Worksheet 08

1. (Statistical Inference) In pairs (or triples), select six marbles of one color (we will call this color C) and four marbles of another color. In one of the bags (we will call this bag F, for the first bag), place 4 C and 1 non-C marbles. In the other bag, place 2 C and 3 non-C marbles. We are going to consider an experiment where you select one bag at random and then randomly select a marble from the choosen bag. The idea is that we want to see how well you can estimate which bag the marble came from based on the color of the marble. (a) Compute the two-by-two table of probabilities where C is the event of selecting a marble of color C and F is the event of selecting the first bag. (b) What are the probabilities $\mathbb{P}(F|C)$ and $\mathbb{P}(F^c|C)$? (c) What are the probabilities $\mathbb{P}(F|C^c)$ and $\mathbb{P}(F^c|C^c)$? (d) What is the best guess for a bag if you have a C marble and what is the best guess if you have a non-C marble? (e) Let R be the event that your guess of the correct bag is right. What is $\mathbb{P}(R)$?

(f) Now, we are going to simulate the game 12 times.[1] To do this, have one person close their eyes. The other person rolls the die. If odd, they select the first bag and hand it to the other person, who then selects a marble. If even, the pick the second bag and give that instead. Keep track of the number of correct guesses. Switch roles mid-way through the simulation. We will aggregate across the class and see if we can get close to the analytical answer.

[1] It may seem silly to go through this exercise, but I find it really helpful to have the perspective of the guesser in which you are updating your understanding of the bag probabilties with the data.

2. (Simpson's Paradox) There are two physicians named Dr. A and Dr. Z. Each of them performs two types of procedures: band-aid removal (B) and heart surgery (H). Their recent performance is given by the following tables:

	Heart	Band-Aid
Success	70	10
Failure	20	0

Dr. A

	Heart	Band-Aid
Success	2	81
Failure	8	9

Dr. Z

For this question, we will use the *empirical probability* of each event. That is, the value of every probability $\mathbb{P}E$ is given by the proportion of procedures from the data for which E occurs. (a) Compute the probability that a procedure done by Dr. A is successful and the probability that a procedure done by Dr. Z is successful. Who seems to be the better physician? (b) Compute the probabilities that each procedure is successful, conditioned on the doctor doing the procedure and which procedure is being done. Who seems to be the better physician now? (c) What paradox seems to exist? Can you explain why this happens?

3. (Monty Hall) This is a probability question so well-known my guess is that most of you have already heard of it. But let's see how we can answer it in a formal, systematic way. There are three doors, one randomly has a car behind it and the other two have goats. A contestant is playing a game in which they want to win the car. In the first round, they stand in front of a door that the are thinking of picking. The host of the game, Monty Hall, selects one of the other doors that he knows has a goat behind it and opens it for everyone to see. The contestant now has to pick the door they actually want to open and then open it. What is the probability that they will win if they switch their choice from the first selection? Hint: We can assume that the contestant selects door 1.[2] Let C_1, C_2, and C_3 be the events that the car is behind door 1, 2, and 3, respectively and W be the event that the contestant wins if they switch their choice.

[2] If not, relabel the doors so their choice is called door 1.

4. (Monty Hall, revisited) Consider a variation of the previous problem where there are seven doors, all equally likely to have the prize. In the second round, Monty Hall randomly selects three goat doors that are you not in front of to open. There is now a closed door you are in front of and three other remaining doors. What is the probability that you will win if you switch to one of the other three doors?

5. (Hard!/Fun?) Consider an airplane with 100 seats assigned to each of 100 passengers. The first person to board has had too much to drink and selects their seat at random. Everyone else sits in their assigned seat unless it is already occupied, in which case they select a seat at random from the remaining empty seats. What is the probability that the last person to board will sit in their own seat?[3]

[3] Try to do this with a much smaller number of passengers and see if you can find a pattern. This question is better to do with some basic logic rather than formal probability manipulations. I include it here because it fits the general theme of the other questions.

6. (Prosecutor's Falacy) There are 10000 people living in a remote rural town. One night, a chicken is stolen from the town barn. The person who stole the chicken accidentally cut themselves on some barb wire escaping the scene, leaving just enough evidence to determine that the blood type of the thief is B positive. The next morning, a man is arrested for the crime who has B positive blood type. Based on knowledge that this type of blood (B positive) is only present in 8.5% of people in the US (and this town in particular), the prosectutor for the case argues that there is a 91.5% chance that arrested man commited the crime. (a) Write out the problem using some probabilistic notation. (b) Find a better measurement of the man's guilt, given that there is no other evidence against him.