Blessings and Challenges of Multiple Causes

April 26th, 2023

Papers

1. Wang, Y., & Blei, D. M. (2019). The blessings of multiple causes. Journal of the American Statistical Association, 114(528), 1574-1596.
2. Athey, S., Imbens, G. W., & Pollmann, M. (2019). Comment on: "the blessings of multiple causes" by Yixin Wang and David M. Blei. Journal of the American Statistical Association, 114(528), 1602-1604.
3. Imai, K., & Jiang, Z. (2019). Comment: The challenges of multiple causes. Journal of the American Statistical Association, 114(528), 1605-1610.
4. Ogburn, E. L., Shpitser, I., & Tchetgen, E. J. T. (2019). Comment on "blessings of multiple causes". Journal of the American Statistical Association, 114(528), 1611-1615.

Overview

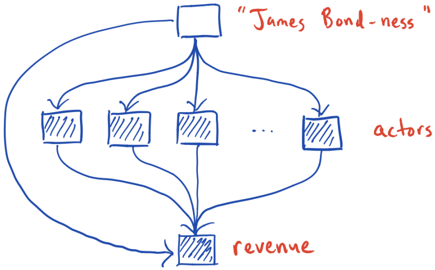
- The Blessing of Multiple Causes
  1. Motivating Example: movie data
  2. Method: deconfounder
  3. Theory: why it works
- Comments
  1. Limitations in the context of financial example. (Athey 2019)
  2. Limitations of factor model and assumptions on ATE identification. (Imai 2019)
  3. Criticism on properties of latent factors. (Ogburn 2019)

BMC: Motivation

Title	Cast	Revenue
Avatar	{Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ... }	\$2788M
Titanic	{Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ... }	\$1845M
The Avengers	{Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ... }	\$1520M
Jurassic World	{Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D'Onofrio, ... }	\$1514M
Furious 7	{Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, ... }	\$1506M
Avengers: Age of Ultron	{Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, ... }	\$1405M
Frozen	{Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, ... }	\$1274M
Iron Man 3	{Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, ... }	\$1215M
Minions	{Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, ... }	\$1157M
Captain America: Civil War	{Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, ... }	\$1153M
⋮	⋮	⋮

- Data about movies: casts and revenue
- Goal: Understand the causal effect of putting an actor in a movie
- Causal: "What will the revenue be if we make a movie with a particular cast?"

BMC: Confounder



- Confounders: variables that affect both the causes and the potential outcome.
- E.g:
  1. James Bond movies are about James Bond, a British spy
  2. Cast James Bond, M, Q, Ms. Moneypenny
  3. M, Q, Ms Moneypenny only appear in Bond movies
  4. Bond movies always do well at the box office

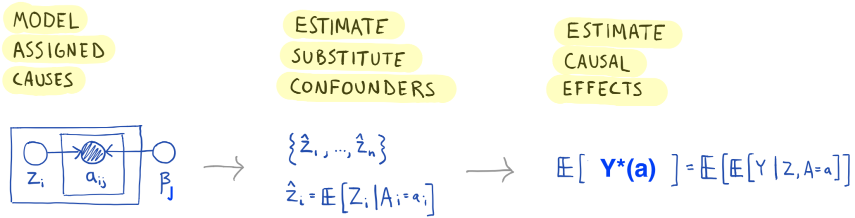
BMC: Unmeasurable Confounders

- If we observe all the confounders $\mathbf{X}$: NUC condition satisfied

$$Y^*(\mathbf{a}) \perp\!\!\!\perp \mathbf{A} | \mathbf{X}, \forall \mathbf{a}. \quad \mathbb{E}(Y^*(\mathbf{a})) = \mathbb{E}[\mathbb{E}(Y | \mathbf{X}, \mathbf{A} = \mathbf{a})]$$

- If not: equality does not hold, resulting biased estimation of $\mathbb{E}(Y^*(\mathbf{a}))$.

BMC: Deconfounder



- Find, fit, and check a factor model of multiple causes. (A predictive check compares the observed assignments with assignments drawn from the model's predictive distribution. If the model is good, then there is little difference.)
- Use the factor model to form substitute confounders for each sample.
- Use the substitute confounders in a causal model.

Discussion: what assumptions needed?

