

Probability Simulation

The Monty Hall Problem

Teacher Notes and Answers

7 8 9 10 11 12



Objectives

The aim is for students to become familiar with how a probability model can be used to simulate a random phenomenon and use the simulation results to explore a problem where the solution is counter intuitive.

Vocabulary

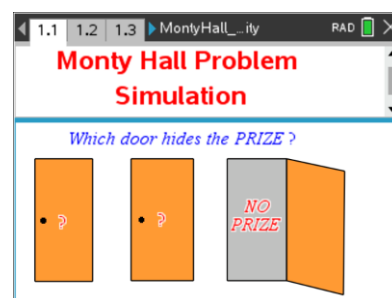
Simulation, trial, long run proportion, conditional probability, Bayes' theorem.

About this lesson

Students use the probability model to simulate a large number of trials and use the long run proportion of successful outcomes to make conjectures about the theoretical probabilities and why the counter intuitive winning strategy and probabilities arise.

Introduction

The probability puzzle which has come to be known as the *Monty Hall problem* is named after the original host of the American game show *Lets Make a Deal*. A solution to this puzzle was published in *Parade* magazine in 1990, sparking considerable controversy. Some mathematicians refused to believe the solution and wrote letters of complaint to the magazine.



Posing the problem

Suppose that a game show includes a game in which a contestant is shown 3 closed doors.

- A prize is hidden behind one of the doors.
- 'No prize' behind the other two doors.
- The contestant is invited to choose a door.
- The host then opens a door **not** chosen by the contestant to reveal 'No Prize'.
- The contestant is invited to either **stay** with the initial choice or **switch** to the other closed door.
- The contestant wins the prize if they choose the door that hides the prize.

What should the contestant do to give themselves the greatest chance of winning the prize?

1. **Switch** their choice of door.
2. **Stay** with their initial choice.
3. 'Staying' or 'switching' makes no difference to the probability of winning the prize.

Question 1

Which of the three options above do you think is correct? Briefly explain why you think that option is correct. (Student answers will vary).

Simulating the problem

In this activity, you will use random number, Lists and Spreadsheet, and Data and Statistics features of TI-Nspire to model the problem and observe graphical displays of the outcomes when the simulated game is repeated many times.

Assumptions in the simulation

- The door with the prize is labelled Door 1 and the doors with 'No prize' labelled Doors 2 and 3 (these are **NOT** necessarily the first, second and third doors, from left to right).
- If the contestant's initial choice is Door 1, the host makes a random choice between Door 2 or Door 3 to be opened.
- If the contestant's initial choice is Door 2, the host opens Door 3, and vice versa.
- A single trial in the simulation consists of selecting a door at random (to represent the contestant's initial choice), followed by randomly selecting another door which is neither the initial choice nor Door 1 (to represent the host's selection). The output shows the outcome for 'staying' and for 'switching'.
- A favourable outcome consists of the contestant's final selection being the door hiding the prize.
- There are 100 trials in each pass of the simulation. These will initially be looked at one trial at a time. Later in the activity, the results of repeated passes of the simulation will be aggregated to observe the long run outcome for a very large number of trials.

Note to teachers on setting up the simple simulation for 100 trials

The simulations in the TI-Nspire document '*Monty Hall_simulation*' have been constructed using built-in features of TI-Nspire (no coding involved) and are ready for students to use.

For teachers who may wish to create their own similar simulations, the instructions below show how a basic simulation can be set up. Further embellishments (such as sliders) can then be made to 'jazz up' the basic simulation.

Two key commands are

- Random integer (syntax: `randInt(lowBound, upBound [, #Trials])`)
- `ifFn` (syntax: `ifFn(BooleanExpr, Value_If_true [, Value_If_false])`)

Simulate 100 trials

A	initial...	B ho...	C stay	D switch	E
	=randint				
1	1	2	win	lose	wins_st...
2	1				
3	3				
4	1				
5					

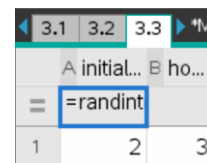
Column A: Contestant's initial choice
 Column B: Door opened by Host
 Column C: Result if contestant stays
 Column D: Result if contestant switches
 Cell e2: Count the wins if 'stay'. Input: `=countif('stay',"win")`
 Cell e4: Count the wins if 'switch'. Input: `=countif('switch',"win")`

Formulas shown in the spreadsheet:

