Qualitative strategies

Macartan Humphreys

Qualitative estimands

- A common qualitative estimand is the "cause of an effect": given $Y = 1$ and $X = 1$ is $Y = 1$ because $X = 1$?
- Randomization (alone) does not justify answers to "causes of effects" questions
- We might be able to say that we think that the effect of a treatment on a given outcome is $1/3$
- But that does not justify claiming that the probability that the outcome is due to the treatment is $1/3$

Qualitative estimands

Here is the logic:

- Say I know that (binary) X increases Y from $\frac{1}{3}$ to $\frac{2}{3}$ on average.
- Say I observe $Y = 1$. What are the chances it is due to X?
    - One possibility is that X has a positive effect for $2/3$ of cases and a negative effect for $1/3$ of cases. In that case whenever $X = Y = 1$ this is due to X.
    - Another is that X never has a negative effect and it has a positive effect for $1/3$ of cases and no effect on the rest. Then there's a 50% chance that $Y = 1$ is due to $X = 1$ (why 50%?)

Learning from clues

- Classically qualitative strategies use *auxiliary* information to understand the relationship between X and Y.
- For instance: You want to know if the swamp caused malaria
  - Quantitative approach: Compare malaria incidence in places with and without swamps
  - Qualitative approach: Look to see whether there mosquitoes are breeding at the swamp
- This approach typically requires a theory: conditional on theory T, observing K increases my confidence that $X = 1$ caused $Y = 1$ in this case.

Process tracing

- In process tracing such clues are called "CPO"–causal process observations
- How informative a clue is is sometimes called its "probative value"
  - sometimes you learn a lot when you *do* see a clue ($K = 1$)
  - sometimes you learn a lot when you do *not* see a clue despite looking for it ($K = 0$)
- Classic tests:
  - A "smoking gun" clue gives a lot of confidence when you find it
  - A "hoop" clue gives a lot of confidence when you do not find it

Qualitative tests: b or d?

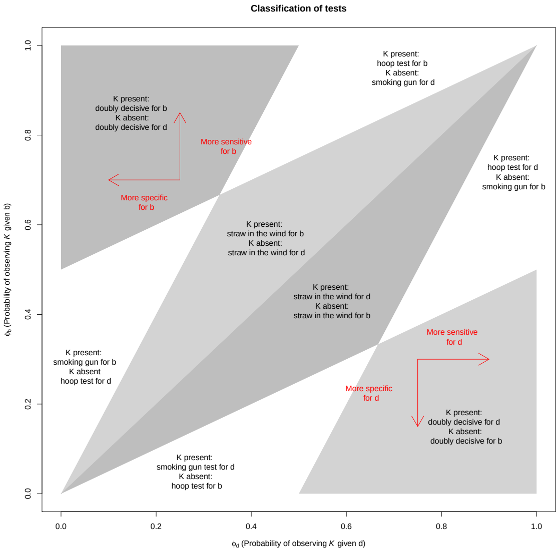

Figure 1: Mapping from the probability of observing a clue if b (ϕ_b) or d (ϕ_d) to a generalization of the Van-Evera tests.

How might beliefs about probative value be supported?

- Most common answer: "theory"
- Better answer: background knowledge + specific assumptions

For instance:

- Say we know in a given population that:
  - $X \rightarrow Y \leftarrow K$
  - $\Pr(Y = 1|X = 0, K = 0) = 1$
  - $\Pr(Y = 1|X = 1, K = 0) = .5$
  - $\Pr(Y = 1|X = 0, K = 1) = 0$
  - $\Pr(Y = 1|X = 1, K = 1) = .5$
- The unit we care about is "exchangeable" with other units in this population

Then for an $X = Y = 1$ case:

- seeing $K = 1 \rightarrow X$ caused Y (why?)
- seeing $K = 0 \rightarrow X$ did not cause Y (why?)

How might beliefs about probative value be supported?

- Thus, doubly decisive clues are possible
- But likely rare: most often best you can do is put bounds on causal effects

Case selection: What do you do with a case?

- Establish that X is indeed X and Y is indeed Y
- Assess whether scope conditions of theory are indeed present
- Assess whether the argument has "face validity"
- Ideally look for pre-specified clues that support or weaken the claim

Case selection

- Case selection depends on the estimand. Are you interested in a case level estimand or a population level estimand?

Common strategies:

1. On the regression line
2. Off the regression line
3. Most likely cases, least likely cases
4. Proportionate to distribution (safe rule of thumb)

less common but good:

5. Follow the probative value:
    - Perhaps K is informative in one case not another
    - For instance in the example above there is no point selecting a $X = Y = 0$ case since you already know that in that case $K = 1$: learning about K will not be informative

Case selection: n

How many cases?

- No good answer
- More always better except insofar as they reduce quality of analysis
- If you are doing causal inference with case comparison methods only then you want as many as possible and *at least* as many as you have explanations that you want to distinguish from each other

Mixed methods [advanced]

Insight:

- If observation of X and Y lets you update your beliefs about a causal effect
- And if observation of K also lets you update your beliefs about a causal effect
- Then you can update jointly from X, Y, K

$$Pr(H|X, Y, K) = \frac{Pr(X, Y, K|H)Pr(H)}{Pr(X, Y, K)}$$