BMC: Identification assumptions

1. SUTVA
2. NUSC: no unobserved single-cause confounders: confounders only affect one cause and the potential outcome. Formally:
   - $\forall j = 1, \ldots, m : \exists V_{ij}$ s.t $A_{ij} \perp\!\!\!\perp Y^*(\mathbf{a})|(\mathbf{X}_i, V_{ij})$, and $A_{ij} \perp\!\!\!\perp A_{i,-j}|V_{ij}$
   - There is no proper subset of the sigma algebra $\sigma(V_{ij})$ satisfies the second condition.
   - $\mathbf{X}_i$ contains all the single cause confounders.
   - V_{ij} is the multiple-cause confounders that affect A_{ij}. V_{ij} induced dependence between causes.
   - Requires marginal Independence rather than joint independence.
3. Overlap: $\mathbb{P}(A_{ij} \in \mathcal{A}|\mathbf{Z}_i) > 0, \forall \mathcal{A}$ s.t. $\mathbb{P}(\mathcal{A}) > 0$
4. Consistency of substitute confounders: $p(\mathbf{z}_i|\mathbf{a}_i, \boldsymbol{\theta}) = \delta_{f_\theta(\mathbf{a}_i)}$ for some function f_θ. Z_i can be estimated from $\mathbf{A}_i$ with certainty.

BMC: Sketch of theory

1. Factor models + no unobserved single-cause confounders $\Rightarrow$ unconfoundedness.
2. Substitute confounder: captures all multiple-cause confounders, does not capture any mediators
3. $\Rightarrow$ if the factor model captures the distribution of the assigned causes then the substitute confounder renders the assignment ignorable. Moreover, such a factor model always exists.
4. Under identification assumptions, the deconfounder identifies the average causal effects and the conditional potential outcomes under different conditions.

BMC: unconfoundedness

$$\begin{tikzcd}
& \text{Kallenberg construction} \arrow[dr, Leftrightarrow] & \\
\text{Factor model} \arrow[ur, Rightarrow] & & \text{Weak unconfoundedness}
\end{tikzcd}$$

BMC: unconfoundedness

Kallenberg construction

Factor model

Weak unconfoundedness

Fitted factor model captures $p(\mathbf{a}_{1:n})$

$$p(z_{1:n}, \mathbf{a}_{1:n}, \theta_{1:m}) = p(z_{1:n}) \prod \prod p(a_{ij} | z_i, \theta_j),$$

$$p(\mathbf{a}_{1:n}) = \int p(z_{1:n}, \mathbf{a}_{1:n}, \theta_{1:m}) dz_{1:n}$$

$$(A_{i1}, \ldots, A_{im}) \perp\!\!\!\perp Y_i^*(\mathbf{a}) | Z_i, \forall \mathbf{a}$$

BMC: unconfoundedness

Distribution of $(A_{i1}, \ldots, A_{im})$ satisfies $\forall \mathbf{a}$:

$$\exists f_j, \text{ s.t } A_{ij} \overset{a.s.}{=} f_j(Z_i, U_{ij}), U_{ij} \sim Unif(0,1), (U_{i1}, \ldots, U_{im}) \perp\!\!\!\perp (Z_i, Y_i^*(\mathbf{a}))$$

Kallenberg construction

Factor model

Weak unconfoundedness

Fitted factor model captures $p(\mathbf{a}_{1:n})$

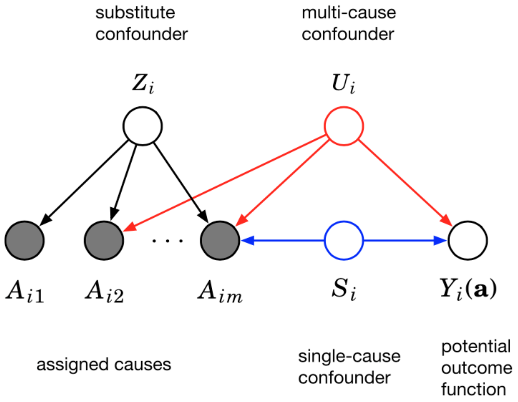
$$p(z_{1:n}, \mathbf{a}_{1:n}, \theta_{1:m}) = p(z_{1:n}) \prod \prod p(a_{ij}|z_i, \theta_j),$$

$$p(\mathbf{a}_{1:n}) = \int p(z_{1:n}, \mathbf{a}_{1:n}, \theta_{1:m}) dz_{1:n}$$

$$(A_{i1}, \ldots, A_{im}) \perp\!\!\!\perp Y_i^*(\mathbf{a}) | Z_i, \forall \mathbf{a}$$

BMC: properties of substitute confounders

1. Captures unobserved confounders, proof is by contradiction.
2. Does not pick up mediators, variables along the path between causes and effects.



