Workshop 3.1. Monte Carlo Simulation

Part 1

Cell Phone Experiment

1. Open the spreadsheet labeled "cell phone plan.xlsm" You'll find yourself on the best estimate tab.

2. You're trying to choose one of 3 cell phone plans for your family, based on the number of minutes you expect to use each month. The 3 plans are described, including how many minutes are included, the cost of the basic minutes, and then the cost per minute of exceeding your allowance.

3. On studying your family's past cell phone usage, you find that your average usage is 375 minutes per month. The spreadsheet shows that the least expensive option is plan 1, which allows 400 minutes.

What do you think of this estimate of future minutes used? Is this a good analysis of which plan to choose?

4. Go to the tab labeled "100 trials" to see a Monte Carlo Analysis of the decision problem. The blue cells (C13:C14) show the mean and standard deviation of the number of minutes used in the past, with mean = 375 minutes and standard deviation = 250 minutes. The distribution is LogNormal (meaning it can go much higher above the mean than it can go below than the mean.)

5. Below, there are 100 trials of the decision problem, with minutes randomly chosen from the defined LogNormal distribution, and the cost of each plan (including overages) computed. The best (cheapest) plan is chosen for each trial, to the right. In red, the mean and standard deviation of the cost of each plan over the 100 trials is computed, as well as the minimum and maximum cost, and the number of times each plan is the best choice is computed.

Based on the 100 trials of the specified distribution of minutes, which is the best plan to choose? With what information did you make that choice?

6. Hit F9 a few times, to choose different random numbers, and see if the results change. Experiment with changing the standard deviation of monthly minutes in the blue cells. Look for thresholds at which the choice of best plan changes.

7. This experiment has shown that when an input is uncertain, it is often not adequate to use the best estimate in your computation.

Save and close the spreadsheet.

Game Show Experiment

8. This Monte Carlo analysis example attempts to answer the question posed by a game show. Imagine there are 3 doors shown to you, and one of them has a prize behind it. You are asked to choose a door. After your choice, the host of the game show opens another door, and shows you that there is no prize behind it. Then, you are given another choice: you may stay with your original choice, or you may switch your choice to the unopened door. Which is the better choice?

9. Open spreadsheet labeled "3doors exercise.xlsm," and you'll be in the tab titled "3 doors"

10. The player's choice of door is entered in the blue cell (B7). A random U[0,1] number is generated in the yellow cell (D7), and that value is used to sample from the probability distribution of which door should have the prize. We are assuming each door is equally likely, and so has probability 1/3, or 0.333. Next we see the door shown by the host (J7). Finally, both possible strategies are depicted: strategy 1 = don't switch doors (L7:M7), and strategy 2 = switch doors (O7:P7).

11. Take some time to study how the spreadsheet computes these values. Note that the "door shown by host" equation looks complicated, but the logic is that if the player has chosen wrong, the host shows the only remaining door that *doesn't* contain the prize, and if the player has chosen right, the host chooses one of the other 2 doors (door 1 if possible.)

12. Hit F9 a few times to see a few "plays" of the game.

Does the strategy to switch or not switch appear to be more successful?

13. It is actually somewhat difficult to tell which is better, seeing 1 outcome at a time. Move to the tab labeled "200 trials" to see more outcomes. One this page, the game has been repeated 200 times. The user choice is always 2, but this can be changed, or randomized, if you choose. Note the red percentage values at the top of the columns. The values above the door1, door2, and door3 columns show us how often each door was chosen. We expect those percentages to each be 33.3%. The percentages above the strategy columns show us how often those strategies won, in 200 trials.

Do you have a better idea of which strategy (to switch or not switch) is more successful?

14. This example has shown how repeated sampling of a random value, and follow-through of the computation based on that value, can help answer a question.

Save and close the spreadsheet.

IF YOU HAVE TIME: Roulette Experiment

15. Roulette involves spinning a wheel that has 18 red numbers, 18 black numbers, and 2 green zeroes, for a total of 38 possible outcomes. We'd like to test a betting strategy that bets on red (paying double the bet if you win) repeats the bet if you win, doubles the bet if you lose, and then goes back to the original bet after winning again. With this strategy, you will either achieve some target total or go broke before you get there.

16. Open spreadsheet "roulette exercise.xlsm" and find yourself on tab "game"

17. Note the random number generation that defines each spin, and specifies whether it comes to red, black or green. Note the details of the strategy at the top. Under label "strategy" is the multiple of the next bet if you lose. "Bet" for the bet size, "start" for the initial money, and "goal" for the target amount to win. The yellow area summarizes the results of a single trial.

18. Hit F9 a few times to see results of different repetitions of the same betting strategy. Adjust the strategy if you want.

19. Run the macro to look at the results of 500 replicates. To do this, go to the "output" tab and put your cursor on the blue cell (B3). Go to the Developer menu block and click the Macros button on the left. Run the macro called "MonteRoulette" and wait for it to finish.

What percent of time did your strategy win?

Save and close the spreadsheet.

Part 2: How many replicates?

20. A Monte Carlo experiment can have 5 replicates or 5 million replicates. 5 million replicates would take quite a long time… How many is enough? We know we have enough when the answer stops changing (or stops changing outside of a defined interval.) Mainly, we want to reduce the "error" in the result.

3 Doors Experiment

21. Go to the 3 doors experiment, in spreadsheet "3doors exercise.xlsm" and choose tab "explore trials"

22. Recall the red values in the top row which, in columns E, F and G, show the sample distribution of the input variable, and in columns M and P show the percentage of time each strategy wins. If we have enough replicates, 2 things will happen. (1) The sample distribution will be correct, showing 1/3 or 33.3% for each door. And (2), the percentage win for each strategy will ***stop changing*** when you hit F9.

23. Copying a row downward adds another replicate to the Monte Carlo Experiment. Add a few rows, and check the sample distribution cells. Hit F9 a few times, and see if the percentage win values change.

24. See how many rows you need to add to make the sample distribution cells correct. See how many rows you need to add to make the percentage win values stop changing.

How many rows did you need?

Save and close the spreadsheet.

Cell Phone Experiment

25. Go to the cell phone spreadsheet, "cell phone plan.xlsm" and go back to the "100 trials" tab. You'll be running a macro, which is done by going to the Developer menu block and hitting the Macros button on the left, and choosing a macro. Run the macro titled "cellMonte100" which will hit F9 and copy the results to tab "output100" 500 times. **It will take a few minutes to complete.** Go to tab "1000 trials" and hit F9 a few times, to observe the results with 1000 trials rather than 100. Run the macro titled "cellMonte1000" which will hit F9 and copy the results to tab "output1000" 500 times. Now compare tab "output100" to tab "output1000." We are interested in the grand mean and the standard error for the cost of each of the plans, which are the values in bold.

How does the grand mean of the cost of each plan compare between 100 and 1000 replicates?

How does the standard error of the cost of each plan (shown below the grand mean) compare between 100 and 1000 replicates?

How do these values inform you about the accuracy of your answers and adequate sample size?

26. Normally, we wouldn't do 500 trials of a 100 member sample, when instead we could just do a 50,000 member sample. The reason for this experiment was to show us how much the estimate of the mean cost changes (standard error) for samples of 100 members, compared to how it changes for samples of 1000 members. Clearly, with 1000 members, the values change less between trials, showing use that the estimates of the mean are more accurate (closer to what we'd get with 5 million replicates).

Save the spreadsheet and close.