The 100 Prisoners Problem

The 100 Prisoners Problem

In a prison there are 100 prisoners, each with a number from 1 to 100.

The 100 Prisoners Problem

In a prison there are 100 prisoners, each with a number from 1 to 100.

One day the warden decides to play a game.

The 100 Prisoners Problem

In a prison there are 100 prisoners, each with a number from 1 to 100.

One day the warden decides to play a game.

A room contains 100 boxes numbered 1 to 100. Into those boxes she randomly distributes pieces of paper with the numbers 1 to 100, one per box.

She tells the prisoners:

She tells the prisoners:

You may enter the room, one at a time, and open up to 50 boxes. After you leave all the boxes will be closed and it will be the next prisoner's turn.

She tells the prisoners:

You may enter the room, one at a time, and open up to 50 boxes. After you leave all the boxes will be closed and it will be the next prisoner's turn.

If every one of you finds your own number, you may all go free.

She tells the prisoners:

You may enter the room, one at a time, and open up to 50 boxes. After you leave all the boxes will be closed and it will be the next prisoner's turn.

If every one of you finds your own number, you may all go free.

But if even one of you fails to find your own number, you all stay in prison.

She tells the prisoners:

You may enter the room, one at a time, and open up to 50 boxes. After you leave all the boxes will be closed and it will be the next prisoner's turn.

If every one of you finds your own number, you may all go free.

But if even one of you fails to find your own number, you all stay in prison.

You may discuss and plan ahead of time, but once the game starts, you may not communicate.

What should they do?

If a prisoner just opens 50 boxes at random, the probability they will find their own number is $\frac{1}{2}$.

If a prisoner just opens 50 boxes at random, the probability they will find their own number is $\frac{1}{2}$.

So if every prisoner does this, their overall probability of success is

$$\frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2} = \frac{1}{2^{100}} \approx 0.0000000000000000000000000000008.$$

If a prisoner just opens 50 boxes at random, the probability they will find their own number is $\frac{1}{2}$.

So if every prisoner does this, their overall probability of success is

$$\frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2} = \frac{1}{2^{100}} \approx 0.0000000000000000000000000000008.$$

They might as well not bother.

However, there exists a strategy where the probability of success is more than 30%.

What is it?

The strategy is as follows:

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.
- If box x contains number x, you are done! Otherwise, box x contained the number y.

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.
- If box x contains number x, you are done! Otherwise, box x contained the number y.
- Open box y. If it contains the number x, you are done! Otherwise, box y contained the number z.

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.
- If box x contains number x, you are done! Otherwise, box x contained the number y.
- Open box y. If it contains the number x, you are done! Otherwise, box y contained the number z.
- Open box z. If it contains the number $x, \ldots$

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.
- If box x contains number x, you are done! Otherwise, box x contained the number y.
- Open box y. If it contains the number x, you are done! Otherwise, box y contained the number z.
- Open box z. If it contains the number $x, \dots$
- Repeat until either you find the number x, or you open 50 boxes without finding x.

The strategy is as follows:

- If you are prisoner number x, start by opening the box labelled x.
- If box x contains number x, you are done! Otherwise, box x contained the number y.
- Open box y. If it contains the number x, you are done! Otherwise, box y contained the number z.
- Open box z. If it contains the number $x, \ldots$
- Repeat until either you find the number x, or you open 50 boxes without finding x.

But why does this work?

A bit of technical language: the arrangement of 100 numbers into 100 boxes is called a permutation.

A bit of technical language: the arrangement of 100 numbers into 100 boxes is called a permutation.

Now note that if you were not restricted to just 50 boxes, you would eventually get to number x no matter what.

A bit of technical language: the arrangement of 100 numbers into 100 boxes is called a permutation.

Now note that if you were not restricted to just 50 boxes, you would eventually get to number x no matter what.

Following this procedure from box to box traces out a loop. It might be a big loop (all 100 boxes) or a trivial loop (only 1 box) or something in between. This loop is called a cycle. A cycle with n steps is an n-cycle.

A bit of technical language: the arrangement of 100 numbers into 100 boxes is called a permutation.

Now note that if you were not restricted to just 50 boxes, you would eventually get to number x no matter what.

Following this procedure from box to box traces out a loop. It might be a big loop (all 100 boxes) or a trivial loop (only 1 box) or something in between. This loop is called a cycle. A cycle with n steps is an n-cycle. (Sadly a 2-cycle is called a transposition instead of a bicycle.)

A bit of technical language: the arrangement of 100 numbers into 100 boxes is called a permutation.

Now note that if you were not restricted to just 50 boxes, you would eventually get to number x no matter what.

Following this procedure from box to box traces out a loop. It might be a big loop (all 100 boxes) or a trivial loop (only 1 box) or something in between. This loop is called a cycle. A cycle with n steps is an n-cycle. (Sadly a 2-cycle is called a transposition instead of a bicycle.)

If we look at all the cycles that everyone traces out, we get the cycle structure of the permutation.

