Task Description

Li Hua(the familiar name as you know) is addicted to collecting stamps. There are N kinds of stamps in the packages of a good(like instant noodles) and each package contains only one stamp. Each kind of stamp is equally likely distributed in each package. What is the average number(or expectation) of the packages he should buy to collect a whole set of stamps?

To solve the solution, as the joke of a mathematician who want to be a firefighter* said, we should transformed a question we are not so familiar to the question we have known. Therefore, we should introduce geometry distribution first.

Geometric Random Variable

Suppose that independent trials, each having a probability $p\,(0<p<1)$ of being a success are performed until a success occurs. If we let X equal the number of trials required, then we have

$$P(X=n)=(1-p)^{n-1}p$$

We call X a geometric random variable with parameter p.And it is easy to confirm that

$$\sum_{n=1}^{\infty}P(X=n)=1$$

And to calculate its expected value, we have

$$E[X]=\sum_{n=1}^{\infty}n(1-p)^{n-1}p=p\sum_{n=1}^{\infty}n(1-p)^{n-1}=\frac{p}{(1-(1-p))^2}=\frac{1}{p}$$

where

$$\sum_{n=1}^{\infty}nx^{n-1}=\frac{1}{(1-x)^2}\ ,\ x\in(-1,1)$$

The variance won't be discussed in the task, so if you are interested enough(if not enough you probably do not want to do this), you can calculate its variance by yourself.

Back to the Task

We first consider the situation that if Li Hua holds $i(0\le i\le n-1)$ kinds of stamps wants to collect a new kind what is the average of packages he should buy in addition.

It is easy to calculate the probability of the event, which is $p=\frac{N-i}{N}$ Because a whole set contains N kinds of stamps and Li Hua have collected i kinds of them, the number of categories of stamps he has not collect are $N-i$ .So the expected value is $\frac{1}{p}=\frac{N}{N-i}$.

As Lao Zi once said everything on the earth is generated by being and being comes from nothing. We first focus on the situation that Li Hua holds no stamps. Then when he buys a new package the probability he gets a new kind of stamp is 1. And if he wants to get a second kind of stamp he needs to consume $\frac{N}{N-1}$ packages on average. When he gets the i_{th} kind of stamps, he needs to consume $\frac{N}{N-i}$ packages on average.

Therefore, if he wants to collect a whole set of stamps, he should buy the following number of packages on average:

$$E(event) = \sum_{i=0}^{N-1} \frac{N}{N-i} = N \sum_{i=0}^{N-1} \frac{1}{N-i} = N \sum_{i=1}^{N} \frac{1}{i}$$

If $N = 12$ then we can get the expected value is $37.2385\ldots$

Also, there is an approximation of harmonic series:

$$\sum_{i=1}^{n} \frac{1}{i} = \ln n + \gamma + o(\frac{1}{n}) \ \ (n \to \infty)$$

Thus,

$$E(event) = N \sum_{i=1}^{N} \frac{1}{i} = N \ln N + N\gamma + o(1) \ \ (N \to \infty)$$

where $\gamma \approx 0.5572\ldots$ is Euler's constant.

Further Discussion

If the packaging machines like gambling(this is just an assumption), when they package one package they would like to roll a dice. It will package the stamps(equally likely and independent) whose number is the result of dice into one package. What is the average number of package Li Hua should to buy to collect a whole set of stamps? To simply the model we assume that there are 20 kinds of stamps.

Attention: if you use$E(event) = (N \sum_{i=1}^{N} \frac{1}{i})/3.5$ to calculate the answer you get is probably wrong(why?).

The Joke Mentioned at the Beginning

*One day, a mathematician decides he wants to be a firefighter. So he quits his job and applies to be a firefighter. But the chief doesn't think he is qualified so he gives him a test.

He takes him to the alley and sets a dumpster on fire.

He asks the mathematician, “What do you do?” The mathematician immediately grabs a hose and puts out the fire.

The chief then asks him, “Now that the fire is out, what do you do?”

Stumped, the mathematician thinks for a minute and says “I can reduce this problem into a problem with known solutions.”

With this he pulls out a match and sets the dumpster on fire.

If you find any mistake, please contact the author.