A causal inference framework for spatial confounding by Gilbert, Datta, Casey, and Ogburn (2023) (ArXiv)

Hyemin Yeon

Department of Statistics
Iowa State University

May 10, 2023

Causal Inference + Spatial Statistics

- Reich et al. (2021), A Review of Spatial Causal Inference Methods for Environmental and Epidemiological Applications

- Causal inference with (unmeasured) spatial confounding:
  $Y_i = \beta X_i + U_i$
  - U_i: areal data (i.e., discrete)
  - $U_i = U(s_i)$: geospatial data (i.e., continuous)

- Causal inference with spatial interference/spillover
  $\{(X_i, Y_i)\}_{i=1}^n$ is a network, $X_i \to Y_{i'}$ for $i \neq i'$

Causal inference under spatial confounding

- Spatial (unmeasured) confounding in spatial statistics
- Unmeasured confounding in causal inference
- Connection between (spatial) confounding in spatial statistics and confounding in causal inference

The goals of the paper

- GILBERT ET AL. (2023), A CAUSAL INFERENCE FRAMEWORK FOR SPATIAL CONFOUNDING, ARXIV

- Formal definition of spatial confounding

- Conditions for the identifiability of a causal estimand

- Non-parametric method for estimation (double machine learning)

Contents

Notation

- Y: outcome
- X: (binary or continuous) treatment with support $\mathcal{X}$
- $Y(x)$: potential outcome for $x \in \mathcal{X}$
- C: covariate with support $\mathcal{C}$
- P_{full}: probability distribution of full data $\{\{Y(x) : x \in \mathcal{X}\}, X, C\}$
- P_{obs}: probability distribution of observed data $\{Y, X, C\}$.
- A parameter is *identifiable* if it is written as a functional of P_{obs}.
- Causal estimand of interest: $\mathrm{E}[Y(x)]$

Identiability assumptions

- (Consistency) $Y_i = Y_i(x)$ if $X_i = x$
- (Positivity) $\forall x \in \mathcal{X}, \forall c \in \mathcal{C}, (x, c) \in \mathrm{supp}(X, C)$
  $\implies \forall x \in \mathcal{X}, \forall c \in \mathcal{C}, f(c|x) > 0$
- (Ignorability) $\forall x \in \mathcal{X},\ Y(x) \perp X|C$
  $\implies \forall x \in \mathcal{X}, \mathsf{E}[Y(x)|X = x, C] = \mathsf{E}[Y(x)|C]$
-
$$\begin{aligned} & \mathsf{E}[Y|X = x, C = c] = \mathsf{E}[Y(x)|X = x, C = c] = \mathsf{E}[Y(x)|C = c] \\ \implies & \mathsf{E}[Y(x)] = \mathsf{E}[\mathsf{E}[Y(x)|C]] = \mathsf{E}[\mathsf{E}[Y|X = x, C]] \end{aligned}$$

- $\mathsf{E}[Y(x)]$ is a functional of P_{obs} $\implies$ identifiable!

Unmeasured spatially varying confounding

- $\forall x \in \mathcal{X}, Y(x) \perp X | C, U$ but $\exists x \in \mathcal{X}, Y(x) \not\perp X | C$

- Arbitrary (unstructured) unmeasured confounding
  $\implies$ hard problem!

- Need to make (untestable) assumptions

- Unmeasured confounding with vs without spatial structure

Notations

- Omit C for brevity
- P_{full}: probability distribution of full data $\{\{Y(x) : x \in \mathcal{X}\}, X, U\}$
- P_{obs}: probability distribution of observed data $\{Y, X\}$
- U is an unmeasured/unobserved spatial confounding variable with support $\mathcal{U}$

Contents

All variation in U is captured by S

- S: the observed spatial location in $\mathbb{R}^d$ (e.g., $d = 2$) with support $\mathcal{S}$

(S) $U = g(S)$ for some fixed measurable function g (A3.3)

- E.g., $U_i = U(s_i)$ for $s_i \in \mathbb{R}^2$, $i = 1, \ldots, n$

- We want S to be a proxy for U!

