Discussion of Papers on the Extensions of Propensity Score

Kosuke Imai

Princeton University

August 3, 2010

The Theme and Papers of this Panel

- Theme: Extensions of the Propensity Score Method
- Wang: Extend propensity score to continuous treatment regimes
- Huang: Extend propensity score to mismeasured covariates
- Hong: Extend propensity score to causal mediation analysis
- Hansen: Provide a theoretical justification for some propensity score matching methods

The Wang Paper

- Extend propensity score to continuous treatment regimes
- Imai and van Dyk (*JASA*, 2004)
  - Parametric model for conditional probability of treatment: $p_\psi(A \mid X)$
  - Generalized propensity function: $e(\cdot \mid X) = p_\psi(\cdot \mid X)$
  - Suppose θ uniquely represents $e(\cdot \mid \theta_\psi(X))$
  - $e(\cdot \mid X)$ depends on X only through $\theta_\psi(X)$
- Two main theoretical results:
  1. Propensity function as balancing score
  2. Ignorable treatment assignment given propensity function
- Practical implications:
  - Causal effects estimation:
    1. Estimate the propensity function $p_\psi(A \mid X)$
    2. Subclassify on $\hat{\theta}(X)$
    3. Estimate causal effects within each subclass and aggregate
  - Model diagnostics: check independence between A and X after conditioning on $\hat{\theta}(X)$

The Wang Paper (continued)

- Main theoretical results:
  1. Depending on the outcome model, one can further reduce the dimension of the propensity function
  2. If the interactions between treatment and some covariates exist in the outcome model, these covariates need to be adjusted in addition to the propensity function
- Two main advantages of the propensity score method:
  1. Robustness: when the knowledge of outcome model is lacking
  2. Diagnostics: when the knowledge of propensity model is lacking
- What are the practical implications of these two theoretical results in light of these advantages?
  1. How do these results help analysts if they do not possess the knowledge of outcome model?
  2. Do these results suggest new diagnostics about propensity or outcome model specification?
  3. Models with interactions: Don't we know the propensity model is incorrect if covariates aren't balanced?

The Huang Paper

- Use of propensity score with mismeasured covariates
- Some existing works on mismeasured treatments (e.g., Lewbel; Imai and Yamamoto) but little is done on mismeasured covariates

- Non-differential measurement error: conditionally independent of potential outcomes given (true) covariates
- In addition, measurement error is assumed to be independent of treatment status
- These are reasonable assumptions
- Nonparametric identification analysis in a simple situation (e.g., binary treatment, covariate, and outcome) may be illuminating

The Huang Paper (continued)

- Theorem 2: true propensity score balances true covariates and measurement error as well as mismeasured covariates
- Identification via the restriction on propensity model
- How can we diagnose propensity model specification? Is balancing mismeasured covariates sufficient?

- Finite mixture model that combines outcome model, propensity model, measurement error model
- This is nice but how does one conduct diagnostics within this approach?
- Are there additional advantages for simultaneously modeling propensity score and exposure effect?

The Hong Paper

- Use of propensity score in causal mediation analysis
- Exploration of causal mechanisms require the estimation of natural direct and indirect effects
- Under standard designs, the mediator is not randomized
- Under sequential ignorability, natural direct/indirect effects are *nonparametrically* identified (Imai *et al.* Stat. Sci. 2010):

$$\begin{aligned} \{Y_i(t', m), M_i(t)\} &\perp\!\!\!\perp T_i \mid X_i = x, \\ Y_i(t', m) &\perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x, \end{aligned}$$

- Nonlinear structural estimation: Imai *et al.* (Psy. Meth. in-press) and Pearl (Working paper, 2010)
- Marginal structural estimation: VanderWeele (Epidemiology, 2009)

The Hong Paper (continued)

- A new approach based on propensity score weighting that can handle post-treatment confounders
- Treatment-by-mediator interaction effects are also handled but this is not a problem for existing methods so long as post-treatment confounders do not exist
- The outcome model is nonparametric
- Robins' no-interaction effect assumption in the presence of post-treatment confounder is difficult to justify
- What are the key identifying assumptions in this paper? And how they should be interpreted by substantive researchers?

$$\begin{aligned} Y_i(t, m) &\perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x, L_i(t) = l \\ Y_i(t, m) &\perp\!\!\!\perp M_i(t') \mid T_i = t, X_i = x, L_i(t) = l \\ L_i(t) &\perp\!\!\!\perp M_i(t') \mid T_i = t, X_i = x \end{aligned}$$

- Need for empirical and simulation examples: what happens if mediator is continuous and/or has skewed distributions?

The Hansen Paper

- Novel analytical results: even if one cannot match exactly on propensity score, in a large sample correct inference can be made so long as covariate balance is "good enough"
- Formal definitions of informally used concepts; adjustability, crude balance
- Formal results showing the conditions under which matching can be justified
- What are the implications for practice?
    - Conduct balance test and then estimate causal effects?
    - Can balance tests be used to diagnose the misspecification of propensity score model? If so, how? The possibility of multiple testing?
    - If two different matching methods give the "same" result in terms of balance tests, which one should one choose?

The Hansen Paper (continued)

- A different perspective for the purpose of discussion (Imai et al. *JRSSA*; Ho et. al. *Pol. Anal.*)
- "balance tests" are often conducted on different matched samples to diagnose the misspecification of propensity model
- multiple testing, different sample size can be problematic
- balance should be maximized without limit for better inference
- the gold standard is the experiment with matched-pair design rather than the experiment with simple randomization
- matching with pre-determined balance
    - matching with fine balance
    - coarsened exact matching
    - maximum entropy matching
- matching as nonparametric preprocessing for making parametric inference robust

Concluding Remarks

- A great set of papers extending the propensity score methods to various situations of practical importance
- Common challenges:
  1. Development of diagnostics for propensity model specification
  2. Connecting interesting theoretical results to practice
- Look forward to seeing future versions of the papers