Notes on Rubin 2005: Causal Inference Using Potential Outcomes: Design, Modeling, Decisions

Gray Stanton and Austin Ellingworth

October 20, 2020

Discussion Questions

- Rubin points out that "causal inference is impossible without making assumptions." As an researcher, how do you determine if your assumptions are too weak, too strong, or just right?
- By incorporating prior information and making stronger modeling assumptions, the Bayesian framework for causal inference that Rubin outlines seems to be able to go beyond just making assumptions on the assignment mechanism. What are the advantages and disadvantages of this approach?
- As Rubin points out in his example about concomitant variables, even very smart people can be mislead by questions of causality. How does this manifest itself presently in research involving estimation of causal effects?

2. The Causal Estimand – "The Science"

- Importance of understanding precisely the quantity you are interested in – the causal estimand.
- Basic example: N units (particular objects, treatment receivers), covariates X which cannot be affected by treatment. Potential outcomes: $Y_i(1)$ and $Y_i(0)$ which is the value for the outcome variable in the universe where the ith unit received the active treatment versus the control treatment.
- The "Science" is contained in the $X, Y(1), Y(0)$.
- Unit level causal effects $Y_i(1) - Y_i(0)$.
- Fundamental problem of causal inference: Cannot observe both $Y_i(1)$ and $Y_i(0)$ because time cannot be unwound.
- Summarizing causal effects, e.g. mean unit level causal effect for some subset of the indices.
- Causal effect must be comparison of $\{Y_i(1), i \in S\}$ with $\{Y_i(0), i \in S\}$. Can't compare different sets of units.

- Important assumption: SUTVA (Stable Unit Treatment Value Assumption). Two parts: no interference between units i.e. $Y_i(1)$ and $Y_i(0)$ is not affected by the treatment assigned to the other units. No hidden treatment versions: active treatment on unit i always leads to $Y_i(1)$
- Third underlying assumption: the science (both in covariates and outcomes) are not affected by how or whether the subjects try to learn about about it.
- "Causal inference is impossible without making assumptions, they are the strand which links statistics to science".

3. Fisher and Neyman on the Potential Outcomes Notation in Randomized Experiments and Beyond

- Importance of potential outcome notation.
- Average causal effect, $\sum_{i=1}^{N} \frac{Y_i(1)-Y_i(0)}{n}$ is important causal estimand. Difference in observed treatment means is unbiased estimator of it and $\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}$ is a positively biased estimator of its variance (Neyman).
- Fisher Sharp Null Hypothesis: For each unit, the treatment does exactly nothing. That is, $Y_i(1) = Y_i(0)$ for all i. Sharp because under it, all potential outcomes are known for all units, regardless of the actual assignment.
- Fisher 1918 passage.
- Terminology alert: "Counterfactuals" versus "Potential Outcomes"

4. The Assignment Mechanism

- Assignment Mechanism: A method for assigning treatments to units, which creates missing potential outcomes.
- W_i is the assignment for unit i. The only random quantity in this setup.
- Assignment mechanism is an assignment of the conditional probability $P(W|X, Y(1), Y(0))$.
- $E(\bar{y}_1-\bar{y}_0|X, Y(1), Y(0)) = \bar{Y}(1)-\bar{Y}(0)$ and $V(\bar{y}_1-y_0) \geq E\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}|X, Y(1), Y(0)\right)$ (Neyman).
- Random experiments are a subset of assignment mechanisms, and under RE, it can be that SUTVA is sufficient assumption for causal inference.
- Random experiments are an ignorable non-trivial assignment mechanisms: Non-trivial in that every unit has some possibility of being assigned each treatment $0 < P(W_i = 1|X, Y_{obs}) < 1$. Ignorable in that potential outcomes don't affect assignment mechanism probabilities, just the observed outcomes: $P(W|X, Y(0), Y(1)) = P(W|X, Y_{obs})$.

- In sequential experiments, that's when dependence on Y_{obs} arises. Classical random experiments are unconfounded $P(W|X) = P(W|X, Y_{obs})$.
- Collapsing potential outcomes into just Y_{obs} mixes up science and what we to try and learn about science.
- Only a model on the assignment mechanism still allows for progress on statistical inference, even for observational studies (propensity scores).

