PSC504: Potential Outcomes

Matthew Blackwell

1/8/2013

Association versus Causation

What is association?

- Let's take two variables, Y_i and A_i. These variables obviously have a joint distribution: $\Pr[Y, A]$. We generally say that these two variables are **independent** if one does not predict the other. To fix ideas, let's have Y be whether or not a Democrat won a Senate race and A be an indicator for whether or not the Democrat went negative during the campaign.
- We can write this with conditional probability: $\Pr[Y = 1|A = 1] = \Pr[Y = 1|A = 0]$. That is, knowing what the value of A is doesn't affect the distribution of Y. We write independence (following Dawid) as $Y \perp\!\!\!\perp A$.
- If the variables are not independent, we say they are dependent or associated: $\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$.
- Associations between variables, very famously, are not necessarily due to causation.

What is causation? Counterfactuals, etc.

- Causation is a complicated concept, even if it appears intuitive. Like any such concept, there has been reams and reams written about it from both a statistical and a philosophical perspective.
- In this class, we're going to predominantly use a counterfactual or potential outcome approach to causal inference. There are other approaches (graphical, probabilistic, necessary/sufficient), but they either have issues with non-deterministic relationships or they end up having the same assumptional bite, so to speak. Our approach to defining causality will not help us avoid difficult assumptions.
- The counterfactual approach is useful for a couple of practical reasons: 1. other people use it. 2. it helps fix ideas and gives us some powerful intuition. Perhaps not great reasons, but hey.
- Our primitives will be what Rubin calls "potential outcomes" (others call them "counterfactual outcomes"). $Y_i(a = 1)$ is value that Y would take if the Democrat went negative. $Y_i(a = 0)$ is the outcome when the Democrat stays positive.
- We call Y the outcome variable and we'll often call A the treatment variable, even if we are not dealing with a randomized control trial or even binary variable.

- For each unit, we observe one of these two possible potential outcomes. We can never observe both of the potential outcomes for the same unit. This is called **the fundamental problem of causal inference**. At a basic level, we cannot observe your job market performance if you take this class *and* your job market performance if you don't take this class. Now, you might say that you could not take this class now, go on the job market, then the next year take this class and go on the job market again (this is called a cross-over study). Don't we observe both potential outcomes for the same unit then? No, because you are not the same unit in both years. At a basic level, you are a year older/wiser than before. We have to make strong assumptions in order to get around this fundamental problem.
- An assumption of temporal stability and causal transience could help us get around FPCI, but these are strong assumptions, especially for the social sciences. Lightbulbs are rare. Lab experiments tend to use unit homogeneity to get around this problem. All of these are assumptions. though.

Consistency/SUTVA

- We need some way of connecting these potential outcomes to the observed outcomes. We will do this with a consistency assumption. This is what epidemiologists call it. Economists and statisticians call the "stable unit treatment value assumption".
- $Y_i(a) = Y_i$ if $A_i = a$.
- Two key points here: no interference, between units. My potential outcome does not depend on other's treatment. Think about the Senate elections: this implies that other democratic senators going negative doesn't affect my potential outcome. This can be very wrong: time, equilibrium effects.
- Ill defined counterfactuals/multiple versions of the treatment. time

Key Questions:

- What experiment do we want to emulate?

Estimands

- Suppose there are a population of units, $i = 1, \ldots, N$ and there is a set $\mathcal{T}_i$ of treatments that a
- Individual causal effect (ICE): $\tau_i = Y_i(1) - Y_i(0)$
- Average treatment effect (ATE): $\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N} Y_i(1) - Y_i(0)$. This is the difference between the average outcome in a world where all Democrats go negative and a world in which all Democrats stay positive.
- Average treatment effect (ATE) for a subpopulation: $\tau_{ATE,f} = \frac{1}{N_f}\sum_{i:X_i=f} Y_i(1) - Y_i(0)$, where N_f is the number of units in the subpopulation.
- Average treatment effect on the treated (ATT): $\tau_{ATE,f} = \frac{1}{N_t}\sum_{i:A_i=1} Y_i(1) - Y_i(0)$.
- We might be interested in the ATT for a couple of reasons. First, and least compelling, sometimes our methods choose it. Matching methods generally estimate an ATT. Second, the ATT requires fewer assumptions to identify than the ATE. Last, we might be interested for scientific reasons: we want to estimate the effect of a policy for those states that actually adopt the policy.

