Coupon Collector Problem

- The Problem: Suppose inside each cereal box, there is one of twelve possible coupons. On average, how many cereal boxes will be required to collect all twelve coupons?

- Considerations:

  1. Are each of the twelve possible coupons *equally* likely to show up in each cereal box?
  2. The required number of boxes would vary each time the game is played. Therefore, this value is a *random* quantity since it is not predictable.
  3. What does "on average" mean?

- "Average value":

  1. If many people participated in the coupon collector game, what would be the average number of cereal boxes required for the group?
  2. Ask students for predictions of average value for Math Circle group.
  3. Have students (in pairs/groups) play game (simulate step 1) with twelve sided dice and compare average value with predictions. Denote by

$$X_i = \text{ required number of boxes by group } i$$

$$S_n = X_1 + X_2 + \cdot + X_n = \text{ total number of required boxes for all the } n \text{ groups}$$

and S_n/n to be the average value.

Trial No.	X_1	X_2	X_3	...	X_n	S_n	S_n/n
1				...			
2				...			

  4. Question: Will this average value change if the group plays the game again?
  5. If time permits, have students play again and compare new average value with previous value.
  6. Conclusion: The average value changes with each set of games; i.e. it is also *random*
  7. Question: What do we know about this average value?
     - Does it vary a lot or a little each time the game is played?
     - How is the variability of the average value affected by the number of groups that plays the game? Direct discussion to the conjecture that as the number of groups increases the variability of the average value decreases toward a particular value.
  8. Test previous conjecture with average value of 12-sided die.
     (a) Record variability of averages of groups of size two $S_2/2$, size six $S_6/6$, size ten $S_{10}/10$. Each student simulate each average values.

Student	$S_2/2$	$S_6/6$	$S_{10}/10$

(b) What value does the average value seem like it is going toward as n increases?

- Expected value:

1. Definition: For a random variable X that takes on k different values

$$\begin{aligned}\text{expected value} &= \text{(possible value 1)(prob. of value 1)} + \text{(possible value 2)(prob. of value 2)} \\ &+ \cdots + \text{(possible value } k\text{)(prob. of value } k\text{)}\end{aligned}$$

2. Compute expected value of outcome of 12-sided die.
3. Important: Average value is random but expected value is NOT (deterministic).
4. Law of Large Numbers:

$$\text{average value} \to \text{expected value}$$

as size of groups n increases.

5. Conclusion: By the Law of Large Numbers, the expected value yields the "true" average value of the random variable. Therefore, when we say "average value", we mean "expected value".
6. Property of expected value:

$$E[X+Y] = E[X] + E[Y]$$

Illustrate with X = outcome of 10 - sided die and Y = outcome of fair coin toss.

- Expected value of Coupon Collector Problem:

1. Let

$$C_i = \text{number of cereal boxes after the } i-1 \text{ to get } i \text{ coupon}$$

2. Question: What is the value of C_1 and $E[C_1]$?
3. If C =number of required cereal boxes to get all ten coupons, then

$$C = C_1 + C_2 + \cdots + C_{12}$$

and by property of expected values

$$E[C] = E[C_1] + E[C_2] + \cdots + E[C_{10}]$$

4. Expected value of C_i: The random variable C_i is an example of a geometric distribution which we will discuss in future session

$$E[C_i] = \frac{12}{12-(i-1)}$$

Expected value of Coupon Collector Problem:

$$E[C] = \frac{12}{12-0} + \frac{12}{12-1} + \cdot + \frac{12}{12-11} = 12\left(\frac{1}{1} + \frac{1}{2} + \cdot + \frac{1}{12}\right) = 37.24$$

- Observations:

1. We can generalize result to any number of possible coupons.
2. Coupon collector problems in disguise (discuss other types of problems that are coupon collector problems)

- Further directions: Ask students to figure out correct experiment using 12 - sided die to investigate parts 1. and 2. below. Carry out experiment and then prove mathematically using above result.

1. Average number of cereal boxes between coupons; e.g. 4th and 5th coupons.
2. Average number of cereal boxes to complete set after 3 coupons.
3. Average number of cereal boxes to complete set of 20 coupons. How does the average value depend on the number of distinct coupons q?

- Coupon Collector Problem applet:

  http://www.math.uah.edu/stat/applets/CouponCollectorExperiment.html