

Probability taught visually

The Monty Hall Problem

A Spreadsheet Simulation

This problem which has had a history of generating controversy is often used to teach conditional probability. The article explains how repeated trials can be simulated using the random number generator in Excel, and how the result can be experimentally verified and then explained.

JONAKI GHOSH

One of the fundamental concepts in statistics is that of probability. It forms an integral part of most mathematics curricula at high school level. The topic of probability can be enlivened using many interesting problems. The related experiments are, however, time consuming and impractical to conduct in the classroom. Simulation can be an effective tool for modeling such experiments. It enables the student to use random number generators to generate and explore data meaningfully and, as a result, grasp important probability concepts. This section discusses a well known problem known as the Monty Hall problem or the three door problem which is based on conditional probability and lends itself to investigation. Its exploration using a spreadsheet such as MS Excel can lead to an engaging classroom activity. It highlights the fact that handheld calculators enabled by spreadsheet capabilities can enable students to visualize, explore and discover important concepts without necessarily getting into the rigor of mathematical derivations.

Exploring the Monty Hall Problem

Students often find it hard to understand the concept of conditional probability. This is aptly highlighted by the Monty Hall Problem (also referred to as the three door problem) which goes as follows

You are a contestant in a game show where the host (Monty) asks you to choose one among three doors. Behind one of the doors is a car (the prize) and behind the other two are goats. After you select a door, the host doesn't open it; instead, he opens one of the other two doors and reveals a goat (the host, who knows what's behind the doors, never reveals the car). The host then asks you whether you want to stick to your original choice or you would like to switch by choosing the other door. The question is which option is more likely to win you the prize?

This puzzle and its correct solution was published by Marilyn vos Savant in the *Sunday Parade* magazine in the 90s. She claimed that the contestant should switch because the odds of winning the car would be $2/3$, whereas if one sticks to the original door, then the odds of winning would be only $1/3$. This solution created a furor and thousands of readers including professional mathematicians retorted angrily by writing that switching cannot matter, since after one of the goats is revealed, there are two doors left and the probability of obtaining the car is $1/2$.

One wrote: "As a professional mathematician, I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error and, in future, be more careful."

Indeed this problem can lead to an interesting classroom discussion. The immediate reaction of most students (like most people) is also that switching cannot matter and that the probability of obtaining the car (after Monty opens one of the doors and reveals a goat) is $1/2$. This is however incorrect. The probability that you will select the door with the car (and therefore win) is $1/3$, and probability that you will lose is $2/3$. If your initial choice was correct then switching would be wrong. However if you chose a wrong door initially (the probability of which is $2/3$) then switching would lead you to win (since the host will reveal the goat behind the other wrong door). Figure 1 shows a tree diagram to explain the problem. Here it is assumed that the player has chosen door 1.

The tree diagram shows that the total probability of winning by switching is $2/3$ whereas the probability of winning by staying with one's original choice is $1/3$. The problem can be analyzed using the concept of conditional probability which will be described towards the end of the article.

