Causal inference in multilevel data structures:

Discussion of papers by Li and Imai

Jennifer Hill

May 19th, 2008

Li paper

Strengths

- Area that needs attention!
- With regard to propensity score strategies advice is often that will only affect s.e.'s, not necessarily point estimates – this shows not true
- Important to see the implications of ignoring clustering since that is often done
- Separates out the implications of mis-specifying the model for the assignment mechanism and the model for the response

Concerns

- What's being estimated/assumed
  - Causal vs. not causal
  - Individual vs cluster-level effects
  - SUTVA violations?

Importance of formalizing as a causal estimand

- While Li et al. have defined a valid estimand and derived properties of their estimators with respect to it, it seems the most important decisions regarding modeling and assumptions arise once we make the jump to a causal estimand
- Li et al. write their "response model" as:

$$y_{hk} = \delta_h + \gamma_h z_{hk} + \alpha d_h + \varepsilon_{hk}$$

where, recall, $d_h = (1/n_h)\Sigma z_{hk}$

- What do these parameters mean?
- How do we translate this into a causal model..?
- It depends…

Causal formalization #1

$y_{hk} = \delta_h + \gamma_h z_{hk} + \alpha d_h + \varepsilon_{hk}$

This model seems to imply that the outcome for person k in cluster h depends not only on his own treatment assignment but also the treatment assignment of others in his cluster (so proportion treated could be interpreted causally as well).

This is a SUTVA violation though a potentially tractable one (particularly if we resign ourselves to making inferences at the cluster level). So potential outcomes then would be defined with respect to both z and d

$y_{hk}(z_{hk}=0,d_h) = \delta_h + \alpha d_h + \varepsilon_{hk}$

$y_{hk}(z_{hk}=1,d_h) = \delta_h + \gamma_h + \alpha d_h + \varepsilon_{hk}$

Which leaves room for any number of different kinds of average causal comparisons…

However we have to make stronger assumptions to identify all of these effects.

Causal formalization #2

$y_{hk} = \delta_h + \gamma_h z_{hk} + \alpha d_h + \varepsilon_{hk}$

However another interpretation is that d_h is simply a proxy for an unobserved confounder (e.g. $d_h \sim N(\beta W, \sigma_d^2)$), but conditional on z, d doesn't affect y

For instance if the treatment z denoted Head Start participation in Head Start and W reflected the overall spending on social services in community h (which is associated with both d_h and subsequent outcomes). In this case we wouldn't interpret α causally, rather d_h is probably something we want to control for in order to satisfy ignorability (so α is a nuisance parameter).

No SUTVA violation now.

Then, potential outcomes could be defined with respect to z alone

$y_{hk}(z_{hk}=0) = \delta_h + \alpha d_h + \varepsilon_{hk}$

$y_{hk}(z_{hk}=1) = \delta_h + \gamma_h + \alpha d_h + \varepsilon_{hk}$

Other models

- Of course this is also a rather limited structure, should also consider, for instances, cases such as
  - When $\alpha = 0$
  - When d_h only impact potential outcomes for the treated
  - When clusters proxy for unobserved heterogeneity in a more general way (i.e. not proxied by measured variabls)

Also…

Choice of weights….

- Why use the basic HT weights rather than stabilized versions?
- Not convinced of the utility of the "population overlap" estimand
- Since TOT and TOC are popular estimands would be interesting to see implications in those scenarios as well.

Imai paper

Strengths

- Formalizes aspects of the MPCR that I haven't seen done before with regard to SUTVA at individual and cluster (both within pair and across pairs) levels
- Points out important misconceptions in previous discussions of MPCR
- Delineates when the commonly used harmonic mean estimators will fail
- Clarity about potential estimands

Throws down the gauntlet:

- "*pair matching should be used whenever feasible*"

Concerns

- SATE/CATE requirements for unbiasedness – can be hard to control cluster sizes in field experiments and matching that is strong enough to match SATE across clusters within pairs seems quite unlikely (so even if better than harmonic mean or rb approaches, still a cause for concern) – also doesn't indiv.level interference make this less likely?
- PATE/UATE need cluster sizes uncorrelated with cluster treatment effects (this made more plausible by above conditions – similarity across clusters within pairs in terms of cluster size and covariate distributions)
- Variance not identified for SATE/CATE (is upper bound calculation small enough to be at all helpful?)

Variance estimation

- If you only care about SATE or CATE why can't you use randomization based inference (then constant treatment effect assumption shouldn't matter should it?)

Simulations

- Designed to see what happens when the assumptions necessary for the harmonic mean estimators are not met
- However there is no simulation condition that tests what happens if the assumptions necessary for the new estimators are not met….? If I can't balance cluster sizes, how similar do those clusters have to be?

Comparison to UMCR

- When comparing the efficiency gains of the matched pairs design over the unmatched cluster randomized design the authors assume that treatment effect estimation is performed without further covariance adjustment
- UMCR with additional covariance adjustment seems to me to be the more reasonable comparison

Alternative designs…?

- Why not take a fully Bayesian approach to estimation? Should be able to account for the design and incorporate additional covariance adjustments.
- Could relax some of the assumptions that the harmonic mean estimator is implicitly built upon (normality, iid distributions, equal variances for the potential outcomes)