}{10}\times\frac{1}{10} \\
&= \frac{2}{10}\times\frac{8}{10} \\
&= \left(\frac{10-3+1}{10}\right)\left(1-\frac{10-3+1}{10}\right)^{2-1}
\end{aligned}
$$

Cont'd

- In general, the probability that it takes i boxes to get the third coupon is

$$P(X_3 = i) = \left(\frac{10-3+1}{10}\right)\left(1-\frac{10-3+1}{10}\right)^{i-1}, \quad i \in \mathbb{N}.$$

Expectation

- obviously the probability that it takes i boxes to get k^{th} coupon given that we already have $k-1$ coupon is

$$P(X_k = i) = \left(\frac{10-k+1}{10}\right)\left(1-\frac{10-k+1}{10}\right)^{i-1} \quad k = 1, 2 \ldots 10.$$

- It is clearly that $X_k \sim \text{Geom}\left(\frac{10-k+1}{10}\right)$
  Since X_k follow a geometric distribution with $\left(\frac{10-k-1}{10}\right)$ as probability of success then

$$\begin{aligned} E(X_k) &= \frac{1}{\frac{10-k+1}{10}} \\ &= \frac{10}{10-k+1} \end{aligned}$$

Cont'd

So, $E(X) = \sum_{k=1}^{10} E(X_k)$

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \ldots\ldots + E(X_{10}) \\ &= \frac{10}{10-1+1} + \frac{10}{10-2+1} + \frac{10}{10-3+1} + \cdots + \frac{10}{10-10+1} \\ &= 1 + \frac{10}{9} + \frac{10}{8} + \cdots + \frac{10}{2} + 10 \\ &\simeq 29.29 \end{aligned}$$

On average, We will need to open around 30 cereal boxes to collect all 10 unique coupons.

Conclusion

Suppose we want to collect n different coupons. The probability of collecting a new coupon after drawing i boxes, given that we already have $k-1$ coupons, is:

$$P(X_k = i) = \left(\frac{n-k+1}{n}\right)\left(1 - \frac{n-k+1}{n}\right)^{i-1}$$

where $X_k \sim \text{Geom}\left(\frac{n-k+1}{n}\right)$ with parameter $P_k = \frac{n-k+1}{n}$

$$E(X_k) = \frac{1}{P_k}$$

$$E(X) = \sum_{k=1}^{n} E(X_k)$$

Cont'd

$$
\begin{aligned}
E(X) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\
&= \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n} \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} \\
&= n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\
&= n.H_n
\end{aligned}
$$

where H_n is the n^{th} harmonic number

For $n \to \infty$, $H_n \approx \log n + \gamma$ where γ is the Euler-Mascheroni constant, approximately 0.577. . Thus , $E(X) = n.\log n$

End

Thank you for your Kind attention