- Cell A1: `=randint(1,3,100)`
- Cell B1: `=ifFn(A1=1,randint(2,3),ifFn(A1=2,3,2))`
- Cell C1: `=ifFn(A1=B1,"win","lose")`
- Cell D1: `=ifFn(A1+B1=5,"win","lose")`

- The formulas for columns B, C & D need to be filled down to cells B100, C100 & D100, as follows.
- Enter the formulas in cells B1, C1 and D1.
- Select cell B1, hold down the <shift> key and arrow across to cell D1.
- Open the context menu (on the software, right mouse click; on the handheld, press <ctrl> + <menu>). Select 'Fill' and press the down arrow to cells B100, C100 & D100. Press <enter>.
- To return to the top of the spreadsheet, press <ctrl> + <7>.

- Note that in cells e2 and e4, the apostrophe(') before 'stay and 'switch indicate that these are column references.
- For a new set of 100 trials, select 'cell A=' (as shown). Press <enter> + <enter>.
- For a dotplot of results, add a 'Data and Statistics' page. Press <Tab> and select 'Stay' from the pop-up menu on the horizontal axis. Press <menu>. Select: 2.Plot Properties > 5.Add X variable > Switch.
- To change the dotplot to a bar chart, press <menu>. Select: 1.Plot Type > Bar chart. Hover over the bars to display the count and percentage of favourable and unfavourable outcomes for the 'stay' and 'switch' strategies.



Success fractions for 'stay' and 'switch' strategies

Open the TI-Nspire document 'Monty Hall_simulation'. Navigate to page 2.3.

Page 2.3 is where the simulation calculations are taking place.

Page 2.4 shows graphically the count of favourable outcomes for the 'stay' and 'switch' strategies.

success fractions for 'stay' and 'switch' strategies, where

$$\text{success fraction} = \frac{\text{favourable outcomes}}{\text{number of trials}}$$

Trial	stay	switch	wins
1	lose	win	28
2	win	lose	72
3	lose	win	...
4	win	lose	72

Use the slider labelled 'game' to observe results one trial (game) at a time on pages 2.4 and 2.5. To reset, edit the value to 'game=0' (zero).

To obtain a new set of 100 trials, increase the 'repeat' slider value by 1.

Observe change in proportion as the number of trials increase.

Question 2

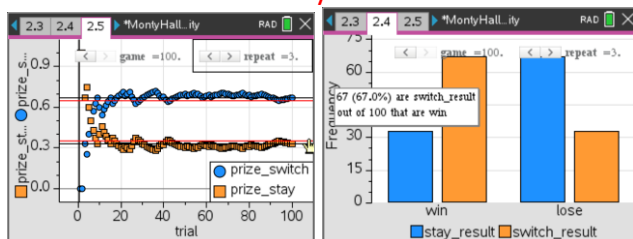
Observe changes in the success fractions on page 2.5 as the number of trials (games) increases from 1 to 100.

- Comment on the appearance and value of the success fractions as the number of trials increases.

Sample answer. As the number of trials increase, the success fraction converges towards 0.33 for the 'stay' option and towards 0.67 for the 'switch' option.

- Use the 'repeat' slider to select 100 new trials. Comment on comparisons between the success fraction graphs for different sets of 100 trials.

Sample answer. Although the patterns differ between sets of 100 trials, in all cases convergence towards 0.33 for 'stay' 0.67 for 'switch' is observed.



Question 3

Based on your observations in the counts and success fractions for several sets of 100 trials, comment on whether the 'stay, or the 'switch' option is more advantageous, or whether there is no advantage to either of these strategies.

Sample answer. The 'switch' option is clearly more advantageous.

Question 4

On page 2.5, click on each of the green horizontal lines.

a. What are the equations of these lines?

Sample answer. $y = 1/3$ and $y = 2/3$.

b. What behaviour are the success fractions for the 'stay' and 'switch' strategy showing with respect to these two lines, for any of the sets of 100 trials? What is the significance of this behaviour?

Sample answer. Convergence of the success fractions towards these lines as the number of trials increases is consistent with the theoretical probabilities are $1/3$ for the 'stay' option and $2/3$ for the 'switch' option.

Question 5

On page 2.5, set the slider 'game=100'. The horizontal red lines show the mean (average) of the success fraction for the 100 trials. Comment on the comparison between the red and green horizontal lines for different sets of 100 trials.