5. Models on the Science

- However models on science can be have a critical role (Bayesian inference).
- Models only on assignment mechanism are more robust (but still require SUTVA), models on science allow for handling greater complexity and logical summaries of results.
- Using bayesian framework to create posterior predictive distribution for the missing half of the potential outcomes.
- Using $P(X, Y(1), Y(0))$ model for the science + $P(W|X, Y(0), Y(1))$ assignment mechanism, then can find PPD for $P(Y_{mis}|X, Y_{obs}, W) \propto P(X, Y(1), Y(0))P(W|X, Y_{obs})$
- Can calculate distribution of any causal estimand, which is function of $X, Y(1), Y(0)$.
- Randomized Experiments have impact on bayesian inference: Any distribution on the science will be exchangeable. Ignorable means that second term is a constant.
- "Rubin's Causal Model" – Extends potential outocomes to all situations, explicitly includes assignment mechanisms with possible depenedence on all potential outcomes, embeds assignment based + bayesian likelihood in common framework.

6. Decisions: Based on current knowledge of science and on costs of decisions

- Posterior distribution of causal estimands is "Summary of current knowledge of the science" from current data and past science (prior dist).
- Fisher perspectieve: unknown to what purpose discoveries will be put, so should not introduce cost functions.
- Likelihood function vs signficance tests vs accept/reject tests.

7 Complex Experiments: "Direct" and "Indirect" Causal Effects

- Concomitant random variable: Outcome variable not of primary interest, but is on causal pathway from treatment to primary outcome. Not a covariate, but may want to adjust for it.
- Fisher suggestion: Analyze Y_{obs} via W and the concomitant C_{obs} with ANCOVA, but this is equivalent to regressing Y_{obs} on W, C_{obs}, which he called naive earlier. Compares $Y_i(1)$ with $Y_i(0)$ for those with common level of C_{obs} which is not a causal effect because C_{obs} is affected by treatment.
- Two thought experiments: When treatment impacts Cbut not Y i.e. no direct impact of treatment on outcome after adjusting for C, but if we condition on C_{obs}, actually appears that treatment plots do worse. Problem still arises even when there is direct treatment effect after adjusting for C.
- Controlling for C_{obs} essentially breaks ignorability of treatment assignments.
- Fisher's ANCOVA is predicated on ignorability of assignment bc assumes set of subjects with fixed C_{obs} are randomly assigned treatment/control, which is not the case.
- Rubin posits that combining Y and C into 1 variable (i.e. Y / C) ior treating (C,Y) as a bivariate outcome is the better route for co-contaminants.

Notes on Pearl 1995: Causal Diagrams for Empirical Research

Gray Stanton and Austin Ellingworth

October 20, 2020

Discussion Questions

- Pearl states that whether the "three rules" of the calculus of interventions are sufficient to derive all identifiable causal effects remains an open question. Is that still true? Do the given inference rules align with our intuition about causality?
- Pearl suggests that "standard probability theory" is too weak to "describe the precise experimental conditions that prevail in a given study". Do we agree with this statement? How does this align with Rubin's potential outcome notation?
- The title of the paper explicitly applies causal diagrams to empirical research. Is the relatively abstract "calculus of intervention" put to good use by researchers in the field? If so, where? If not, what obstacles to adoption are there?

1. Introduction

- Classic Cochran example of eelworm control via fumigants can be expressed/analyzed via DAGs
- Causal diagrams help to: 1. Explicitly encode causal assumptions underlying model. 2. Decide if assumptions are sufficient to consistently estimate target causal effect. 3. If yes, gives closed form expression for estimator, If no suggests what observations are necessary.
- Hollow circle indicates unobserved quantity, full circle indicates observed.
- Dashed arrows connect nodes where at least one quantity is not observed, full arrows if both observed.
- Causal assumptions about how one quantity affects another are encoded by present arrows + direction. Missing arrow indicates that one quantity cannot directly affect another.

- Check notation: $\check{x}$ signifies that random variable X is set to fixed value x by external intervention.

2. Graphical Models and the Manipulative Account of Causation

2.1 Graphs and conditional independence.

- DAGs useful for representing conditional independence assumptions, as can identify conditional independence restrictions that are implicit in factorization of joint probability:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pa_i)$$

  where pa_i is some subset of (X_1, X_{i-1}). If construct DAG where variables are in pa_i are represented by the parents of node X_i, then independences implicit in the above factorization can be recovered by d-separation test.

- D-separation: X, Y, Z are three disjoint node sets in a DAG, and p is an (undirected) path between a node in X and on in Y. Z is said to block p if either some node w on p has converging arrows along p and neither w or its descendent are in Z, or there is some node w on p without converging arrows along p and $w \in Z$.

- If for every p between X and Y, we have that Z blocks p, then Z d-separates X, Y: $(X \perp Y | Z)_G$.

- There is 1-1 correspondence between conditional independence assumptions, and triples of nodes that are d-separated.

