Kernel Single Proxy Control for Deterministic Confounding

Liyuan Xu, Arthur Gretton

Gatsby Unit

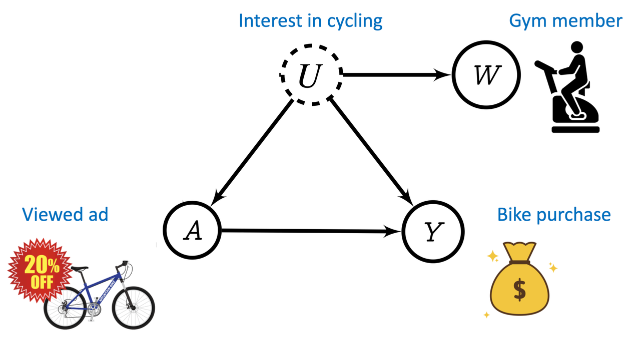
Abstract

- Causal effect estimation with an unobserved confounder from a single proxy variable that is associated with the confounder.
- Causal recovery is possible if the outcome is generated deterministically, even though it is shown to be impossible in general.
- Two kernel-based methods for handling nonlinear relationship between variables

Preliminaries

Causal Inference

- Y: Outcome (Bike Purchase)
- A: Treatment (Viewing Ads)
- U: Hidden Confounder (Interest in cycling)



Goal: Estimate structural function

$$f_{\mathrm{struct}}(\tilde{a}) \overset{\mathrm{def}}{=} \int \mathbb{E}[Y|A=\tilde{a},U]\,\mathrm{d}P(U)$$

Assumptions for Single Proxy Variable W

We observe additional proxy variable W (e.g. Gym membership)

(1): $W \perp\!\!\!\perp Y|U$ (2): W is informative for hidden confounder U

It is not possible to estimate f_{struct} from (W,A,Y) in general (Kuroki and Pearl [2014])

Single Proxy Causal Control

Consider binary treatment $A\in\{0,1\}$ and potential outcomes $U=(Y^{(0)},Y^{(1)})$, which give

$$Y = AY^{(1)} + (1-A)Y^{(0)}$$

Park et al. [2024]; Result 3

Let bridge function $b(w)$ be the solution of

$$y = \mathbb{E}[b(W)|A=0,Y=y] \quad \forall y.$$

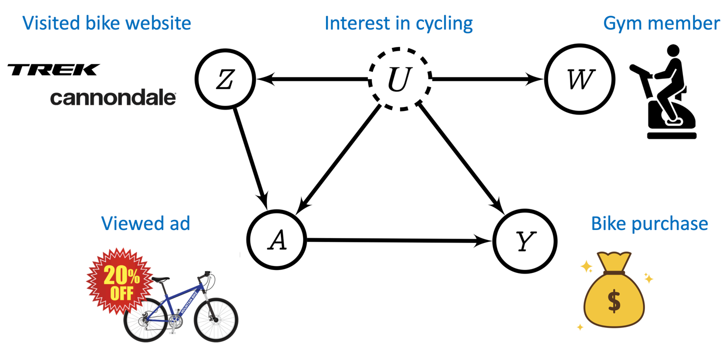
Then, if $W \perp\!\!\!\perp Y^{(0)}|A$, we have $f_{\mathrm{struct}}(0) = \mathbb{E}\left[Y^{(0)}\right] = \mathbb{E}_W[b(W)]$

Proxy Causal Learning with Two Proxies

Assume we have another proxy Z that;

(1) $Z \perp\!\!\!\perp Y|A,U$

(2) Z is informative for hidden confounder U



Miao et al. [2020]

Let bridge function $h(a,w)$ be the solution of

$$\mathbb{E}[Y|Z=z] = \mathbb{E}[h(a,W)|A=a,Z=z] \quad \forall a,z.$$

Then, we have $f_{\mathrm{struct}}(\tilde{a}) = \mathbb{E}[h(\tilde{a},W)]$.

Identification Results

Theorem 3.3 (informal)

Let bridge function $h(A,W)$ be the solution of

$$y = \mathbb{E}[h(a,W)|A=a,Y=y] \quad \forall a,y. \tag{1}$$

Then, we have $f_{\mathrm{struct}}(\tilde{a}) = \mathbb{E}[h(\tilde{a},W)]$ if

(1) Y is deterministic given A,U

(2) $W \perp\!\!\!\perp Y^{(a)}|A$ or Y is informative* for hidden confounder U

* informative means $\mathbb{E}[Y|a,U]$ is a.s. invertible

Connection to Single Proxy Control

- Y is deterministic given A and potential outcomes $U=(Y^{(0)},Y^{(1)})$
- Use the same assumption $W \perp\!\!\!\perp Y^{(a)}|A$

Generalization of single proxy control to other U

Connection to Proxy Causal Learning

- If Y is deterministic, $Y \perp\!\!\!\perp Y|A,U$
- We assume Y is informative for hidden confounder U

We can use outcome Y as "additional proxy Z"

Consistency Results

We may use any Proxy Causal Learning methods by replacing Z with Y.
We adopt two PCL methods from

- Single Kernel Proxy Learning (SKPV) (extending KPV (Mastouri et al. [2021]))
  Learn bridge function h by minimizing L2 deviation
  $$\hat{h} = \arg\min_{h\in\mathcal{H}} \mathbb{E}_{Y,A}\left[(Y-\mathbb{E}[h(A,W)|A,Y])^2\right]$$
  by solving two-stage regression.
- Single Proxy Maximum Moment Restriction (SPMMR) (cf. PMMR (Mastouri et al. [2021]))
  Learn bridge function h by solving suddle-point problem
  $$\hat{h} = \arg\min_{h\in\mathcal{H}} \max_{\|g\|_{\mathcal{H}}\le 1} (\mathbb{E}[(Y-h(A,W))g(A,Y)])^2.$$

Both methods have closed-form solutions,

where we improved the numerical stability of (Mastouri et al. [2021])

Consistency Results

Under certain regularity conditions, we have

$$\|\hat{h}_n - h_0\|_{\mathcal{H}} \to 0 \quad n\to 0$$

where $\hat{h}_n$ is SKPV/SPMMR estimation with n samples, and h_0 is true bridge function.

