On Positivity Condition for Causal Inference

Inwoo Hwang* Yesong Choe* Yeahoon Kwon Sanghack Lee

Seoul National University

ICML 2024

Table of Contents

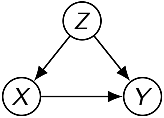
Table of Contents

Background

- Identifying and estimating a **causal effect** is a fundamental task when inferring a causal effect using observational study *without experiments*.
- Just assuming the **strict positivity** ($P(\mathbf{V}) > 0$) of the given distribution under the unconfounded assumption has been a long convention.
- We examine the graphical counterpart of the conventional positivity condition to license the use of identification formula **without strict positivity**.

Motivating Example 1

Backdoor formula:



$$P_x(y) = \sum_z P(y \mid x, \mathbf{z}) P(\mathbf{z})$$

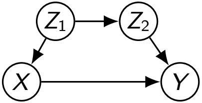
$$\Rightarrow \forall \mathbf{z}(P(\mathbf{z}) = 0 \vee P(x \mid \mathbf{z}) > 0) \equiv \text{adj}(\mathbf{x}; \mathbf{Z})$$

- Under the strict positivity, we can identify the causal effect—i.e., we can get the intervened distribution of y ($P_x(y)$) from the observed distribution $P(\mathbf{V})$.
- To estimate average treatment effect for each value of the covariate in the population, there are some subjects that received the treatment—i.e., $P(X \mid \mathbf{z}) > 0$ for all $\mathbf{z}$ with $P(\mathbf{z}) \neq 0$ (Hernán & Robins, 2006).

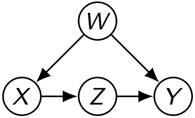
Motivating Example 2

Multiplicity of identification formulae and conditions:

- One may estimate the causal effect with a formula but **not with the other**, which was **not the case** under strict positivity.



$$\sum_{z_1} P(y \mid x, z_1)P(z_1) \quad \text{or} \quad \sum_{z_2} P(y \mid x, z_2)P(z_2)$$



$$P_x(y) = \sum_w P(y \mid x, w)P(w) \qquad \text{Backdoor}$$
$$P_x(y) = \sum_z P(z) \sum_{x'} P(y \mid x', z)P(x') \qquad \text{Front-door}$$
$$P_x(y) = \sum_z P(z \mid x)P(y \mid z, w) \qquad \text{IDENTIFY}$$

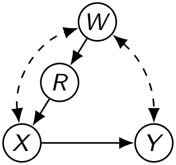
Table of Contents

Causal Identification with Strict Positivity

- The causal effect $P_x(y)$ is **identifiable** if it can be uniquely computed from $P(\mathbf{V})$ in any causal model which induces $\mathcal{G}$.
- How to identify $P_x(y)$?
  1. Do-calculus (Pearl, 1995)
  2. Q-decomposition (Tian, 2003)

  $\Rightarrow$ These well-known methods of identification heavily rely on $P(\mathbf{V}) > 0$.

  $\Rightarrow$ Their validity and mathematical correctness are unclear under **relaxed positivity**.
- e.g., Napkin



$$P_x(y) = \frac{\sum_w P(y, x \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

Do-calculus with Strict Positivity

- This calculus (Pearl, 1995) facilitates the identification of causal effects in non-parametric models.

- The following transformation are valid for any positive do-distribution induced by a model:

Definition (do-calculus)

- Rule 1 (addition/deletion of observation):
  $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$

- Rule 2 (exchange of action and observation):
  $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$

- Rule 3 (addition/deletion of action):
  $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}, \overline{\mathbf{Z}(\mathbf{W})}}}$,
  where $\mathbf{Z}(\mathbf{W}) = \mathbf{Z} \setminus An(\mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.

Q-decomposition with Strict Positivity

- c-factors derived from the given observational distribution $P(\mathbf{V})$ are used to answer the c-factors derived from the query $P_x(y)$ (Tian, 2003).