BMC: causal identification

1. ATE: $\mathbb{E}Y_i^*(\mathbf{a}) - \mathbb{E}Y_i^*(\mathbf{a}')$ given
    - SUTVA, NUSC, consistency of factor model
    - the substitute confounder is a piece-wise constant function of the (continuous) causes: $\nabla_{\mathbf{a}} f_\theta(\mathbf{a}) = 0$ up to a set of Lebesgue measure zero
    - the outcome is separable, $\forall(\mathbf{a}, \mathbf{x}, \mathbf{z})$ and some continuously differentiable functions f_1, f_2, f_3, f_4,

$$\mathbb{E}[Y_i^*(a)|\mathbf{Z}_i = \mathbf{z}, \mathbf{X}_i = \mathbf{x}] = f_1(\mathbf{a}, \mathbf{x}) + f_2(\mathbf{z}),$$

$$\mathbb{E}[Y_i|\mathbf{A}_i = \mathbf{a}, \mathbf{Z}_i = \mathbf{z}, \mathbf{X}_i = \mathbf{x}] = f_3(\mathbf{a}, \mathbf{x}) + f_4(\mathbf{z}).$$

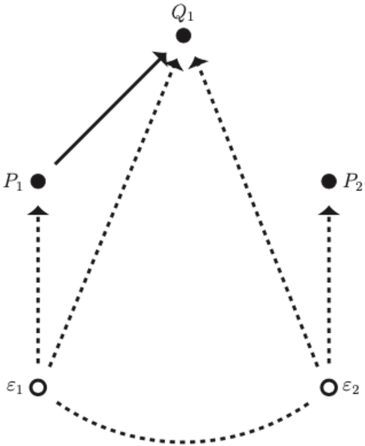
2. ATE for first k causes given: SUTVA, NUSC, consistency of factor model and $\mathbb{P}((A_{i1}, ..., A_{ik}) \in \mathcal{A}|\mathbf{Z}_i) > 0, \forall \mathcal{A}$ s.t. $\mathbb{P}(\mathcal{A}) > 0$
3. Conditional mean potential outcome: $\mathbb{E}(Y_i^*(\mathbf{a}')|\mathbf{A}_i = \mathbf{a}) = \mathbb{E}\mathbb{E}(Y_i|\mathbf{Z}_i, \mathbf{X}_i, \mathbf{A}_i = \mathbf{a}')$ given: SUTVA, NUSC, consistency of factor model and $\mathbb{P}(\mathbf{Z}_i|\mathbf{A}_i = \mathbf{a}) = \mathbb{P}(\mathbf{Z}_i|\mathbf{A}_i = \mathbf{a}')$

Discussion: what's your concern about Deconfounder?

Limitation of BMC in financial example (Athey 2019)

1. Set up
   - Two products $i = 1, 2$ in several markets indexed by t, price P_{ti}, sold quantity Q_{ti}.
   - Demand function $Q^D_{ti}(p) = \alpha_i + \beta_i p + \epsilon_{ti}$
   - Cost function $C^D_{ti}(q) = q(c_i + \eta_{ti})$
   - Profit $\Pi_{ti}(p) = pQ^D_{ti}(p) - C_{ti}(Q^D_{ti}(p))$
   - Best price $P_{ti} = \frac{c_i}{2} - \frac{\alpha_i}{2\beta_i} - \frac{\epsilon_{ti}}{\beta_i} + \frac{\eta_{ti}}{2}$
   - Interested in estimate β_1
2. Violation of weak unconfoundedness: P_{ti} is function of ϵ_{ti}
3. Multiple cause: shared unobserved confounders $\mathrm{Cor}(\epsilon_{\mathrm{t}1}, \epsilon_{\mathrm{t}2}) = \rho_\epsilon > 0$

Limitation of BMC in financial example (Athey 2019)

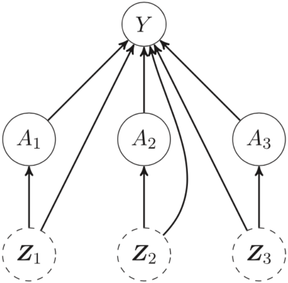


Limitation of BMC in financial example (Athey 2019)

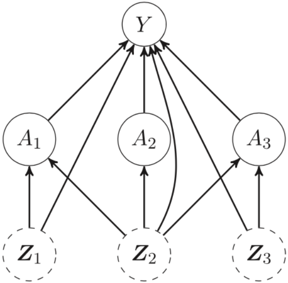
Method	Prob Limit of estimators	Unbiasd if
$Q_{t1} \sim P_{t1}$	$\beta_1 - 2\beta_1 \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \beta_1^2 \sigma_\eta^2} > \beta_1$	
$Q_{t1} \sim P_{t1}, P_{t2}$	$\beta_1 - 2\beta_1 \frac{(\sigma_\epsilon^2 + \sigma_\eta^2)\sigma_\epsilon^2 - (\rho_\epsilon \sigma_\epsilon^2 + \beta_1^2 \rho_\eta \sigma_\eta^2)\rho_\epsilon \sigma_\epsilon^2}{(\sigma_\epsilon^2 + \sigma_\eta^2)^2 + (\rho_\epsilon \sigma_\epsilon^2 + \beta_1^2 \rho_\eta \sigma_\eta^2)^2}$	biased
instrument P_{t2}	$\beta_1 - 2\beta_1 \frac{\rho_\epsilon \sigma_\epsilon^2}{\rho_\epsilon \sigma_\epsilon^2 + \rho_\eta \beta_1^2 \sigma_\eta^2}$	$\rho_\epsilon = 0, \rho_\eta > 0$
$Q_{t1} \sim P_{t1}, \bar{P}_t$	$\beta_1 - 2\beta_1 \frac{(1-\rho_\epsilon)\sigma_\epsilon^2}{(1-\rho_\epsilon)\sigma_\epsilon^2 + (1-\rho_\eta)\beta_1^2 \sigma_\eta^2}$	$\rho_\epsilon = 1, \rho_\eta \neq 1$