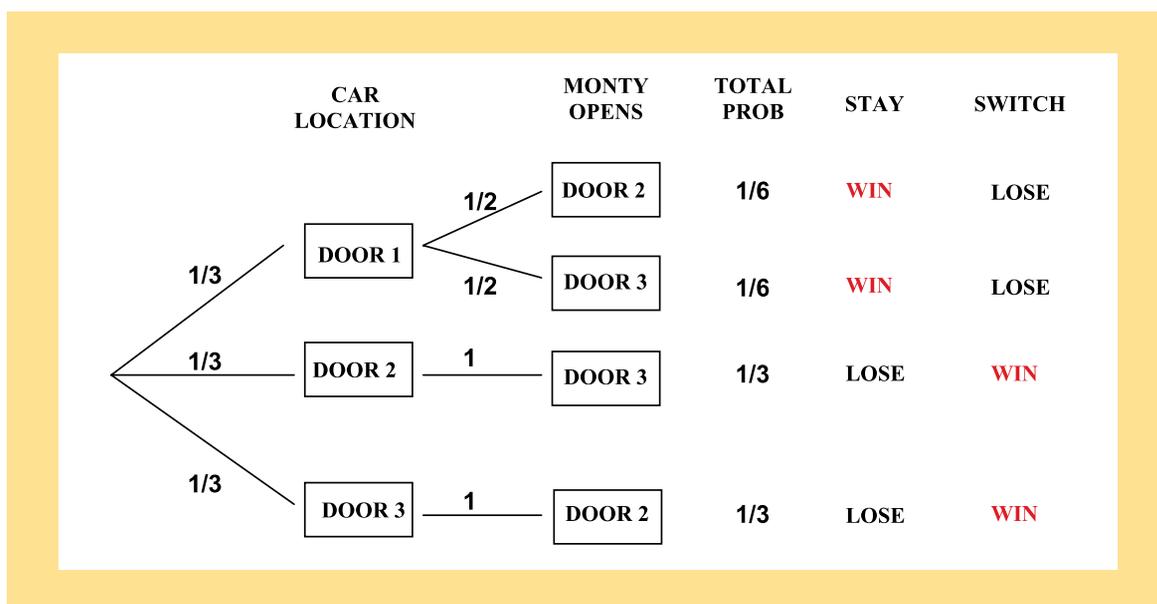


Figure 1 A tree diagram to explain the Monty Hall problem; the player's initial choice is door 1.

Spreadsheet verification in Excel also reveals that the probabilities of winning by switching are $2/3$ and winning by staying is $1/3$. In Figure 2 the experiment has been simulated 100 times with the assumption that the player's initial choice is door 1. The numbers in column A indicate the door behind which the car is placed. This is done by randomly generating the integers 1 through 3, both inclusive. We assume that the player's initial choice is door 1 in all the simulations and enter 1 throughout in column B. To get the entries in column C, we first check whether the entries for A and B are the same (i.e., the player has selected door with the car). If so, we generate a random number between 0 and 1. If the entries of column A and B are both equal to 1, that is, the car is behind door 1 (which is also chosen by the player), then Monty may open either door 2 or door 3 (with probability $1/2$ each) since he cannot reveal the car. This is indicated by the random number between 0 and 1. A random number less than 0.5 indicates that Monty opens door 2 in which case 2 is entered in column C. However, if the random number is 0.5 or greater, then 3 is entered in column C indicating Monty's opening door 3. These entries in column C are filled using a simple macro.

In column D we identify the cases where the player wins by switching. If the entries of

column A and B are different, then YES appears in the corresponding cell. Similarly in column E we identify the cases where the player wins by staying with his original choice. Thus if the entries of columns A and B are the same then YES appears in the corresponding cell. Finally counting the YES's in column D reveals the total number of times the player wins by switching in 100 trials. Interestingly, this number, on an average turns out to be around 67. Thus one may conclude that the probability of winning by staying with one's choice is around 33% and that by switching is 67%.

Simulation of the Monty Hall Problem on Excel

The problem can be simulated on Excel by using the **RAND()**, **INT()** and **If()** functions. The steps of simulation are as follows

Step 1: The first step is to simulate the position of the car behind one of the three doors. This is done by generating 100 integers among 1, 2 and 3 in column A. Enter **=INT(3*RAND()+1)** in cell A2 and drag till cell A101. When creating a column of numbers by dragging, we need to drag the small box in the lower right hand corner of the cell. 100 randomly generated integers will appear in column A. These represent with door behind which the car is placed (See Figure 2).