Theorem (Q-decomposition)

Given $\mathbf{H} \subseteq \mathbf{V}$, *let* $\mathbf{H}_1, \ldots, \mathbf{H}_k$ *be the c-components of* $\mathcal{G}[\mathbf{H}]$. *Let* $\prec$ *be a topological order over the variables in* $\mathbf{H}$ *according to* $\mathcal{G}[\mathbf{H}]$ *such that* $V^{(1)} \prec V^{(2)} \cdots \prec V^{(|\mathbf{H}|)}$. *Let* $\mathbf{H}^{\preceq i}$ *be the variables in* $\mathbf{H}$ *that come before* $V^{(i)}$ *including* $V^{(i)}$. *Let* $\mathbf{H}^{\succ i}$ *be the variables in* $\mathbf{H}$ *that come after* $V^{(i)}$. *Given* $Q[\mathbf{H}] > 0$,

$$Q[\mathbf{H}_j] = \prod_{V^{(i)} \in \mathbf{H}_j} \frac{Q[\mathbf{H}^{\preceq i}]}{Q[\mathbf{H}^{\preceq i-1}]},$$

where $Q[\mathbf{H}^{\preceq i}] = \sum_{\mathbf{h}^{\succ i}} Q[\mathbf{H}]$.

$$e.g., Q[W, X, Y] = \frac{Q[W, R, X, Y]}{Q[W, R, X]} \cdot \frac{Q[W, R, X]}{Q[W, R]} \cdot \frac{Q[W]}{Q[\emptyset]}$$

Table of Contents

Post-hoc Analysis (Appendix)

- **We can examine** a positivity condition under which the identification formula is well-defined.
- e.g., Napkin

$$\exists r \frac{\sum_w P(y, x \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)} \geq 0 \quad \Leftarrow \exists r (① \geq 0 \wedge ② > 0)$$

$$① \geq 0 \Leftarrow \text{adj}(r; W)$$

$$② > 0 \Leftarrow \text{adj}(r; W) \wedge P(x, r) > 0$$

$$\therefore \exists r (\text{adj}(r; W) \wedge P(x, r) > 0)$$

- While it is true that the positivity condition derived directly from a formula ensures that the formula is well-defined, yet its validity is unclear for now since the formula is derived under strict positivity.
- Post-hoc analysis yields a sufficient positivity condition for the identification formula derived through **Identify+**.

Table of Contents

Causal Identification with Relaxed Positivity: Do-Calculus

We develop a general and principled approach for deriving a positivity condition by examining the conditions for do-calculus (Pearl, 1995).

Definition (Positivity Relaxed do-calculus)

Let $\mathcal{G}$ be the directed acyclic graph (DAG) associated with a causal model, and let $P(\cdot)$ be the probability distribution induced by the model. Then,

(R1) $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})}$ and $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

(R2) $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\underline{\mathbf{Z}}}}$ and $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

(R3) $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\overline{\mathbf{Z}(\mathbf{W})}}}$ and $P_{\mathbf{x}}(\mathbf{w}) > 0$

e.g.,

$$
\begin{aligned}
P_x(y) &= P_{w,r}(y \mid x) && \text{if } P_{w,r}(x) > 0 \\
&= P_{w,r}(y, x)/P_{w,r}(x) && \text{if } P_{w,r}(x) > 0 \\
&= \frac{\sum_{w'} P(y,x|r,w')P(w')}{\sum_{w'} P(x|r,w')P(w')}. && \text{if } \mathrm{adj}(r; W)
\end{aligned}
$$

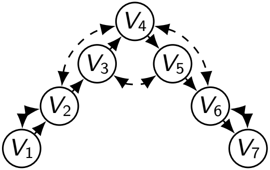
$$\therefore \exists r(\mathrm{adj}(r; W) \wedge P(x, r) > 0)$$

Table of Contents

Causal Identification with Relaxed Positivity: Q-Decomposition

- Given the theoretical underpinnings of the sufficient positivity conditions over do-calculus, we now investigate the feasibility of creating an identification algorithm capable of simultaneously taking a non-strictly positive observational distribution into account.
- The intuition behind the generalization is that the product of fractions often can be shortened by canceling out depending on the topological order.

