Coupon Collector Problem

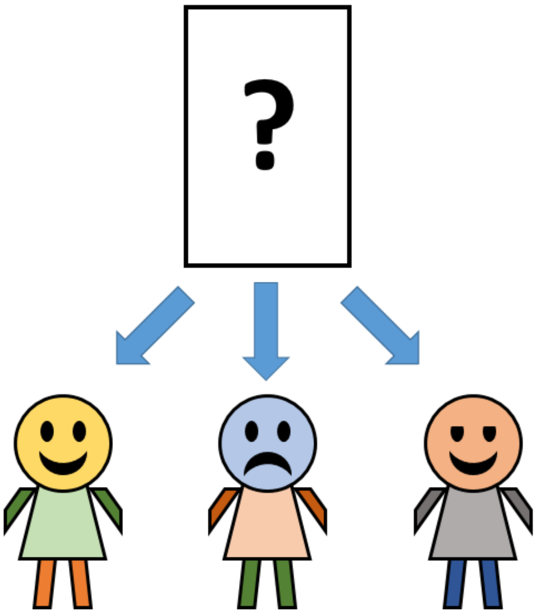
In the **coupon collector problem**, the goal is to purchase distinct objects in order to make a complete set of objects. Ea gives a random object, and the contents are independent of all other purchases. "Coupon" is just a placeholder word; the collected can be any kind of object.

Mathematically, the goal of the problem is to quantify the effort required to complete the collection. This is done by calcul expected value of purchases of an object in order to obtain a full set of those objects.

TRY IT YOURSELF

Each sealed *Storelings* pack contains a single *Storeling* toy chosen at random. There are 3 distinct *Storelings* toys in the set, and each toy has an equal chance to be in each pack. Also, each pack is independent of every other pack.

Submit your ans



What is the expected value of the number of packs one would need to open in order to obtain at least one copy of each toy?

One of the most interesting observations of the coupon collector problem is that a collection becomes *more difficult* to co one approaches a full collection. The final item in a collection typically requires the *most* effort to obtain.

The mathematical principles behind this problem are useful for problems involving any number of different types of things collectible card games (CCGs), sports cards, and, as seen in the example above, collectible toys.

Contents

General Case
Variance

General Case

There are many possible variations of the coupon collector problem. The following is the most basic case of the coupon c problem:

The Coupon Collector Problem

There is a bin that contains n distinct objects. Each "purchase" consists of selecting an object out of the bin at random replacing it. X is the discrete random variable that represents the number of purchases until each of the n objects is s least once.

The goal of the coupon collector problem is to find $\mathrm{E}[X]$.

Typically, the "bin" is interpreted to be some pack in which the contents are randomized. One of the key aspects of this p that the selections are mutually independent (selection with replacement). Practically speaking, this means that objects c selected more than once. However, in this problem, multiple selections of the same object do not matter; only selections objects matter.

The linearity of expectation page contains several problems relating to the coupon collector problem. These problems sh attempted before moving on to the generalization of the coupon collector problem, below.

General Solution to the Coupon Collector Problem

For the coupon collector problem as stated above, the expected value of the number of purchases required in order to each of the n objects at least once is:

$$\mathrm{E}[X] = nH_n$$

Where H_n is the n^{th} harmonic number.

PROOF

Show Proof

Let X be the discrete random variable that represents the number of purchases until each of the n objects is selecte least once.

Let X_k be the discrete random variable that represents the number of purchases after the $(k-1)^{\text{th}}$ distinct object t the k^{th} distinct object. As a base case, $X_1 = 1$, because the first object selected will always be distinct.

By linearity of expectation,

$$\mathrm{E}[X] = \sum_{k=1}^{n} \mathrm{E}[X_k]$$

.

After the $(k-1)^{\text{th}}$ distinct object is selected, there are $n-k+1$ objects remaining to be selected. Let A_k be the e that one of those objects is selected in the next purchase. Then by probability by outcomes,

$$P(A_k) = \frac{n-k+1}{n}$$

.

X_k follows a geometric distribution (of trials). It's expected value is the reciprocal of $P(A_k)$:

$$\mathrm{E}[X_k] = \frac{1}{P(A_k)} = \frac{n}{n-k+1}$$

From before, $\mathrm{E}[X]$ is equal to the sum of all these expectations:

$$\mathrm{E}[X] = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{k=1}^{n} \frac{1}{k}$$

Thus, $\mathrm{E}[X] = nH_n$, where H_n is the n^{th} harmonic number.

There is no closed form expression for the n^{th} harmonic number. However, there does exist a very good estimate for H_n

For relatively large n, the n^{th} harmonic number can be approximated as:

$$H_n \approx \ln n + \gamma + \frac{1}{2n}$$

Where γ is the Euler-Mascheroni Constant, $\gamma \approx 0.577216$.

This gives an approximate solution to the coupon collector problem:

Approximate Solution to the Coupon Collector Problem

For the coupon collector problem as stated above, the expected value of the number of purchases required in order to each of the n objects at least once is approximated as:

$$\mathrm{E}[X] \approx n(\ln n + \gamma) + \frac{1}{2}$$

Where γ is the Euler-Mascheroni constant, $\gamma \approx 0.577216$.

This approximation tends to be very accurate, and is more accurate with larger n. For $n \geq 5$, the approximation is accur decimal place.

EXAMPLE

The collectible card game *Arcane: The Congregation* just came out with a new set of cards. Cards are sold in sealed p each pack contains a single random rare card. The contents of each pack is independent of every other pack, and eacl equally likely.

If there are 55 distinct rare cards in the newest set, what is the expected number of packs needed to obtain at least one each rare card? Round your answer to the nearest integer.

