

The condemned prisoners and the boxes.

A group of 100 condemned prisoners are offered the chance to play the following game. They will be each be allowed to enter a room one at a time in any order they wish until all 100 have gone in once. In the room there is a long table with a row of 100 identical wooden boxes numbered in order from 1 to 100. Each box can be opened and inside it there is the name of one of the prisoners, such that each of the 100 names appears exactly once. However, the ordering of the names in the boxes is random.



Once inside the room, each prisoner is allowed to open and look inside 50 boxes—any 50 that he wants. After he has opened his 50 boxes he must leave the room, making sure that the boxes are in exactly the same state as he found them, and he is no longer able to communicate with any other prisoner. The reprieve that they have been granted is that at the end of the game they will all be spared on condition that every prisoner manages to find the box that contains his own name. On the other hand, if at least one of the prisoners fails to find his name, they will all be executed at dawn. Let's be clear about this—in order to escape execution, it is necessary that *every* prisoner find his own name.

The prisoners are allowed to get together before-hand and work out a strategy. How good a strategy can they find and what is the resulting probability of success?

This is a fascinating problem. If you spend a bit of time reflecting on the situation (and you should!) you will start to feel that the odds of them all surviving seem microscopic. For example, it seems clear enough that no matter what strategy is used, the first prisoner will have only a 50% chance of finding his name. And if he does, he can leave no clue as to whether or not he did find his name or what boxes he opened or what he found in them. And then the second prisoner comes in. What can he possibly do to raise his probability much above 50%? So we're down to something close to $(0.50)^2 = 0.25$ after only two prisoners. *And there are 98 prisoners left to go!* Just to look at an extreme case, if every prisoner simply opened 50 boxes at random, the probability they'd all win would be 0.50 raised to the power 100 and that's a very small number.

Here is the remarkable result. There is a strategy which will save them all from execution *with probability greater than 30%*.

Even more remarkable—the solution is elegant and simple. So simple it is easily understood.

But simple as it is, the solution is not, I think, easily found. For me as a mathematician, the absorbing problem is not one of finding the solution (I had to be told the answer) but of understanding it. Having been given the solution, *understand how it works*. That already is a wonderful little project.

Reposted Jan 3, 2010. I'm grateful to Michal Dobrogost for pointing out a technical error in the earlier solution.

And again reposted Oct 1 2012 with technical thanks to Roy Dyckhoff of the University of St Andrews.

This is just so hard to believe—such a strategy seems preposterous. How could you get a 30% chance that *every* prisoner will find his name?

Okay. Here's the solution—here's that simple elegant strategy I promised you. I'll present it for the case of 8 prisoners who are each permitted to open 4 boxes.

What the prisoners should first do is construct an ordering in which they will enter the room. This assigns to each prisoner a number between 1 and 8. It doesn't matter what it is, but each should be provided with a list of this ordering. That's the only preparation they need to do as a group. Given this, you can think of each box as having two numbers, the number on the top which every prisoner can see, and the number inside, the number of the prisoner's name that's hidden inside the box. This is an unknown (random) rearrangement of the numbers from 1 to 8.

box	inside
1	7
2	3
3	8
4	6
5	1
6	4
7	5
8	2

Now here's what each prisoner does. The first prisoner, #1, opens box 1. If he happens to find his own name there, he's happy and he leaves, otherwise *he takes the name he finds in that box and opens the box labeled with that prisoner's number*. In the example at the right, the name he finds is that of prisoner #7, and therefore he next opens box 7. He continues in this way. Following the example at the right, the name he finds in box 7 is that of prisoner #5, so he next opens box 5. Luckily the name he finds there is that of prisoner #1 and that's him! and he is happy. If he hadn't found his name there, he would have been able to open one more box, his fourth. Then prisoner #2 comes in and does the same thing except he starts with box 2. He takes the name he finds there (#3), and opens that box. Etc. And then prisoner #3 enters. Etc. Finally, all 8 prisoners will have visited the room.

In the above example do they all find their own name? Yes they do—let's check it out by following each prisoner. Start with #1 and list the boxes he opens. We get:

1 7 5 win

Now #2:

1	2	3	4	5	6	7	8
---	---	---	---	---	---	---	---

2 3 8 win

Then #3:

7	3	8	6	1	4	5	2
---	---	---	---	---	---	---	---

3 8 2 win

Actually we didn't need to do #3. We had already handled #3 in doing #2. That's a significant observation. *The fact that #2 wins implies that all those whose names belong to the boxes that #2 opened will also win,* in this case, #3 and #8. The path 2 to 3, 3 to 8, and 8 back to 2, is called a cycle of length 3. All prisoners on that cycle find their name after opening 3 boxes.

That observation gives us 6 of the 8 prisoners who win. The remaining two are #4 and #6 and they form a cycle of size 2: (46). So they both also win. In fact this gives us an interesting way of representing the original permutation: as a product of cycles:

$$(175)(238)(46)$$

Can you see how this cyclic representation holds the key that unlocks the success of this strategy? Indeed, your job is to calculate the probability of success for this strategy, and then attempt to understand why it works so (surprisingly) well.

What I do with the class at this point is have each student generate a random permutation of the 8 numbers, and then work out whether or not it yields success. With 60 students I get a pretty good statistical sample and usually get close to the true theoretical value. This exercise also gives the students a chance to sit back and "internalize" the algorithm.

$$(175)(238)(46)$$

This cyclic representation is a compact way of specifying the permutation. It tells us that box 1 contains name 7, box 7 contains name 5, box 5 contains name 1, box 2 contains name 3, etc.