Sensitivity Analysis

What if we lift deterministic assumption?

Theorem 3.4 (Informal)

Assume that

- Bounded noise: $Y=\gamma_0(A,U)+\varepsilon$ where $\mathbb{E}[\varepsilon]=0,\ |\varepsilon|\le M$
- Y is (sufficiently) informative for U
  $$\sup_u |\ell(u)| \le \Xi \sup_y \mathbb{E}[\ell(U)|A=a,Y=y]$$
- Bridge function h_0 satisfying (1) exists.

Then, for all $a\in\mathcal{A}$, we have

$$|f_{\mathrm{struct}}(a) - \mathbb{E}_W[h(a,W)]| \le M\Xi$$

Violation of deterministic assumption only incurs bounded bias

Experiments

Data Generation Process

$$A=\Phi(U)+\varepsilon_1,\quad W=\exp(U)+\varepsilon_2,\quad Y=\sin(\pi U/2)+A^2-0.3$$

where $U\sim\mathrm{Unif}[-1,1], \varepsilon_1\sim\mathcal{N}(0,(0.1)^2), \varepsilon_2\sim\mathcal{N}(0,(0.05)^2)$

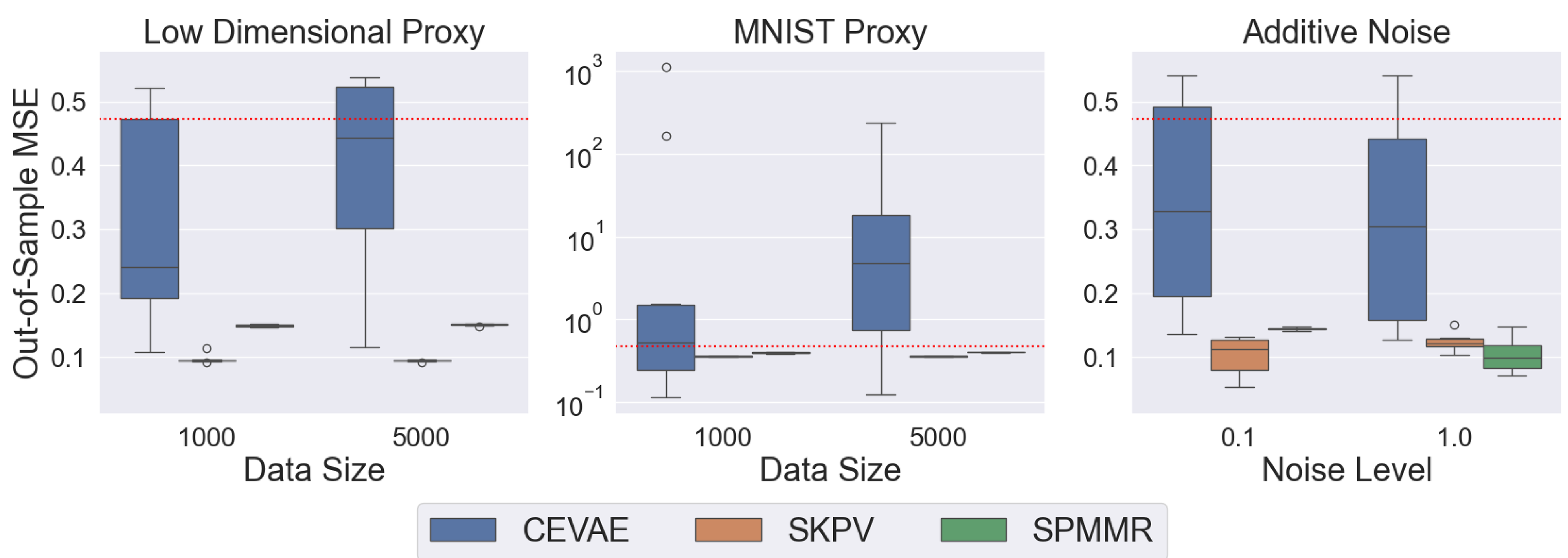
Three variants

- Low-dim proxy: Use original data generation process
- MNIST proxy: Use MNIST image as proxy W, where the digit label is chosen as $\lfloor 5U+5 \rfloor$
- Additive noise: Add Gaussian noise to Y with different variances

Baseline methods

- Kernel regression: Learn $\mathbb{E}[Y|A=a]$, ignoring confounding
- CEVAE Louizos et al. [2017]: Fit VAE model to learn U from observed (Y,A,W)

Results



- SKPV and SPMMR outperform baselines in all three settings
- The effect of violating deterministic assumption is minimal

Reference:

- Chan Park, David B Richardson, Eric Tchetgen Tchetgen. Single proxy control, Biometrics, 2024
- Wang Miao, Zhi Geng, Eric Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika, 2020
- Mastouri, Afsaneh, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt J. Kusner, Arthur Gretton and Krikamol Muandet. Proximal Causal Learning with Kernels: Two-Stage Estimation and Moment Restriction, ICML, 2021
- Christos Louizos, Uri Shalit, Joris M. Mooij, David Sontag, Richard Zemel, Max Welling. Causal Effect Inference with Deep Latent-Variable Models, Neurips, 2017