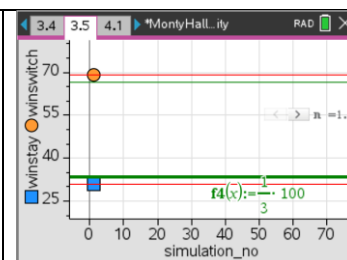
Sample answer. For 100 trials, the mean success fractions coincide with the lines with equations $y = 1/3$ and $y = 2/3$.

Long run proportion

Long run proportion is the ratio of favourable outcomes to the total number of trials in the experiment after conducting a **very large** number of trials.

On page 3.4 the results of repeated passes of the simulation of sets of 100 trials is captured. The results are aggregated (combined) to observe the behaviour of favourable outcomes for a very large number of trials.

On page 3.5, use the slider labelled 'n' to add new passes of the simulation of 100 trials. The blue dots on the graph represent the number of wins in 100 trials using the 'stay' strategy and the orange dots represent the number of wins in 100 trials using the 'switch' strategy. Observe the changes in the success fraction for a large number of sets of 100 trials.



To **RESET**: Edit slider to 'n=0'. Then, on page 3.4 select each cell with 'capture' in the formula and press <enter> + <enter>. This is a quick way to clear data.

Question 6

a. What are the equations of the two horizontal green lines on page 3.5?

Sample answer. $y = \frac{1}{3} \times 100$ and $y = \frac{2}{3} \times 100$

b. Based on the long run proportion observed in this part of the simulation, what is your best estimate of the theoretical probability of winning the prize for the

i. 'stay' strategy **Sample answer.** $1/3$

ii. 'switch' strategy **Sample answer.** $2/3$

c. In the long run, which of the two strategies will give contestants the greater chance of winning the prize?

Sample answer. In the long run, the contestants are twice as likely to win the prize if they 'switch'.

Question 7

Imagine that a student gave the following answer to Question 1 above.

'Staying' or 'switching' makes no difference to the probability of winning the prize. The reason is that, suppose the contestant initially chooses the middle door and the host opens the door on the left to reveal 'No prize'. There are two closed doors remaining. Therefore, whether the contestant stays with the middle door or switches to the door on the right, there is a probability of 0.5 that the prize will be behind that door.

- a. Do the results of the simulation support the student's prediction that 'stay' or 'switch' are equally likely to result in winning the prize?

Sample answer. Not consistent with simulation results (which are not close to 0.5).

- b. If not, explain why the prize is **not** equally likely to be behind either door after the host opens a door with 'No prize'?

Sample answer. This is because the host does not open the door at random. The door opened by the host depends on the contestant's initial choice and on the location of the prize. Therefore, once a door is opened to reveal 'No prize', there is a reduced sample space; it is now a matter of conditional probability (i.e. the contestant can consider their choice given the new information). The assumption of independence for each of the doors no longer holds.

Extension questions**Conditional probability and Bayes' Theorem****Question 8**

The definition of conditional probability for events A and B , where $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$ may be expressed as $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$ or $\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$.

Use the above equations to show that $\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$. This is known as Bayes' theorem.

Question 9

Assume the following.

- The contestant chooses door 1.
- The host then opens door 2 to reveal **no** prize.
- Event A is 'the prize is behind door 1'
- Event B is 'the host opens door 2 to show **no** prize'

- a. Determine the following for events A and B .

i. $\Pr(A)$

Answer. $\frac{1}{3}$

ii. $\Pr(B|A)$ (i.e. $\Pr(\text{host opens door 2 ... given prize is behind door 1})$)

Answer. $\frac{1}{2}$ (host opens one of either door 2 or door 3)

- b. Use the fact that there are 3 doors and that the host will **not** reveal the prize and will **not** open the door chosen by the contestant to explain why

$$\Pr(B) = \left(\frac{1}{3} \times \frac{1}{2}\right) + \left(\frac{1}{3} \times 0\right) + \left(\frac{1}{3} \times 1\right)$$

Answer. This is the likelihood of the host opening door 2 considering the possibilities of the prize being behind door 1 or door 2 or door 3.

- c. Use Bayes' theorem to determine the probability that the prize is behind door 1 given that the host opens door 2. That is, $\Pr(A|B)$.

Answer. $\frac{1}{3}$

- d. Hence determine the probability that the contestant wins the prize if they stay with the initial choice (door 1) and the probability of winning the prize if they switch to the other closed door.

Answer. stay $\frac{1}{3}$, switch $1 - \frac{1}{3} = \frac{2}{3}$.