- Alternative test for d-separation: Delete all nodes from G except those in X, Y, Z and their ancestors. If any two nodes share a common ancestor, connect with undirected arrow. Remove all direction from arrows. If Z is a cut-set of the graph which separates X, Y, then Z d-separates X, Y.

2.2 Graphs as Models of Intervations

- Non parametric structural equations: for each RV, have $X_i = f_i(pa_i, \epsilon_i)$ where f_i is some fixed function and all the ϵ_i are assumed mutually independent disturbances (exogenous factors). If any ϵ_i could plausibly affect multiple X_i, then it should be included in model as unobserved variable with those X_i as children.

- Correspondence between causal diagrams and potential outcomes: Read equality as "is determined by".

- Because each ϵ_i is independent of non-descendants of X_i, child-parent characterization as determinstic function instead of conditional probability leads to same independence constraints. But also provides language of how to specify resulting distribution changes with respect to external interventions.
- Atomic intervention $\text{Set}(X_i = x_i)$ changes functional mechanism determining X_i from $X_i = f_i(pa_i, \epsilon_i)$ to $X_i = x_i$ while keeping all other mechanisms the same.
- Causal effect: for disjoint sets of variables, X, Y, causal effect of X on Y is function from $\mathcal{X}$ to space of distributions on $\mathcal{Y}$. For each $x \in \mathcal{X}$, $P(y|\check{x}) = P(Y = y)$ under new collection of mechanisms which arises from deleting all equations for X_i in X and substituting values x_i into all other mechanisms.
- Graphically, $\text{Set}(X_i = x_i)$ amounts to removing lines from pa_i to X_i and keeping rest of network the same.

3. Controlling Confounding Bias

- Setup is: DAG G, observed variables V_0 from non-experimental context, want to estimate impact of $\text{Set}(X_i = x_i)$ on X_j: $P(x_j|\check{x}_i)$.
- Set of variables Z satisfies back-door criterion relative to X_i and X_j if: No node in Z is a descendent of X_i, and Z blocks every path between X_i and X_j which contains an arrow into X_i (i.e. the back door)
- Point of back-door is not every path between X_i, and X_j need be blocked. Only those which enter from the back-door.
- If Z satisfies back-door criterion for X, Y, then causal effect of X on Y is identifiable, and given by

$$P(y|\check{x}) = \sum_z P(y|x, z)p(z).$$

- Front door criterion: If Z intercepts all directed paths between X and X and: No back door paths between X and Z, every back door path between Z, X is blocked by X. Then the causal effect between X and X is identifiable and given by

$$P(y|\check{x}) = \sum_x P(z|x) \sum_x{}' P(y|x', z)P(x')$$

- Figure 3 looks nearly identical to the diagram of confounding we teach in intro stats courses. I (Austin) am having a tough time buying that front door criterion holds. If the unobserved affects both X and Y, how can we possibly derive a form for the causal effect from X to Y? I am missing something from this portion.

4. A Calculus of Intervention

- Use diagrams to manipulate causal effects $P(y|\check{x})$
- If can reduce $P(y|\check{x})$ into an expression involving standard (unchecked) probabilities of observed quantities, then causal effects of X on Y is identifiable.
- Three rules governing: insertion/deletion of observables, action/observation exchange, insertion/deletion of actions.
- Not clear if those three rules are sufficient to derive all identifiable causal effects
- Example of using those three rules to derive the front-door expression.
- If we want to estimate $P(y|\check{x})$, but not identifiable. If we can't run experiment on X directly, can we identify another set of variable Z which can be controlled?
- If we can transform $P(y|\check{x})$ into expression with only z checked, then can recover causal effect of X on Y via experiment using Z.

5. Graphical Tests of Identifiability

- Bow pattern is and equation $Y = f_Y(X, U, \epsilon_Y)$ where U is unobserved. It does not permit causal inference.
- Presence of a variable $Z \to X$. Where Z is connected to X, but not to U. This will facilitate the relationship.
- Confounding arc: If there is a back-door path which contains only unobserved variables and no converging arrows along the path, can replace whole path with confounding arc.
- If confounding arc present between X, Y, then $P(y|\check{x})$ cannot be identified, except when linearity assumptions are made (Instrumental Variables)
- For non-parametric models, addition of arcs can impede but not help identifiability, bc reduces possible d-separation.
- Examples + analysis of simple graphs where $P(y|\check{x})$ is identifiable.
- Example similar to Rubin, where identifiability of $P(y|\check{x})$ requires adjusting for concomitant RV, but adjustment has unusual form.
- Local identifiability not necessary for global identifiability.
- Also examples of non-identifiable DAGs.