Limitations of BMC theoretically (Imai 2019)

1. NUSC:
   - $\forall j = 1, \ldots, m : \exists V_{ij}$ s.t $A_{ij} \perp\!\!\!\perp Y^*(\mathbf{a}) | (\mathbf{X}_i, V_{ij})$, and $A_{ij} \perp\!\!\!\perp A_{i,-j} | V_{ij}$ There is no proper subset of the sigma algebra $\sigma(V_{ij})$ satisfies the second condition.
   - Comment: difficult to distinguish single cause or multiple causes confounders without the knowledge of causal relationships among the variables.



(a) only unobserved single-cause confounders exist



(b) both unobserved single-cause and multiple-cause confounders exist

Limitations of BMC theoretically (Imai 2019)

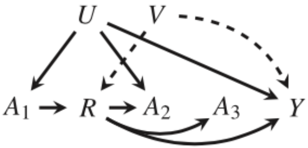
2. Identifiability of factor model: different factor models give different ATE estimation. To identify factor model, $\mathbf{Z}_i$ must be causally affect $\mathbf{A}_i$. Faithfulness assumption in DAG.
3. Identification of ATE:
   - estimated latent variables $\hat{\mathbf{Z}}_i = \mathbb{E}_{\text{Model}}(\mathbf{Z}_i|\mathbf{A}_i) \rightarrow$ function of observed causes rather than $\mathbf{Z}_i$.
   - requires $p(\hat{\mathbf{Z}}_i|\mathbf{A}_i = \mathbf{a})$ and $p(\hat{\mathbf{Z}}_i)$ have the same support.
   - $p(\hat{\mathbf{Z}}_i|\mathbf{A}_i = \mathbf{a})$ degenerates. And overlap does not take effect since it's about $\mathbf{Z}_i$.
   - BMC deal with this with additional conditions:

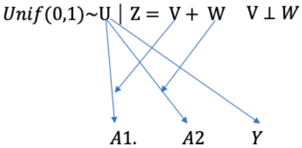
$$\nabla_{\mathbf{a}} f_\theta(\mathbf{a}) = 0$$
$$\mathbb{E}[Y_i^*(\mathbf{a})|\hat{\mathbf{Z}}_i = \mathbf{z}, \mathbf{X}_i = \mathbf{x}] = f_1(\mathbf{a}, \mathbf{x}) + f_2(\mathbf{z}),$$
$$\mathbb{E}[Y_i|\mathbf{A}_i = \mathbf{a}, \hat{\mathbf{Z}}_i = \mathbf{z}, \mathbf{X}_i = \mathbf{x}] = f_3(\mathbf{a}, \mathbf{x}) + f_4(\mathbf{z}).$$

Validity of conditions critically depends on the choice of factor model.

Errors of BMC (Ogburn 2019)

1. Conditional independent causes DoNot ensure conditional ignorability:
    - **Z** can include variables that bias the ATE: like single cause mediator, single cause collider and M-bias collider.
    - Even if **Z** renders conditional independence, it cannot control confounding.





Limitations of BMC (Ogburn 2019)

2. Identification of ATE requires parametric assumptions that the confounding variable is a clustering indicator and that the treatment effects are constant across clusters (no treatment-confounder interaction)
3. Strong overlap assumption (as pointed out by Imai 2019)
4. Consistency of $\mathbf{Z}$ requires the number of causes goes to infinity. Assumptions that are asymptotically satisfied is not guaranteed in finite samples.

Important messages

1. The idea of incongruence between outcome model and substitute confounders may serve as a general approach to causal identification.
2. Identifying a latent substitute confounder from the observed data on multiple causes essentially requires the assumption that learning structure on the causes suffices to learn about any joint structure linking the causes with the outcome.
3. The multiplicity of the causes and the consistent estimability of factor models enable us to effectively observe such multi-cause confounding. It is these two features that form the basis of the deconfounder

Thank You