The Monty Hall Problem

Karl Stratos

1 Original Version

Let $X, Y, Z \in \{1, 2, 3\}$ denote random variables representing

- Z: The door behind which the prize car lies
- X: The door I *choose* (but have not opened)
- Y: The door the host opens to reveal a goat

The generative process is

$$x, z \overset{\text{iid}}{\sim} \text{Unif}(\{1, 2, 3\})$$
$$y \sim p_{Y|XZ}(\cdot|x, z)$$

where

1. If the door I choose is the prize door, the host can open either of the two remaining doors.

$$p_{Y|XZ}(y|x, z) = \begin{cases} \frac{1}{2} & \text{if } y \in \{1, 2, 3\} \setminus \{x\} \\ 0 & \text{otherwise} \end{cases} \qquad x = z$$

2. If the door I choose is not the prize door, the host must select the only door that is neither my door nor the prize door.

$$p_{Y|XZ}(y|x, z) = \begin{cases} 1 & \text{if } y \in \{1, 2, 3\} \setminus \{x, z\} \\ 0 & \text{otherwise} \end{cases} \qquad x \neq z$$

The host never opens my door or the prize door. Assume $X = 1$ and $Y = 3$ without loss of generality. We have

$$\Pr(X = 1, Y = 3, Z = 1) = \left(\frac{1}{3}\right)^2 \times \underbrace{p_{Y|XZ}(3|1, 1)}_{\frac{1}{2}} = \frac{1}{18}$$
$$\Pr(X = 1, Y = 3, Z = 2) = \left(\frac{1}{3}\right)^2 \times \underbrace{p_{Y|XZ}(3|1, 2)}_{1} = \frac{1}{9}$$
$$\Pr(X = 1, Y = 3, Z = 3) = \left(\frac{1}{3}\right)^2 \times \underbrace{p_{Y|XZ}(3|1, 3)}_{0} = 0$$

thus $\Pr(X = 1, Y = 3) = \frac{1}{6}$ and

$$\Pr(Z = 1|X = 1, Y = 3) = \frac{1/18}{1/6} = \frac{1}{3} \qquad < \qquad \Pr(Z = 2|X = 1, Y = 3) = \frac{1/9}{1/6} = \frac{2}{3}$$

Intuitively, if my door (door 1) *is* the prize door, the host could have opened door 2 instead; the fact that he didn't makes it more likely that he opened door 3 because he was forced to.

2 Generalization

Assume now there are $n \geq 3$ doors. The random variables are

- $Z \in \{1 \dots n\}$: The door behind which the prize car lies
- $X \in \{1 \dots n\}$: The door I *choose* (but have not opened)
- $Y \subset \{1 \dots n\}$: The $n-2$ doors the host opens to reveal goats

The generative process is

$$\begin{gathered} x, z \overset{\text{iid}}{\sim} \text{Unif}(\{1 \dots n\}) \\ y \sim p_{Y|XZ}(\cdot|x,z) \end{gathered}$$

where

1. If the door I choose is the prize door, the host can open any $n-2$ doors out of the $n-1$ remaining doors. Since there are $\binom{n-1}{n-2} = n-1$ choices,
$$p_{Y|XZ}(y|x,z) = \begin{cases} \frac{1}{n-1} & \text{if } y \subset \{1 \dots n\} \setminus \{x\} \text{ and } |y| = n-2 \\ 0 & \text{otherwise} \end{cases} \qquad x = z$$
2. If the door I choose is not the prize door, the host must select the remaining $n-2$ doors:
$$p_{Y|XZ}(y|x,z) = \begin{cases} 1 & \text{if } y = \{1 \dots n\} \setminus \{x, z\} \\ 0 & \text{otherwise} \end{cases} \qquad x \neq z$$