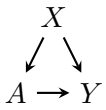
- Last, note that we can also define estimands in terms of the potential outcomes. Let's say that Y is the Democratic share of the two-party vote. Then we might be interested in the effect of negativity for Democrats that would win whether or not they went negative: $i : Y_i(0) > 50, Y_i(1) > 50$. Denote the number of such candidates N_w. Here is the effect we might be interested in: $\tau_{ATE,w} = \frac{1}{N_w} \sum_{i:Y_i(0)>50,Y_i(1)} Y_i(1) - Y_i(0)$. You can probably see that we can't directly observe this subgroup. You might wonder why we would want to define such a parameter. All will be reveals when we get to instrumental variables.

- Note that the potential outcomes here are pre-treatment variables: they are fixed attributes of the units.

Graphical causal models

DAGs and some graph theory

- We can encode assumptions about causal relationships in what are called causal Directed Acyclic Graphs or DAGs. Here is an example:



- Each arrow represents the presence of a direct causal effect (that is, an individual causal effect as above). The lack of an arrow represents the lack of a causal effect.

- These are **directed** because each arrow implies a direction (aspirin causes pain relief, not the other way around). They are **acyclic** because there are no cycles: a variable cannot cause itself, either directly or through cycles. We have r.v.s $V = (V_1, \ldots, V_M)$ which directed edges and no cycles. PA_m are the set of **parents** of V_m: these are the nodes that have a direct arrow into V_m. In the above examples, $M = 3$ and we have $V_1 = X, V_2 = A$ and $V_3 = Y$, then $PA_3 = (L, A)$.

- Causal Markov assumption: conditional on its direct causes, a variable V_j is independent of its non-descendents: $\Pr(V_j|V_1, \ldots, V_{j-1}) = \Pr(V_j|PA_j)$. This allows us to factorize the joint distribution of the data using a Markov structure:

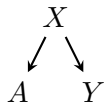
$$f(v) = \prod_{j=1}^{M} f(v_j|pa_j)$$

- Note that for a graph to be causal, all common causes, measured or unmeasured, of any pair of variables in the graph must also be included in the graph.

Causal DAGs and associations.

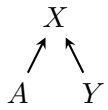
- DAGs are a convenient way to encode causal assumptions about the problem at hand, but they also can tell us about potential associations between variables in the graph. This is what makes them extremely useful.

- A **path** between two variables (C and D) in a DAG is a route that connects the variables following nonintersecting edges. A path is causal if those edges all have their arrows pointed in the same direction. Otherwise it is noncausal. Here is an example of a noncausal path between A and Y (a classic example here is the relationship between drowning deaths and ice cream sales):



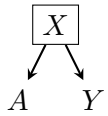
- Two variables connected by common causes will have a marginal associational relationship. That is, in the above example $\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$. There is a correlation. Otherwise, there will be no association between them.

- Let's look at another situation:



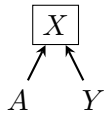
Here, X is a **collider**: a node that two arrows point into. What happens here? Are A and Y related? No. Imagine that A is getting the flu and Y is getting hit by a bus. Both of these might cause us to be in the hospital, but knowing that I have the flu doesn't give me any information about whether or not I've been hit by a bus. The flow of association is blocked by a collider.

- Above we have shown how marginal associations flow over paths, but what about relationships between variables within levels of a third variable? We can represent conditioning on a variable by drawing a box around it.



Conditioning on a variable is on a causal path or on a variable that is a common cause (above), will block the association that flows over that path.

- Conditioning on a collider (a common consequence) actually opens the flow of association over that path, even though before there was none:



To see why this is the case, let's go back to the flu, getting hit by a bus example. In this case the common consequence was being in the hospital. If we condition on this by, say, looking only at people who are in the hospital, we will induce a negative relationship between getting the flu and getting hit by a bus. If I know that a person in the hospital for the flu, it is less likely that they were hit by a bus: $\Pr[Y = 1|A = 0, X = 1] \neq \Pr[Y = 1|A = 1, X = 1]$. This is even though there is no effect of getting the flu on getting hit by a bus. This also comes up with different criteria for getting into college/grad school.

- To sum up: associations flow over paths (causal or noncausal) that don't contain a collider. These associations can be blocked by conditioning a variable on the path that is not a collider. We'll come back to these properties later when we talk about the back-door criteria.