Every permutation has a unique (up to order) representation as a product of cycles. If we write the top-inside permutation as a product of cycles, then the key observation is that the length of the cycle containing any prisoner's number is the number of boxes he would have to open to find his name, and therefore *the prisoners who find their name are precisely the ones whose cycle has length at most 4.*

We conclude that the permutation will meet success precisely when its cyclic representation has all cycles of length ≤ 4 . All such permutations have the property that all 8 prisoners will find their names.

It remains to calculate the probability that a permutation of 8 symbols will have all cycles of length ≤ 4 . To do this it helps to know basic things about permutations and combinations.

Let's start by counting the number of permutations of 8 symbols that have a cycle of length 5. To get a cycle of length 5 we need to choose

5 of the symbols and there are $\binom{8}{5}$ ways to do that. Then note that

there are $4!$ possible cycles with 5 symbols—start with any one and then write the others in all possible orderings. Finally, for each of these there are $3!$ ways to permute the remaining 3 symbols. The total count is:

$$\binom{8}{5} 4! 3! = \frac{8!}{5!3!} = \frac{8!}{5}.$$

Now the total number of permutations of 8 symbols is $8!$. And hence the *probability* that a random permutation will have a 5-cycle is $1/5$. A surprisingly simple result, but it's not easy to "see" right away why it ought to be true. [At least I haven't found a simple argument.]

Similarly the probability that a random permutation will have a 6-cycle is $1/6$, etc. Now we want the probability that all cycles have length ≤ 4 . That's the same as the probability of having no cycles of length 5 or 6 or 7 or 8. Well we can calculate the probability that a random permutation will *have* a cycle of length 5 or 6 or 7 or 8. It will be:

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \approx 0.63$$

and hence the probability of no cycle of length greater than 4 is 1 minus this which is 0.37. This strategy delivers a whopping 37% chance of success.

What about more prisoners? Our argument generalizes exactly. The probability of failure for 10 prisoners (having a cycle of length > 5) is

$$\frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} \approx 0.64.$$

Actually, without doing the calculation, we can easily see that the second sum exceeds the first, as $1/5$ is clearly less than $1/9 + 1/10$. Generalizing this, we see that the failure probability increases as the number of prisoners increases (for even numbers of prisoners). Therefore the success probability *decreases* and the question then is: how small does the success probability get?

Note that these are disjoint possibilities—for example, you couldn't have both a cycle of length 5 *and* a cycle of length 7. Thus we can sum the individual probabilities.

Of course we are already able to write down the answer for 100 prisoners. The failure probability is

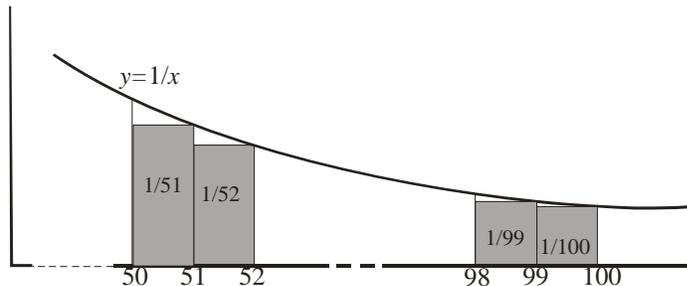
$$\frac{1}{51} + \frac{1}{52} + \frac{1}{53} + \dots + \frac{1}{99} + \frac{1}{100}$$

and with a programmable calculator or a spreadsheet we could easily calculate this. It turns out to be 0.688 which is less than 0.7 and that confirms our result that the probability that all 100 prisoners will find their names exceeds 30%.

But it's of considerable interest to discover what happens as the number of prisoners gets very large. Does the success probability approach zero, or does it have a positive limiting value? This is actually a very nice problem.

A lovely way to represent these sums is as the area of a family of rectangles. Anyone who has worked with Riemann sums might well suspect that such an approach might work. What we want is an area representation that allows us to see easily how the different sums compare.

The obvious graph to work with is $y = 1/x$. It takes a bit of playing around to see how to set things up, but the final argument is quite elegant. Consider the graph below. The curve has equation $y = 1/x$, and fifty rectangles are erected on the interval $[50, 100]$ each with base 1. The sum of their areas is exactly the above sum.



Now the sum of these areas is less than the area under the curve on $[50, 100]$, and this is:

$$\int_{50}^{100} \frac{1}{x} dx = \ln x \Big|_{50}^{100} = \ln 100 - \ln 50 = \ln \frac{100}{50} = \ln 2 = 0.69$$

The failure probability is less than this so the success probability exceeds $1 - \ln 2 = 0.31$. And the proof is complete.

Error analysis: It's interesting to estimate the error in the above inequality. The success probability exceeds $1 - \ln 2$ but by how much? It's simplest to look at the failure probability

$$\text{failure probability} = \frac{1}{51} + \frac{1}{52} + \frac{1}{53} + \dots + \frac{1}{99} + \frac{1}{100} < \int_{50}^{100} \frac{1}{x} dx = \ln 2.$$

The error in the above inequality is the sum of the triangular regions and this is less than what we'd get if we made those triangular regions into real triangles (by giving them straight hypotenuses). Now these triangles all have base 1 so their area is half their height and the sum of these areas is then:

$$\begin{aligned} \frac{1}{2} [\text{sum of heights}] &= \frac{1}{2} \left[\left(\frac{1}{50} - \frac{1}{51} \right) + \left(\frac{1}{51} - \frac{1}{52} \right) + \dots + \left(\frac{1}{98} - \frac{1}{99} \right) + \left(\frac{1}{99} - \frac{1}{100} \right) \right] \\ &= \frac{1}{2} \left[\frac{1}{50} - \frac{1}{100} \right] = \frac{1}{200}. \end{aligned}$$

So the failure probability is within $1/200$ of $\ln 2$.