The host never opens my door or the prize door. Assume $X = 1$ and $Y = \{3 \dots n\}$ without loss of generality. We have

$$\Pr(X=1, Y=\{3\dots n\}, Z=1) = \left(\frac{1}{n}\right)^2 \times \underbrace{p_{Y|XZ}(\{3\dots n\}|1,1)}_{\frac{1}{n-1}} = \left(\frac{1}{n}\right)^2 \left(\frac{1}{n-1}\right)$$

$$\Pr(X=1, Y=\{3\dots n\}, Z=2) = \left(\frac{1}{n}\right)^2 \times \underbrace{p_{Y|XZ}(\{3\dots n\}|1,2)}_{1} = \left(\frac{1}{n}\right)^2$$

$$\Pr(X=1, Y=\{3\dots n\}, Z=z) = \left(\frac{1}{n}\right)^2 \times \underbrace{p_{Y|XZ}(\{3\dots n\}|1,z)}_{0} = 0 \qquad \forall z \in \{3 \dots n\}$$

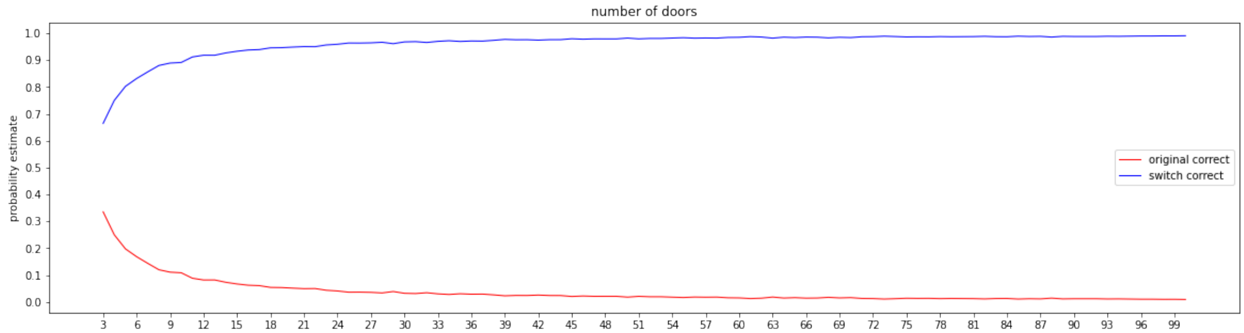
thus $\Pr(X=1, Y=\{3\dots n\}) = \left(\frac{1}{n}\right)^2 \left(\frac{n}{n-1}\right)$ and

$$\Pr(Z=1|X=1,Y=\{3\dots n\}) = \frac{\left(\frac{1}{n}\right)^2 \left(\frac{1}{n-1}\right)}{\left(\frac{1}{n}\right)^2 \left(\frac{n}{n-1}\right)} = \frac{1}{n} \quad < \quad \Pr(Z=2|X=1,Y=\{3\dots n\}) = \frac{\left(\frac{1}{n}\right)^2}{\left(\frac{1}{n}\right)^2 \left(\frac{n}{n-1}\right)} = \frac{n-1}{n}$$

Intuitively, if my door (door 1) *is* the prize door, the host could have opened any set of $n-2$ doors other than $\{3 \dots n\}$ out of $\{2 \dots n\}$; the fact that he didn't, even though he would have had so many options, makes it overwhelmingly likely that he opened doors $\{3 \dots n\}$ because he was forced to.

3 Discussion

Switching the door (to the only other alternative door), instead of adhering to the original door, after the host reveals $n-2$ goats makes it $n-1$ more likely that I will win the prize. Here is a simulation that estimates the probability of winning the prize with either the original door or the switched door over 10,000 games for each $n = 3, 4, \dots, 100$:



Even after proving and simulating the statement, I *still* find it counterintuitive, because of my deeply rooted belief that the host, who acts after I make a selection and does not even tell me where the prize is, should not affect my original decision. As noted by Glymour *et al.* (2016), we must approach the problem in terms of what options the host had before revealing goats.

- My door, no matter what it is, was never an option for the host to open.
- The other door (neither mine nor one of the host's revealed doors), *was* an option for the host to open but he didn't.
- My door was never put through a test of refutation, while the other door withstood the test. Therefore, the latter is the more likely answer.

References

Glymour, M., Pearl, J., and Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.