Discussion of “The Blessings of Multiple Causes” by Wang and Blei

Kosuke Imai Zhichao Jiang

Harvard University

JASA Theory and Methods Invited Papers Session

Joint Statistical Meetings

July 29, 2019

Exciting Paper that Opens Up New Research Frontiers

1. Causal heterogeneity
   - Existing work: heterogeneous effects of a treatment
   - This work: effects of heterogenous treatments

2. Causal inference and machine learning
   - Existing work: estimation of propensity score, heterogenous effects
   - This work: control for unobserved confounding

Thought-provoking paper on an extremely important topic

The Deconfounder Method

- Setup:
  - multiple causes: $\mathbf{A}_i = (A_{i1}, A_{i2}, \ldots, A_{im})$
  - unobserved multi-cause confounders: $\mathbf{A}_i \perp\!\!\!\perp Y_i(\mathbf{a}) \mid \mathbf{U}_i$
  - no unobserved single-cause confounder:
- Methodology:
  1. Factor model

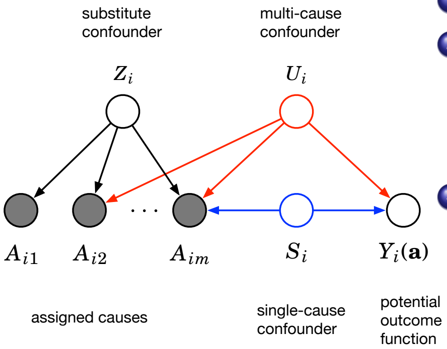
$$p(A_{i1}, A_{i2}, \ldots, A_{im}) \; = \; \int p(\mathbf{Z}_i) \prod_{j=1}^{m} p(A_{im} \mid \mathbf{Z}_i) \; d\mathbf{Z}_i$$

  2. Substitute confounder

$$\mathbb{E}\{Y_i(\mathbf{a}) - Y_i(\mathbf{a}')\} \; = \; \mathbb{E}\{\mathbb{E}(Y_i \mid \mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i) - \mathbb{E}(Y_i \mid \mathbf{A}_i = \mathbf{a}', \mathbf{Z}_i)\}$$

- Advantages:
  1. checkable assumption about unobserved confounders
  2. easy to implement

Assumptions



1. SUTVA
2. No single-cause confounder
   $A_{ij} \perp\!\!\!\perp Y_i(\mathbf{a})$
   $\rightsquigarrow$ This should be
   $\mathbf{A}_i \perp\!\!\!\perp Y_i(\mathbf{a}) \mid \mathbf{U}_i$, $A_{ij} \perp\!\!\!\perp A_{ij'} \mid \mathbf{U}_i$
3. Overlap:
   $p(A_{ij} \in \mathcal{A} \mid \mathbf{Z}_i) > 0$ for all sets $\mathcal{A}$ with $p(\mathcal{A}) > 0$

- $\mathbf{Z}_i$ is a function of causes $\rightsquigarrow$ different from usual propensity score
  1. factor model: $\mathbf{Z}_i = \hat{h}(\mathbf{A}_i) \to h(\mathbf{A}_i)$ as the sample size grows
     $\rightsquigarrow$ overlap assumption might be difficult to satisfy
  2. A_{ij} causally affects $A_{ij'}$
     $\rightsquigarrow$ factor model no longer applicable, must know causal ordering

Nonparametric Identification

- Two-step proof:
  1. existence of factor model for multiple causes (Proposition 5)
  2. nonparametric identification of causal effects given the consistency of substitute confounders (Theorem 6)

- D'Amour (2019)
  1. provides an example where many factor models are compatible with observed data, yielding different causal effect estimates
  2. shows even the existence of unique factor model does not guarantee identification

