Real World Mini-Project 2: Modeling Assumptions

Due Date: This assignment is due at Monday May 17 at 11:59 PM (Seattle time, i.e. GMT-7).
You will submit as a PDF to gradescope.

The tools of this class are useful to computer scientists, but many of them are useful beyond just "classic" computer science. In order to use the powerful tools of probability, we need to make assumptions to let our mathematical tools model the real world. Things like "this coin is perfectly fair" or "the coin flips are all independent." These are usually not perfectly true[1] Indeed, occasionally these assumptions are ways that people "lie with statistics" or provide evidence for claims that aren't actually true.

In this project you will critique the modelling assumptions made in an analysis and see if other modelling assumptions would lead to a different result.

1.   Find a Model

Find an analysis that uses probability and statistics tools you're familiar with from this course. By "analysis," we mean any estimate of a "real-world" probability, along with the assumptions that lead to that number. You might want to look at the examples in the final section for what we mean.

We expect most of the answers to this section will be short (2-3 sentences), but you are free to write more if your resource is more complicated.

  (a) Provide a link to (or somehow let us access) the analysis you're critiquing.

  (b) What is the fundamental claim of the analysis? I.e., what conclusion do they draw at the end of their analysis?

  (c) What modelling assumptions do they use? (For example, do they assume some occurences are independent? Do they assume a set of events all have equal probability? Do they assume they know the probability? Do they use a variable from the zoo?)

2.   Improve the model

Now, see if their modelling assumptions are reasonable or if other ones would lead to a different conclusion. We expect parts a,b,d will be a few sentences each (though you can write more if you have more to say).

  (a) Identify at least one weakness of the modelling assumptions they have made (e.g. a potential dependence on events that are supposed to be independent).

  (b) Now create your own model for the same problem. You might do this by coming up with different probability estimates for the events, or by using a different random variable (e.g. a binomial distribution instead of a Poisson), or by incorporating some outside knowledge about the problem that you think sheds more light. Briefly describe what your model will be, and how it differs from the previous one.

  (c) Under your new model, calculate the probability of the event your source calculated.

  (d) Does the calculation change significantly? If it does, does the conclusion of the analysis change?

---

[1]e.g., if you flip a coin repeatedly, the result of the last flip is probably how the coin will appear on your hand before you flip it, which will make the results not quite independent.

3. Some Ideas

You are free (and encouraged!) to find your own examples outside this list if you have a topic you are passionate about, but if you can't think of anything, you may use any of these as starting points. In many cases, there are already critiques of poor statistical/probability analyses online – it's ok to look at these critiques, as long as you tell us if you're using any and still do the new probability calculation independently and put everything in your own words.

- Every year millions of people predict the outcomes of the NCAA men's basketball tournament. It is commonly said that the probability of a perfect bracket is $\frac{1}{2^{63}}$, (since there are 2^{63} ways the 63 games could play out) and therefore no one will ever predict a perfect bracket. (Yes, Robbie has a blog post about this one) Here is a source using that number

- Shortly after the 2020 presidential election, there were many assertions (including by people with PhD's...) that the probability of the election night shift was so low as to be impossible. Page 20 here is one example.

- A video-game streamer named "Dream" did a speedrun of Minecraft where they had incredible luck in a few parts of the game. So lucky, that a speedrunning organization declared that Dream had to be using a modified version of the game, and that the run was therefore invalid. The analysis that lead to the rejection of the run.