Select Reading Question Responses (9/2/2021)

> In a more broad sense, how is probability different from statistics, and how do the two work together to process and convey data? How is this course divided between the two disciplines, and how important will both be when looking forward in this class?

The way I see it, statistics is about analyzing data, and probability provides the theoretical/mathematical underpinning for statistics. We'll only study a bare minimum of probability

> Regarding 3.7(f), how do you know if the two variables are independent or not?

> I'm currently struggling with 3.7, particularly part (f).

Let I be the event that someone is independent and S the event that someone is a swing voter. What we know is that $P(I) = 0.35$ and $P(S) = 0.23$ and $P(I \text{ and } S) = 0.11$. By definition, S and I are said to be independent if $P(I \text{ and } S) = P(I)P(S)$, but

$$P(I)P(S) = 0.35 \cdot 0.23 = 0.0805 \neq 0.11 = P(I \text{ and } S).$$

Thus the events are *not* independent.

> I had a difficult time choosing the proper formula for problem 3.13(b), part (ii). I wanted to do $P(A \text{ or } B) = P(A) + P(B)$, which translates into 0.3+0.7, but the answer key in the textbook says it should be 0.79.

> Getting mixed up on how to solve $P(A \text{ or } B)$. Help please.

In general, we have

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

so in 3.13(b), part (ii), we have

$$P(A \text{ or } B) = 0.3 + 0.7 - P(A \text{ and } B) = 1 - 0.21 = 0.79$$

where the 0.21 comes from your calculation from part (i) of the same problem.

There's a "special case" of the above formula which happens when A and B are disjoint, which just means that $P(A \text{ and } B) = 0$. In this special case, the formula above reduces to $P(A \text{ or } B) = P(A) + P(B)$. But I would not recommend remembering this separately; you should just know and understand the general formula that we used above since that can be used even when A and B happen to be disjoint.

> Clarification on 3.13(d) please

By definition, we have

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}.$$

We know that $P(A \text{ and } B) = 0.1$ and $P(B) = 0.7$, so $P(A \mid B) = 0.1/0.7 \approx 0.143$.

> A question I have today is more of a logic question. When calculating the probability for A or B, the probability for A and B (the overlap in the Venn diagram) is also counted in the final result. This doesn't make that much sense to me. To take a real life example, suppose a mom tells her kid that he can buy either a car or a superman. That doesn't mean he can choose to buy both! In math, why does this area of A and B get counted in the probability of A or B?

This is a great question and an astute observation. The word "or" has two senses: the "inclusive or" which includes the overlap, or the "exclusive or" which excludes the overlap. In math, "or" means the inclusive or; when someone wants to use an exclusive or, they'll explicitly specify this by saying something like "but not both." This does in fact differ from the day-to-day usage of "or," which is often an exclusive or like in the example you give.

I don't know *why* there's this discrepancy between the mathematical usage of "or" and the day-to-day usage. It is a little unfortunate and confusing, but it's quite well established by this point so there's no option but to learn to put up with it.

You might like taking a look at this discussion on Wikipedia about the exclusive or in natural language.

> At first I thought it was strange that drawing an ace or a heart were independent because they seem connected in that if you draw one you won't get the other. But I guess that is the point– they have no overlap. No ability to be together in a single outcome.

There is an ace of hearts. In other words, there is one card in the deck that's both an ace and a heart. They're independent, if we let A represent drawing a heart and B represent drawing an ace, then $P(A \text{ and } B) = 1/52$ since there's just one ace of hearts in a deck, and $P(A)P(B) = (1/4)(1/13) = 1/52$ as well.

Actually, a piece of your original intuition is correct — two events that cannot both happen must be dependent! More precisely, if A and B are both possible but A and B is impossible, then A and B are dependent. For example, drawing a heart and drawing a spade are both possible (both have a $1/4$ chance). But drawing both is impossible: the joint probability is 0. If A represents drawing a heart and B represents drawing a spade, we have $P(A \text{ and } B) = 0$ but $P(A)P(B) = (1/4)(1/4) = 1/16 \neq 0$, so the two events are dependent.