Mechanics of the Substitute Confounder

- Substitute confounder has the property: $\mathbf{A}_i \perp\!\!\!\perp \mathbf{U}_i \mid \mathbf{Z}_i$

$$
\begin{aligned}
& \mathbb{E}\{Y_i(\mathbf{a}, \mathbf{U}_i)\} \\
= & \int Y_i(\mathbf{a}, \mathbf{U}_i = \mathbf{u}) p(\mathbf{U}_i = \mathbf{u}) d\mathbf{u} \\
= & \int\int Y_i(\mathbf{a}, \mathbf{U}_i)\ p(\mathbf{U}_i = \mathbf{u} \mid \mathbf{Z}_i = \mathbf{z}) p(\mathbf{Z}_i = \mathbf{z}) d\mathbf{u} d\mathbf{z} \\
= & \int\int Y_i(\mathbf{a}, \mathbf{U}_i)\ p(\mathbf{U}_i = \mathbf{u} \mid \mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i = \mathbf{z}) p(\mathbf{Z}_i = \mathbf{z}) d\mathbf{u} d\mathbf{z} \\
= & \int \mathbb{E}(Y_i \mid \mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i = \mathbf{z}) p(\mathbf{Z}_i = \mathbf{z}) d\mathbf{z}
\end{aligned}
$$

- Implied estimator:

$$
\mathbb{E}\{\widehat{Y_i(\mathbf{a}, \mathbf{U}_i)}\} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i \mid \widehat{\mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i} = \widehat{\mathbf{Z}}_i) \text{ where } \widehat{\mathbf{Z}}_i = \hat{h}(\mathbf{A}_i)
$$

The Support Condition of the Substitute Confounder

- The support of $p(\mathbf{Z}_i)$ must be the same as that of $p(\mathbf{Z}_i \mid \mathbf{A}_i = \mathbf{a})$ $\rightsquigarrow$ otherwise, we can't compute $\mathbb{E}(Y_i \mid \mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i = \mathbf{z})$ for some $\mathbf{z}$
  - Equivalent to the support condition $p(\mathbf{A}_i = \mathbf{a} \mid \mathbf{Z}_i = \mathbf{z}) > 0$ for all $\mathbf{a}$ and $\mathbf{z}$ such that $p(\mathbf{A}_i = \mathbf{a}) > 0$ and $p(\mathbf{Z}_i = \mathbf{z}) > 0$.
  - Given $\mathbf{A}_i$, $\mathbf{Z}_i = h(\mathbf{A}_i)$ is constant
- Average causal effect of changing the first k ($k < m$) causes (Theorem 7)
  - Requires the calculation of $\mathbb{E}(Y_i \mid \mathbf{A}_{i,1:k} = \mathbf{a}_{1:k}, \mathbf{Z}_i = \mathbf{z})$
  - The same support condition problem applies
  - Example $Z_i = \sum \alpha_j A_{ij}$ violates the support condition when A_{ij}'s are binary

Different Assumptions?

- "D'Amour (2019) do not make the same assumptions as in Theorem 6" (p. 54)
  1. The outcome is separable:

     $$\mathbb{E}(Y_i(\mathbf{a}) \mid \mathbf{Z}_i = \mathbf{z}) = f_1(\mathbf{a}) + f_2(\mathbf{z})$$

     - Then, $\mathbb{E}(Y_i(\mathbf{a}) \mid \mathbf{Z}_i = \mathbf{z}) - \mathbb{E}(Y_i(\mathbf{a}') \mid \mathbf{Z}_i = \mathbf{z}) = f_1(\mathbf{a}) - f_1(\mathbf{a}')$
     - But, in large sample, $\mathbf{Z}_i = h(\mathbf{A}_i) \rightsquigarrow f_2(\mathbf{z}) = f_2(h(\mathbf{a}))$ separability does not hold in general
  2. The substitute confounder is a piece-wise constant function of the (continuous) causes, $\nabla_{\mathbf{a}} f_\theta(\mathbf{a}) = 0$ where $p(\mathbf{Z}_i \mid \mathbf{A}_i = \mathbf{a}, \theta) = \delta_{f_\theta(\mathbf{a})}$
     - Changing $\mathbf{a} \to \mathbf{a}'$ does not change $f_2(\mathbf{z})$
     - But again, in large sample, $f_2(\mathbf{z}) = f_2(h(\mathbf{a}))$ and so this assumes away confounding issue all together
- Binary causes
  1. Separability: $\mathbb{E}(Y_i(\mathbf{a}) - Y_i(\mathbf{a}') \mid \mathbf{Z}_i = \mathbf{z}) = f_1(\mathbf{a} - \mathbf{a}') + f_2(\mathbf{z})$
  2. There exist $\mathbf{a}_{new}, \mathbf{a}'_{new}$ s.t. $\mathbf{a}_{new} - \mathbf{a}'_{new} = \mathbf{a} - \mathbf{a}' \overset{?}{\to} f(\mathbf{a}_{new}) = f(\mathbf{a}'_{new})$