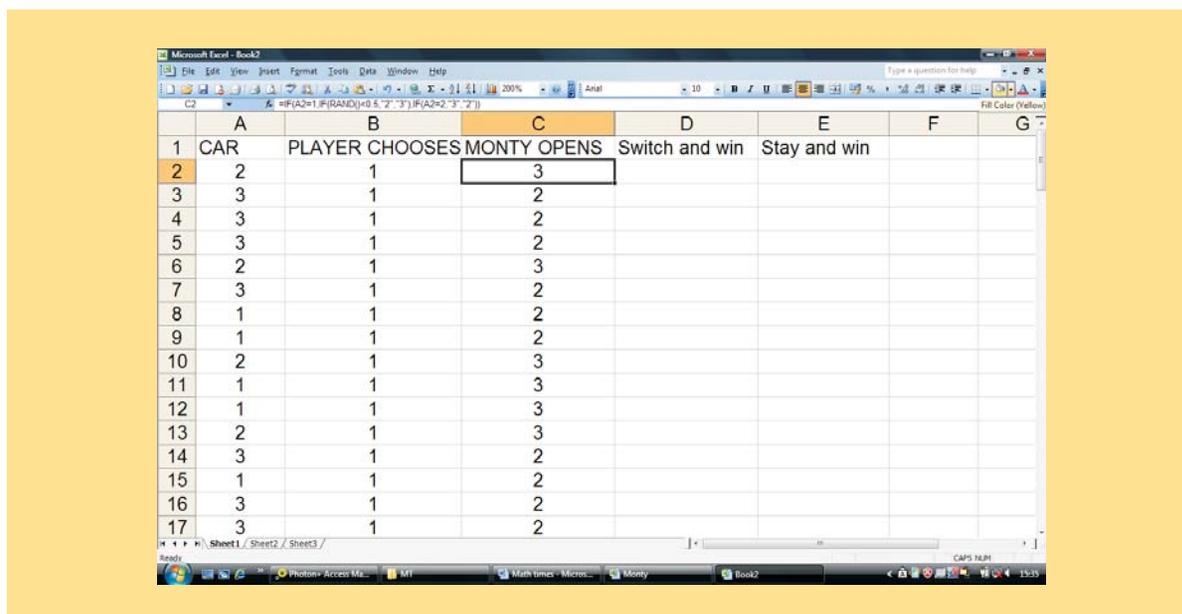


Figure 2 Simulating the Monty Hall problem using Excel.

Step 2: In this simulation we are assuming that the player's choice is door 1. Enter 1 in cell B2 and drag till B100 or simply double click on the corner of cell B2.

Step 3: The next step is to simulate the door opened by Monty. This is done by entering `=IF(A2=1, IF(RAND()<0.5, "2", "3"), IF(A2=2, "3", "2"))` in cell C2 followed by a double click on the corner of the cell. This step simulates the door opened by Monty. If the car is behind door 1 and the random number generated by `RAND()` is less than 0.5 then Monty opens door 2, otherwise he opens door 3. On the other hand if the car is behind door 2, then Monty opens door 3 and if the car is behind door 3, Monty opens door 2.

Step 4: In column D we identify the cases in which the player wins by switching and in column E we identify the cases where the player wins by staying with his original choice. In column D cell should show 'YES' if the players choice is not the same as the position of the car and 'NO' otherwise. This is done by entering `=IF(A2<>B2, "YES", "NO")` in cell D2 followed by a double click in the corner of the same cell. In column E a cell should indicate 'YES' if the player's choice is the same as the position of the case and 'NO' otherwise. This can be achieved by entering `=IF(A2=B2, "YES", "NO")` in E2 followed by a double click in the corner of the cell.

Step 5: Finally we need to count the number of 'YES's in columns D and E to find the number of cases in which the player wins by switching and staying with his choice respectively. This can be obtained by entering `=COUNTIF(D2:D101, "YES")/100` and `=COUNTIF(E2:E101, "YES")/100` respectively in separate cells.

The Monty Hall Problem—An Analysis using Conditional Probability

Let C_i denote the event that the car is behind door i . Clearly $P(C_i) = \frac{1}{3}$ where $i = 1, 2, 3$. Also let M_{ij} denote the event that Monty opens door j when the player chooses door i .

$$P(M_{ij}/C_k) = \begin{cases} 0, & \text{if } i = j \\ 0, & \text{if } j = k \\ 1/2, & \text{if } i = k \\ 1, & \text{if } i \neq k, j \neq k \end{cases}$$

Note that here i denotes the door chosen by the player, j denotes the door opened by Monty and k denotes the door behind which the car is placed.

Now, $P\left(\frac{M_{ij}}{C_k}\right) = 0$ if $i = j$, since Monty will never open the door chosen by the player. Also $P\left(\frac{M_{ij}}{C_k}\right) = 0$ if $j = k$, since Monty will never open the door which has the car behind it.

Figure 3 Simulating the probabilities of winning by switching and staying with one's choice.

If the player chooses the door with the car behind it (that is, $i = k$), then Monty can open any of the other two doors with probability $1/2$ each, thus for $i = k$.

For the case $i \neq j \neq k$, that is, the case when the player chooses a door which does not have the car behind it, then Monty has no option but to open the third door (the one which was not chosen by the player and the one which does not have the car behind it). Hence $P\left(\frac{M_{ij}}{c_k}\right) = 1$ for $i \neq j \neq k$.