So now the overall probability of success is the probability that all of the cycles in the cycle structure have length at most 50. (If there was a cycle of length 51 or longer, nobody on that cycle would find their own number in 50 steps.)

So now the overall probability of success is the probability that all of the cycles in the cycle structure have length at most 50. (If there was a cycle of length 51 or longer, nobody on that cycle would find their own number in 50 steps.)

So we can write

$P(\text{success}) = 1 - P(\text{failure})$

So now the overall probability of success is the probability that all of the cycles in the cycle structure have length at most 50. (If there was a cycle of length 51 or longer, nobody on that cycle would find their own number in 50 steps.)

So we can write

$$
\begin{aligned}
P(\text{success}) &= 1 - P(\text{failure}) \\
&= 1 - P(\text{cycle of length} \geq 51)
\end{aligned}
$$

So now the overall probability of success is the probability that all of the cycles in the cycle structure have length at most 50. (If there was a cycle of length 51 or longer, nobody on that cycle would find their own number in 50 steps.)

So we can write

$$
\begin{aligned}
P(\text{success}) &= 1 - P(\text{failure}) \\
&= 1 - P(\text{cycle of length} \geq 51) \\
&= 1 - [P(\text{51-cycle}) + P(\text{52-cycle}) + \cdots + P(\text{100-cycle})]
\end{aligned}
$$

So now the overall probability of success is the probability that all of the cycles in the cycle structure have length at most 50. (If there was a cycle of length 51 or longer, nobody on that cycle would find their own number in 50 steps.)

So we can write

$$\begin{aligned} P(\text{success}) &= 1 - P(\text{failure}) \\ &= 1 - P(\text{cycle of length} \geq 51) \\ &= 1 - [P(\text{51-cycle}) + P(\text{52-cycle}) + \cdots + P(\text{100-cycle})] \end{aligned}$$

That looks pretty complicated!

But now for some magic! First consider P(100-cycle).

But now for some magic! First consider P(100-cycle).

(See Nick's terrible handwriting.)

But now for some magic! First consider P(100-cycle).

(See Nick's terrible handwriting.)

So $P(\text{100-cycle}) = \frac{1}{100}$.

But now for some magic! First consider $P(\text{100-cycle})$.

(See Nick's terrible handwriting.)

So $P(\text{100-cycle}) = \frac{1}{100}$.

Similarly $P(\text{99-cycle}) = \frac{1}{99}$.

But now for some magic! First consider $P(100\text{-cycle})$.

(See Nick's terrible handwriting.)

So $P(100\text{-cycle}) = \frac{1}{100}$.

Similarly $P(99\text{-cycle}) = \frac{1}{99}$.

And $P(98\text{-cycle}) = \frac{1}{98}$.

But now for some magic! First consider P(100-cycle).

(See Nick's terrible handwriting.)

So $P(\text{100-cycle}) = \frac{1}{100}$.

Similarly $P(\text{99-cycle}) = \frac{1}{99}$.

And $P(\text{98-cycle}) = \frac{1}{98}$.

And all the way down to $P(\text{51-cycle}) = \frac{1}{51}$.

Note that this fails for 50: in fact $P(\text{50-cycle}) = \frac{99}{5000} < \frac{1}{50}$.

Note that this fails for 50: in fact $P(\text{50-cycle}) = \frac{99}{5000} < \frac{1}{50}$.

And in general $P(k\text{-cycle}) < \frac{1}{k}$ for $k \leq 50$.

Note that this fails for 50: in fact $P(\text{50-cycle}) = \frac{99}{5000} < \frac{1}{50}$.

And in general $P(k\text{-cycle}) < \frac{1}{k}$ for $k \leq 50$.

This is because for those values of k we could have more than one k-cycle, so the calculation becomes more complicated.

Note that this fails for 50: in fact $P(50\text{-cycle}) = \frac{99}{5000} < \frac{1}{50}$.

And in general $P(k\text{-cycle}) < \frac{1}{k}$ for $k \leq 50$.

This is because for those values of k we could have more than one k-cycle, so the calculation becomes more complicated.

However we don't need those values! Going back to our main formula,

$$
\begin{aligned}
P(\text{success}) &= 1 - [P(51\text{-cycle}) + P(52\text{-cycle}) + \cdots + P(100\text{-cycle})] \\
&= 1 - \left[\frac{1}{51} + \frac{1}{52} + \cdots + \frac{1}{100}\right] \\
&\approx 1 - 0.688172 \\
&= 0.311828
\end{aligned}
$$

This method works for any even number n of prisoners: the probability of success is always more than 30%.

This method works for any even number n of prisoners: the probability of success is always more than 30%.

Bonus problem: Show that as $n \to \infty$, we get

$$P(\text{success}) \to 0.306852...$$

and find the exact value of this number.

This method works for any even number n of prisoners: the probability of success is always more than 30%.

Bonus problem: Show that as $n \to \infty$, we get

$$P(\text{success}) \to 0.306852...$$

and find the exact value of this number.

Hint: You will need to remember how sums become integrals. Also remember

$$\int \frac{1}{x} dx = \log x$$