Numbers in Boxes

Sebastian Liu

Problem

A group of 100 individuals play the following game: in a room, there are 100 boxes, and inside each box, there is a slip of paper containing the number of that box (Box 1 has a slip of paper inside that reads "1", Box 2 has a slip of paper inside that reads "2", and so on). The slips of paper are then randomly mixed so that each box contains a random number (still one slip of paper per box). The goal of the game is for each individual to find his number (individual 1 must find the slip of paper which reads "1", individual 2 must find the slip of paper which reads "2", and so on). Only one individual is allowed in the room at a time, and he is only allowed to open up to 50 boxes. After he is done (either by finding his number or by opening up 50 boxes), the room is reset exactly as it was before he went inside. The group of 100 individuals wins if every single person finds his own number. How do they maximize their chances of winning?

A Losing Strategy

A sure way for the group to lose is if they decide that everyone should open the first 50 boxes. A justification of this claim is left as an exercise to the reader.

A Lower Bound

A naive solution is for each individual to generate 50 random numbers, $1-100$ and open those boxes. Each individual then has a $\frac{1}{2}$ chance of finding his number. We leave the task of calculating the probability that each of the 100 individuals finds his own number as an exercise to the reader.

A Winning Strategy

As surprising as it may be, there is a strategy that allows the group of 100 to win over 30% of the time. Each individual starts at his box number. He opens his box and either he finds his own number or he finds another number. If he finds his number then he's done. Otherwise, he goes to the box number on the slip of paper inside his current box. He repeats this process until either he finds his number, or until he has opened 50 boxes.

Notice that if he ever returns to his box, then he has found his number. The justification of this claim is left for the reader to fill in.

Here is an example of what a sequence of box openings might look like for individual 1: $1 \to 17 \to 72 \to 33 \to 1$, where each number here represents the box number. In such a scenario, individual 1 has found his number. Notice that this sequence begins and ends at 1. Let us call this a "cycle". The length of the cycle here is 4 (we don't double count 1). It should be clear that given our strategy, if all cycles have length less than 50, then the group wins. It remains to be shown that this probability is anywhere close to 30%.

The probability that every cycle is of length 50 of shorter is equivalent to the complement of the probability that there is a cycle of length 51 or 52 or 53 and so on until 100. Notice that there can exist at most one cycle of length greater than 50, since if there is a cycle of length n, the remaining individuals can create a cycle of length at most $100 - n$. Therefore the probability that there is a cycle of length, say 51 is disjoint from the probability that there is a cycle of length, say 52.

So what is the probability that there is a cycle of length 51? Well we have 100 numbers to choose from, and we must pick 51 numbers to be in our cycle, so there are exactly $\binom{100}{51}$ to do this. But given 51 numbers, there are 50! ways to form cycles with these numbers. Simply fix the first number (which represents the individual who is opening boxes), and arrange the remaining 50 in whatever order. For each cycle we form with the 51 numbers, there are $(100 - 51)!$ ways to arrange the remaining numbers. Taken together, the number of ways to generate a cycle of 51 numbers from a pool of 100 numbers is:

$$\binom{100}{51} 50!\,49! = \frac{100!}{51!\,49!} \cdot 50!\,49! = \frac{100!\,50!\,49!}{51 \cdot 50!\,49!}$$

which is simply $\frac{100!}{51}$. Now this is the number of ways to generate a cycle of length 51 from 100 numbers. The probability that such a cycle exists is the number of ways to generate this cycle divided by the number of ways to arrange 100 numbers. But the number of ways to arrange 100 numbers is 100!, so the probability that a cycle of length 51 exists is

$$\frac{100!}{51} \cdot \frac{1}{100!}$$

and this is just $\frac{1}{51}$. We leave the computations for the probability of a cycle being length 52, 53, and so on as an exercise to the reader.

Since the probabilities of cycles of length greater than 50 are disjoint, the probability that there is a cycle of length greater than 50 is given by

$$\frac{1}{51} + \frac{1}{52} + \frac{1}{53} + \cdots + \frac{1}{98} + \frac{1}{99} + \frac{1}{100} = \sum_{i=51}^{100} \frac{1}{i}$$

which we can approximate by the following:

$$\sum_{i=51}^{100} \frac{1}{i} \approx \int_{51}^{100} \frac{1}{x}\,dx = \ln(x)\Big|_{51}^{100} = \ln(100) - \ln(51) = \ln\left(\frac{100}{51}\right) \approx \ln(2) \approx 0.69$$

The complement of this, the probability that all cycles are of length 50 or less, is $1 - 0.69 = 0.31$, or 31%.

We can generalize this problem where there are a total of n number of boxes and where each individual is allowed to open m number of boxes. This is left as an exercise to the reader.

Concluding Remark

Why does this method work better than simply having each individual randomly pick 50 boxes to open? Well suppose we have the following cycle:

$$1 \to 17 \to 72 \to 33 \to 1$$

Once 1 finds his number, we have guaranteed that individuals 17, 72 and 33 will find their respective numbers. A justification of this is left to the reader. Such a guarantee does not exist when each individual is randomly selecting 50 boxes to open.