Now let us compute the probability of the event M_{13} that is the event that Monty opens door 3 when the player chooses door 1. Using conditional probability we have

$$\begin{aligned} P(M_{13}) &= P(M_{13} \cap C_1) + P(M_{13} \cap C_2) \\ &\quad + P(M_{13} \cap C_3) \\ &= P(M_{13}/C_1)P(C_1) + P(M_{13}/C_2) \\ &\quad + P(M_{13}/C_3)P(C_3) \\ &= \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2} \end{aligned}$$

We would now like to compute the probability of the event C_1/M_{13} , that is, the event that the car is actually behind door 1 given that Monty opens door 3 when the player chooses door 1. This would give us the probability of the player winning the car by staying with his choice. Similarly we would like to compute the probability of the event, that is, the event that car is actually behind door 2 given that Monty opens door 3 when the player chooses door 1. This would give us the probability of the player winning the car by switching. The computations are as follows

$$P(C_1/M_{13}) = \frac{P(M_{13} \cap C_1)}{P(M_{13})} = \frac{P(M_{13}/C_1)P(C_1)}{P(M_{13})}$$

$$= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$\begin{aligned} P(C_3/M_{13}) &= \frac{P(M_{13} \cap C_3)}{P(M_{13})} = \frac{P(M_{13}/C_3)P(C_3)}{P(M_{13})} \\ &= \frac{0 \times \frac{1}{3}}{\frac{1}{2}} = 0 \end{aligned}$$

$$\begin{aligned} P(C_2/M_{13}) &= 1 - \{P(C_3/M_{13}) + P(C_3/M_{13})\} \\ &= \frac{2}{3} \end{aligned}$$

Thus $P\left(\frac{c_1}{M_{13}}\right) = \frac{1}{3}$ indicates that the probability of the player winning the car by staying with his choice is $1/3$ and $P\left(\frac{c_2}{M_{13}}\right) = \frac{2}{3}$ indicates that the probability of the player winning the car by switching is $2/3$.

It would be a good exercise to analyse the problem if the number of doors in the game is more than 3. Suppose there were 100 doors, you choose one of them, and Monty then opens 98 of the other doors, would you switch?

Conclusion

The topic of probability has a plethora of interesting problems which can be made accessible to high school students through spreadsheets. The experiments related to these problems may be impractical to conduct manually but simulation can be an effective modeling tool for imitating such experiments. Microsoft Excel proves to be a very handy tool for conducting the explorations and investigations in the classroom. The Monty Hall Problem discussed in this article can be conducted with students of grades 9 and 10 without getting into the mathematical derivations. However in grades 11 and 12 the spreadsheet verification of the problem can be followed by an analysis of the underlying concepts which are rooted in probability theory.

References

- [1] Microsoft Excel Training & Word 2007 Tutorial, Retrieved from <http://www.free-training-tutorial.com/format-cells.html>
- [2] Monty Hall Problem from Wikipedia, the free Encyclopedia, Retrieved from http://en.wikipedia.org/wiki/Monty_Hall_problem



JONAKI GHOSH is an Assistant Professor in Dept. of Elementary Education, Lady Sri Ram College, University of Delhi where she teaches courses related to math education. She obtained her Ph.D in Applied Mathematics from Jamia Milia Islamia University, New Delhi and her masters degree from IIT, Kanpur. She has taught mathematics at DPS, R K Puram for 13 years, where she was instrumental in setting up the Mathematics Laboratory & Technology Centre. She has started a Foundation through which she conducts professional development programmes for math teachers. Her primary area of research is in the use of technology in math instruction. She is a member of the Indo Swedish Working Group on Mathematics Education. She regularly participates in national and international conferences. She has published articles in proceedings and journals and has authored books for school students. She may be contacted at jonakibghosh@gmail.com.

SOLUTION FOR number crossword-1 by D.D. Karopady

	¹	²		³	⁴	⁵		
	1	9	0	9	7	8		
⁶	1	7	9	⁷	1	⁸	⁹	4
¹⁰	7	2		¹¹	4	¹²	¹³	2
		¹⁴	2	6	5	¹⁵	8	
¹⁶	1	0	8	0	¹⁷	2	0	4
		¹⁸	4	¹⁹	4	3	4	
²⁰	²¹		²²	2	5	6	²³	²⁴
	8	8					6	7
²⁵	1	2	²⁶	3		²⁷	1	6
				5				0
	²⁸	1	2	0	²⁹	2	1	6