This problem, as described, fits the format of a coupon collector problem. Packs are independent, the contents are ran each rare is equally likely.

If X is the random variable that represents the number of packs opened until each of the 55 rares is collected at least $\mathrm{E}[X]$ could be calculated with:

$$\mathrm{E}[X] = 55H_{55}$$

However, it would be very tedious to calculate the 55^{th} harmonic number. Fortunately, the problem as stated only requ answer to the nearest integer. Therefore, the approximation can be used:

$$E[X] \approx 55(\ln 55 + 0.577216) + \frac{1}{2} \approx 252.650$$

The expected number of packs needed to obtain at least one copy of each rare card, rounded to the nearest integer, is

Incidentally, using computer software to compute the exact answer, it is $E[X] \approx 252.6486716559$.

TRY IT YOURSELF

Submit your ans

Sealed packs of the collectible card game *Arcane: The Congregation* have a $\frac{1}{8}$ chance to contain one of the 15 distinct mythic cards, chosen at random. If a pack contains a mythic card, each of the 15 mythic cards is equally likely.

If each pack of *Arcane: The Congregation* is independent of all other packs, what is the expected value of the number of packs needed to obtain at least one copy of each mythic card? Round your answer to the nearest integer.

Note: The coupon collector problem page contains an approximation that makes the solution easier to compute, and the approximation is sufficient to obtain the correct answer.

TRY IT YOURSELF

Submit your ans

Madeleine has a fair 100 sided die, each face numbered with a distinct number from 1 to 100.

She decides to roll it and record the number until she has rolled all the numbers at least once.

What is the expected value for the number of rolls she will need to make?

Please round your answer to the nearest integer.

Other Expected Value Quizzes

Image credit: toyspedia.blogspot.com

TRY IT YOURSELF

Submit your ans

Jorge has an N-sided fair die, and wonders how many times he would need to roll it until he has rolled all the numbers from 1 to N (in any order).

He does a quick calculation and discovers that the expected value for the number of rolls is 91 when rounding to the nearest integer.

How many sides does his die have?

Clarification: One distinct number from 1 to N is printed on each face of the die.

Other Expected Value Quizzes

Image credit: ravnerdwars.info.

Variance

It should be noted that the solution to the coupon collector problem is not a guarantee. The expected value of purchases need to make in order to complete a collection is not the same as what will actually happen in practice. In fact, there is si variability in the number of purchases needed to complete a collection. This variability can be quantified with the statistics and standard deviation.

Variance of the Coupon Collector Problem

For the coupon collector problem as stated in the general case section above, the variance of the number of purchases in order to select each of the n objects at least once is:

$$\text{Var}[X] = n^2 \sum_{k=1}^{n} \frac{1}{k^2} - nH_n$$

Where H_n is the n^{th} harmonic number.

PROOF

Show Proof

Let X be the discrete random variable that represents the number of purchases until each of the n objects is selecte least once.

Let X_k be the discrete random variable that represents the number of purchases after the $(k-1)^{\text{th}}$ distinct object t
the k^{th} distinct object.

Given that each X_k is independent,

$$\operatorname{Var}[X] = \sum_{k=1}^{n} \operatorname{Var}[X_k]$$

.

After the $(k-1)^{\text{th}}$ distinct object is selected, there are $n-k+1$ objects remaining to be selected. Let A_k be the e
that one of those objects is selected in the next purchase. Then by probability by outcomes,

$$P(A_k) = \frac{n-k+1}{n}$$

.

X_k follows a geometric distribution. It's variance is:

$$\operatorname{Var}[X_k] = \frac{1-P(A_k)}{[P(A_k)]^2} = \frac{1-\frac{n-k+1}{n}}{\left(\frac{n-k+1}{n}\right)^2} = \frac{n(k-1)}{(n-k+1)^2}$$

From before, $\operatorname{Var}[X]$ is equal to the sum of all these variances:

$$\begin{aligned} \operatorname{Var}[X] &= \sum_{k=1}^{n} \frac{n(k-1)}{(n-k+1)^2} \\ &= n \sum_{k=1}^{n} \frac{n-k}{k^2} \\ &= n^2 \sum_{k=1}^{n} \frac{1}{k^2} - n \sum_{k=1}^{n} \frac{1}{k} \\ &= n^2 \sum_{k=1}^{n} \frac{1}{k^2} - nH_n \end{aligned}$$

Thus, $\operatorname{Var}[X] = n^2 \sum_{k=1}^{n} \frac{1}{k^2} - nH_n$, where H_n is the n^{th} harmonic number.

You might notice that the variance calculation contains a very inconvenient finite sum in $\sum_{k=1}^{n} \frac{1}{k^2}$. Fortunately, there is a v
approximate this sum using the Basel Problem.

The **Basel Problem** gives the following value for the convergent series:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

Thus, for relatively large n,

$$\sum_{k=1}^{n} \frac{1}{k^2} \approx \frac{\pi^2}{6}$$

With this approximation, and the previous approximation for nH_n, an approximation can be given for $\text{Var}[X]$:

Approximate Variance of the Coupon Collector Problem

For the coupon collector problem as stated in the general case section above, the variance of the number of purchases in order to select each of the n objects at least once is approximated as:

$$\text{Var}[X] \approx \frac{\pi^2}{6}n^2 - n(\ln n + \gamma) - \frac{1}{2}$$

Where γ is the Euler-Mascheroni constant, $\gamma \approx 0.577216$.

Cite as: Coupon Collector Problem. *Brilliant.org*. Retrieved 13:46, April 27, 2018, from https://brilliant.org/wiki/coupon-collector-problem/