Counterexamples against (S)

Counterexample 1:

- U_i: personal income
- Different U_1 and U_2 may have the same location S
  $\implies$ U is not a "function" of S

Counterexample 2:

- idea: g is measurable $\iff$ g is nearly continuous (Lusin's theorem)
- U: household income
- $S \mapsto U$ may be discontinuous, and hence, non-measurable

Ignorability

(I1) $\forall x \in \mathcal{X}, Y(x) \perp X | U$ (A3.1)

(I2) $\forall x \in \mathcal{X}, Y(x) \perp X | S, U$ (A3.4)

(I3) $\forall x \in \mathcal{X}, (Y(x), X) \perp S | U$ (A3.5)

(I$*$) $\forall x \in \mathcal{X}, Y(x) \perp X | S$

Proposition 1

(a) *Conditions (I2) and (S) imply Condition (I$*$).*

(b) *Conditions (I1), (I3), and (S) imply Condition (I$*$).*

Proposition 1(a)

- Prop 1(a): Conditions (I2) and (S) imply Condition (I$*$).
- Proof:
  - $\sigma(S) \subseteq \sigma(S, U)$ because $S \subseteq \{S, U\}$
  - $\sigma(S) \supseteq \sigma(S, U)$ because $(S, U) = (S, g(S))$ is a function of S

Proposition 1(b)

- Prop 1(b): Conditions (I1), (I3), and (S) imply Condition (I$*$).
- Proof: for $s \in \mathcal{S}$ and $u = g(s) \in \mathcal{U}$,

$$
\begin{aligned}
& \mathrm{P}(Y(x) = y, X = x | S = s) \\
= & \mathrm{P}(Y(x) = y, X = x, S = s) / \mathrm{P}(S = s) \\
= & \mathrm{P}(Y(x) = y, X = x, S = s | U = u) \mathrm{P}(U = u) / \mathrm{P}(S = s) \\
\underset{(I3)}{=} & \mathrm{P}(Y(x) = y, X = x | U = u) \mathrm{P}(S = s | U = u) \mathrm{P}(U = u) / \mathrm{P}(S = s) \\
\underset{(I1)}{=} & \mathrm{P}(Y(x) = y | U = u) \mathrm{P}(X = x | U = u) \mathrm{P}(U = u | S = s) \\
\underset{(S)}{=} & \mathrm{P}(Y(x) = y | S = s) \mathrm{P}(X = x | S = s) \text{ need 1-1 ??}
\end{aligned}
$$

Positivity

(P1) $\forall x \in \mathcal{X}, \forall u \in \mathcal{U}, (x, u) \in \mathrm{supp}(X, U)$ (A3.2)

(Pa) $X \perp S | U$ (A3.7)

- implied by (I3) (A3.5)

(P$*$) $\forall x \in \mathcal{X}, \forall s \in \mathcal{S}, (x, s) \in \mathrm{supp}(X, S)$ (A3.6)

Proposition 2

Conditions (P1), (Pa), and (S) imply Condition (P$*$).

Proposition 2

- Prop 2: Conditions (P1), (Pa), and (S) imply Condition (P$*$).
- Proof: Since $\sigma(S) = \sigma(S, U)$, we have

$$0 \underset{(P1)}{<} \mathrm{P}(X = x|U) \underset{(Pa)}{=} \mathrm{P}(X = x|U, S) \underset{(S)}{=} \mathrm{P}(X = x|S).$$

- This is not sufficient to prove (P$*$):
  $\forall x \in \mathcal{X}, \forall s \in \mathcal{S}, (x, s) \in \mathrm{supp}(X, S)$??

Estimability/Identifiability

Suppose Condition (S) and either combinations of conditions:

- (I$*$), (P$*$)
- (I2), (P1), (Pa)
  - (I2), (S) $\implies$ (I$*$)
  - (P1), (Pa), (S) $\implies$ (P$*$)
- (I1), (P1), (I3)
  - (I1), (I3), (S) $\implies$ (I$*$)
  - (P1), (I3), (S) $\implies$ (P$*$) (because (I3) $\implies$ (Pa))
- (I2), (P$*$)
  - (I2), (S) $\implies$ (I$*$)

$$\mathrm{E}[Y(x)] = \mathrm{E}[\mathrm{E}[Y|X = x, S]]$$

Shift interventions

- Shift intervention by $\delta > 0$: $\Delta = \mathrm{E}[Y(X+\delta) - Y(X)]$

(PS) $\forall (x, u) \in \mathrm{supp}(X, U)$, $(x+\delta, u) \in \mathrm{supp}(X, U)$ (A3.8)

(IS) $\forall y \in \mathcal{Y}$, $\forall (x, u) \in \mathrm{supp}(X, U)$,

$$\begin{aligned} &\mathrm{P}(Y(x+\delta) = y | X = x, U = u) \\ =&\mathrm{P}(Y(x+\delta) = y | X = x+\delta, U = u) \end{aligned} \tag{A3.9}$$

- Under the assumptions above
  but with (I1) replaced by (IS) and (P1) replaced by (PS),

$$\mu_\delta \equiv \mathrm{E}[Y(X+\delta)] = \mathrm{E}[\mathrm{E}[Y | X+\delta, S]]$$