Relaxed Q-decomposition Motivating Example



- $Q[\mathbf{H}] = Q[\mathbf{H}_1] \cdot Q[\mathbf{H}_2]$ where $\mathbf{H}_1 = \{V_1, V_2, V_4, V_6, V_7\}$ and $\mathbf{H}_2 = \{V_3, V_5\}$.
- Denoting $Q[\mathbf{H}^{\preceq i}]$ as Q_i for brevity, if $Q[\mathbf{H}] = Q_7 > 0$, then $Q[\mathbf{H}_1] = \frac{Q_7}{Q_6} \cdot \frac{Q_6}{Q_5} \cdot \frac{Q_4}{Q_3} \cdot \frac{Q_2}{Q_1} \cdot \frac{Q_1}{Q_0}$ and $Q[\mathbf{H}_2] = \frac{Q_5}{Q_4} \cdot \frac{Q_3}{Q_2}$ by Tian (2003, Lemma 4). Since Q_6 and Q_1 can be canceled out, we can write $Q[\mathbf{H}_1] = \frac{Q_7}{Q_5} \cdot \frac{Q_4}{Q_3} \cdot \frac{Q_2}{Q_0}$.
- We show that this expression is valid if $Q_5 > 0$, and further show that it is still possible to identify $Q[\mathbf{H}_1]$ when some of the denominators are 0, i.e., $Q_5 = 0$ or $Q_3 = 0$, relaxing the strict positivity condition of $Q[\mathbf{H}] > 0$ in (Tian, 2003).

Relaxed Q-decomposition

- We generalize Q-decomposition under relaxed positivity.

Theorem (Positivity Relaxed Q-decomposition)

Let $\mathbf{H}' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}])$ *where* $I_{\mathcal{G}[\mathbf{H}],\prec}(\mathbf{H}') = \{(l_d, r_d)\}_{d=1}^{T}$. *Then, the following holds:*

- *If* $Q[\mathbf{H}^{\preceq l_T - 1}] > 0$, *then*

$$Q[\mathbf{H}'] = \prod_{d=1}^{T} \frac{Q[\mathbf{H}^{\preceq r_d}]}{Q[\mathbf{H}^{\preceq l_d - 1}]}.$$

- *If* $Q[\mathbf{H}^{\preceq r_m}] = 0$ *and* $Q[\mathbf{H}^{\preceq l_m - 1}] > 0$ *for some* m, *then*

$$Q[\mathbf{H}'] = 0.$$

Relaxed Q-decomposition with Example (Napkin)

$$Q[W,X,Y] = \frac{Q[W,R,X,Y]}{\cancel{Q[W,R,X]}} \cdot \frac{\cancel{Q[W,R,X]}}{Q[W,R]} \cdot \frac{Q[W]}{Q[\emptyset]}$$

$$Q[W,X,Y] = \frac{Q[W,R,X,Y]}{Q[W,R]} \cdot \frac{Q[W]}{Q[\emptyset]} \quad \text{if } Q[W,R] > 0$$

$$Q[W,X,Y] = 0 \quad \text{if } Q[W] = 0$$

$$\Longrightarrow P_x(y) = \frac{\sum_w Q[W,X,Y](x,y,w,r)}{\sum_{y',w} Q[W,X,Y](x,y',w,r)}$$

$$\therefore \exists r(\text{adj}(r;W) \wedge P(x,r) > 0)$$

- We devise **Identify+**, a sound algorithm that returns an identification formula with sufficient positivity.

Table of Contents

Discussion & Conclusion

- We provide positivity conditions for **do-calculus** and **generalized Q-decomposition**, forming a basis for causal effect identification without $P(\mathbf{V}) > 0$.
- We devise **Identify+** algorithm, incorporating a relaxed version of **generalized Q-decomposition** into an existing identification method.
- Towards a positivity-aware identification algorithm— 3 key factors: topological order, fixing values, and latent projection.
- Since we have established sufficient conditions for both do-calculus and identification of marginal effects, our results indeed generalize to conditional causal effects as well.

Table of Contents

References

- Hernán, M. A. and Robins, J. M. Estimating causal effects from epidemiological data. Journal of Epidemiology & Community Health, 60(7):578–586, 2006.
- Pearl, J. Causal diagrams for empirical research. Biometrika, 82(4):669–688, 1995.
- Tian, J. and Pearl, J. On the identification of causal effects. Technical Report R-290-L, Department of Computer Science, University of California, Los Angeles, CA, 2003.