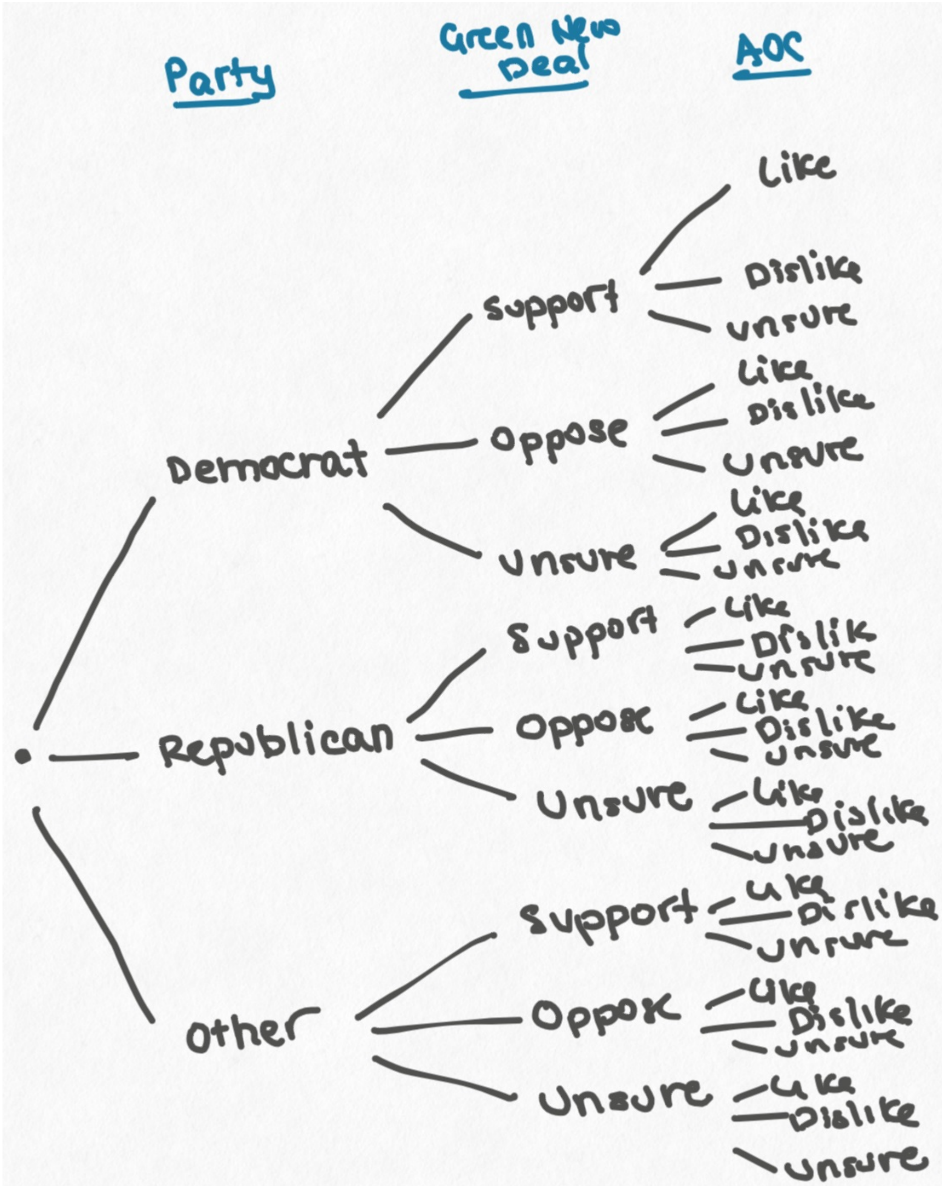
> From my perspective it seems that Bayes' Theorem is just a faster way to calculate than the tree diagrams. However, I find the tree diagrams much easier to use and understand. Is this view incorrect? Do the two approaches accomplish different things and would I sometimes need to use Bayes' theorem instead of the tree diagrams?

> Why would you ever use a tree diagram instead of Bayes' theorem?

Some people find Bayes' theorem easier, and some people find tree diagrams easier. There's nothing "incorrect" about either view. Anything you do with Bayes' theorem, you could theoretically do with tree diagrams as well, and conversely. In fact, tree diagrams are really just a way of "visualizing" Bayes' theorem!

> Do tree diagrams always contain only two branches (primary and secondary)? Or is it possible that more complex tree diagrams with tertiary branches will come up?

You can have tree diagrams with an arbitrary amount of branching at each level, and also an arbitrary number of levels! For example, let's say I was doing a study of the political affiliations and opinions of many Americans. More specifically, let's say I ask everyone what political party they belong to (Democrat, Republican, Other), what they think about a having a Green New Deal (Support, Oppose, Unsure), and what they think about AOC (Like, Dislike, Unsure). In that case, I'd have a tree diagram with three levels, and ternary branching at each level.



Party
Green New Deal
AOC
Democrat
Support
Like
Dislike
Unsure
Oppose
Like
Dislike
Unsure
Unsure
Like
Dislike
Unsure
Republican
Support
Like
Dislike
Unsure
Oppose
Like
Dislike
Unsure
Unsure
Like
Dislike
Unsure
Other
Support
Like
Dislike
Unsure
Oppose
Like
Dislike
Unsure
Unsure
Like
Dislike
Unsure