Identifiability of shift intervention

$$\mu_\delta \equiv \mathrm{E}[Y(X+\delta)] = \mathrm{E}[\mathrm{E}[Y|X+\delta, S]]$$

Proof.

$$\begin{aligned}
\mathrm{E}[Y|X=x+\delta, S=s] &= \mathrm{E}[Y(x+\delta)|X=x+\delta, S=s] \\
&\underset{(IS)}{=} \mathrm{E}[Y(x+\delta)|X=x, S=s] \\
&= \mathrm{E}[Y(x+\delta)|S=s]
\end{aligned}$$

$$\Longrightarrow \mathrm{E}[Y(X+\delta)] = \mathrm{E}[\mathrm{E}[Y(X+\delta)|S]] = \mathrm{E}[\mathrm{E}[Y|X+\delta, S]]$$

??

■

Contents

Remaining parts

- Existing approaches
- Estimation/Inference
  - Spatial structure
  - CLT for $\mu_\delta \equiv \mathrm{E}[Y(X+\delta)]$
    - Doubly robust estimation
    - Double machine learning
- Simulations
- Data application

- It seems that the paper is still being revised. ??

Existing approaches to spatial confounding

(1) For a (non-random) function g (random S_i),

$$Y_i = \alpha + \beta X_i + g(S_i) + \varepsilon_i$$

(2) For a random field W (fixed s_i),

$$Y_i = \alpha + \beta X_i + W(s_i) + \varepsilon_i$$

- Spatial confounding is from $g(S_i)$ or $W(s_i)$, leaving ε_i to be independent

Spatial structure

- $\exists r > 0$ such that $|S_i - S_j| > r$ are (marginally) independent
- $\{S_i\}$ are sampled randomly from a geographic domain

- Smith (1980), Definition 2.
  A (regular) random field $W = \{W(s) : s \in \mathbb{R}^d\}$ is said to be *locally covariant* if $\exists r > 0$ such that
  - $\forall s, s' \in \mathbb{R}^d$ with $\mathrm{dist}(s, s') < r$, $\mathrm{cov}[W(s), W(s')] \geq 0$;
  - $\forall B, B' \subseteq \mathbb{R}^d$ with $\mathrm{dist}(B, B') \geq r$, $X_B \perp X_{B'}$.

- The paper probably considers the case when $U_i = W(s_i)$ for a locally covariant random field W and fixed locations $\{s_i\}$ in $\mathbb{R}^d$.

- Asymptotic scheme: increasing domain as $n \to \infty$??

Doubly robust estimation of $\mu_\delta \equiv \mathrm{E}[Y(X+\delta)]$

- $m(x,s) \equiv \mathrm{E}[Y|X=x, S=s]$, $\tau(x,s) \equiv f_{X|S}(x|S=s)$

$$\hat{\mu}_\delta \equiv n^{-1} \sum_{i=1}^{n} \frac{\hat{\tau}(X_i-\delta, S_i)}{\hat{\tau}(X_i, S_i)} \{Y_i - \hat{m}(X_i, S_i)\} + \hat{m}(X_i+\delta, S_i)$$

(1) $\hat{\tau}(X,S) - f(X|S) = o_{\mathrm{P}}(n^{-1/2})$, $\hat{m}(X,S) - g(X,S) = o_{\mathrm{P}}(n^{-1/2})$, $\exists g$

(2) $\hat{\tau}(X,S) - t(X|S) = o_{\mathrm{P}}(n^{-1/2})$, $\hat{m}(X,S) - m(X,S) = o_{\mathrm{P}}(n^{-1/2})$, $\exists t$

Doubly robust estimation of $\mu_\delta \equiv \mathrm{E}[Y(X+\delta)]$

Theorem 1

Under regularity conditions, if one of (1) and (2) holds, then $\sqrt{n}(\hat{\mu}_\delta - \mu_\delta) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2)$ *for some* $\sigma^2 \in (0, \infty)$.

- Similar definition/proof to Kennedy et al. (2017)
- $\hat{\mu}_\delta - \mu_\delta = V_n + R_{1n} + R_{2n}$
  - V_n: dominating/variance term converging to normal by spatial CLT (Smith, 1980)
  - R_{1n}, R_{2n}: negligible
  - Need to check more details due to spatial dependence ??

Double machine learning

- No exact description of DML that they use
- Probably, need more details ??

Data application

- X: ambient PM_2.5
- Y: birthweight

- U: greenspace

- Greenspace is a relatively smooth function of spatial location $\implies U_i = W(s_i)$
  - $W(s)$: peak normalized difference vegetation index (NDVI) measuraed at the census block level (from the year 2013)
  - s: (latitude, longitude) recorded in the Web Mercator projection

The End

THANK YOU