  $\rightsquigarrow$ the same problems as above apply here

Connection to the Control Function in Econometrics

- Control function is a variable that, when adjusted for, renders a treatment variable exogenous $\rightsquigarrow$ deconfounder!
- But, it requires instrumental or proxy variable W_i
- Classic two-stage least squares:
  1. Regress A_i on W_i and obtain residuals $\hat{\epsilon}_i$
  2. Regress Y_i on A_i and $\hat{\epsilon}_i$
- Nonparametric identification of triangular system (Imbens and Newey, 2009):

$$\begin{aligned} Y_i &= g(A_i, \epsilon_i) \\ A_i &= h(W_i, \eta_i) \end{aligned}$$

  where $W_i \perp\!\!\!\perp (\epsilon_i, \eta_i)$ and $h(\cdot, \cdot)$ is strictly monotonic in a continuous disturbance η_i
  - Control function is $V_i = F_{A|W}(A_i, W_i)$
  - Even in this case, for the identification of causal effects, we require the support of $p(V_i)$ is the same as that of $p(V_i \mid A_i)$

A Possible Parametric Strategy

1. Assume that the joint distribution of $(\mathbf{A}_i, U_i)$ is uniquely identifiable
   - If U_i is binary and we have 3 binary causes $\mathbf{A}_i = (A_{i1}, A_{i2}, A_{i3})$ such that A_{ij} is independent of $A_{ij'}$ given U_i, then the joint distribution of $(U_i, \mathbf{A}_i)$ is identifiable up to label switching
   - Kruskal (1977) for general results on discrete variables and Allman *et al* (2009) and Stanghellini *et al.* (2013) for recent generalization to correlated variables
2. Assume a parametric outcome model:

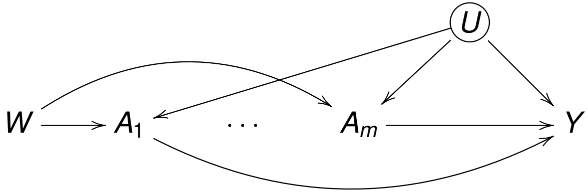
$$Y_i = \sum_{k=1}^{K} \beta_k f_k(\mathbf{A}) + \gamma g(U_i)$$

where $f_k(\cdot)$ and $g(U_i)$ are pre-specified functions.

- If γ is known, the average treatment effects are identifiable so long as $(f_1, \ldots, f_K)$ are linearly independent
- If γ is unknown, the average treatment effects are identifiable so long as $(f_1, \ldots, f_K, \mathbb{E}\{g(U_i \mid \mathbf{A}_i)\})$ are linearly independent

Nonparametric Strategy Using Instrumental Variables

- Cox and Donnely (2011):
  if an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved nonparametrically, then it is usually dangerous to resolve it parametrically.
- Nonparametric identification via instrumental variables:



- The separable outcome model:

$$Y_i = g(\mathbf{A}_i) + U_i$$

  with the instrumental variable satisfying $\mathbb{E}(U_i \mid W_i) = 0$

1. All binary causes

$$\mathbb{E}(Y_i \mid W_i) = \sum_{a_1,\ldots,a_m=0,1} g(a_1,\ldots,a_m) \times \Pr(A_{i1}=a_1,\ldots,A_{im}=a_m \mid W_i).$$

where W_i must have more than 2^m levels

2. All continuous causes

$$\mathbb{E}(Y_i \mid W_i) = \int g(a_1,\ldots,a_m) \times p(A_{i1}=a_1,\ldots,A_{im}=a_m \mid W_i) da_1 \cdots da_m.$$

where for any function g, if $\mathbb{E}\{g(A_{i1},\ldots,A_{im}) \mid W_i = w\} = 0$ for all w, then $g(a_1,\ldots,a_m) = 0$ for all $a_1,\ldots,a_m$.

Concluding Remarks

- The Wang and Blei paper opens up new research frontiers:
  - causal inference with many causes
  - use of factor models in causal inference

- Key insight: factorization $\rightsquigarrow$ deconfounder (checkable)
- Difficulty: the support condition for the substitute confounder

- Two possible ways to make progress:
  1. parametric strategies to identify factor and outcome models
  2. nonparametric strategies based on instrumental and proxy variables $\rightsquigarrow$ connections to the control function method