Doubly robust identification of treatment effects from multiple environments

Piersilvio de Bartolomeis[1], Julia Kostin[1], Javier Abad[1], Yixin Wang[2], Fanny Yang[1]

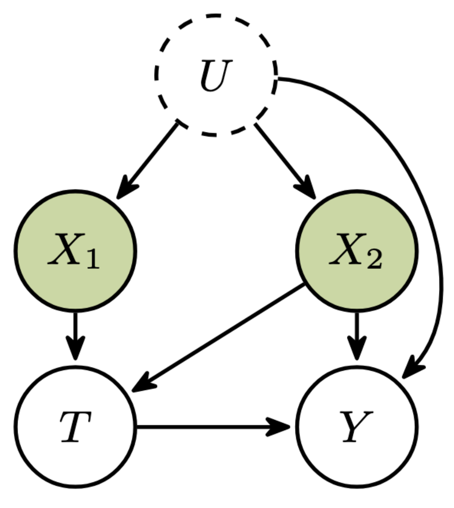
[1]ETH Zurich [2]University of Michigan

Problem setting

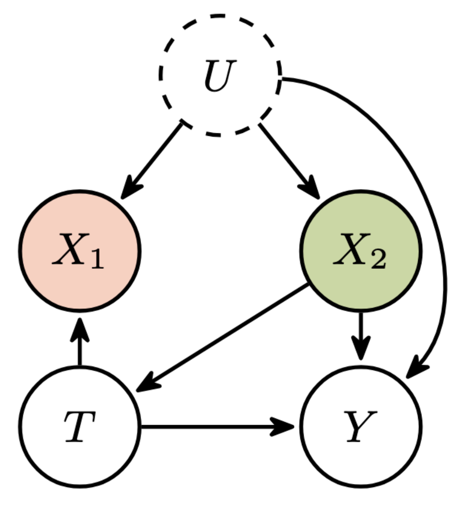
- **Identifying the average treatment effect (ATE)** from observational data is possible using a valid adjustment set S

$$\text{ATE} = \mathbb{E}\left[\mathbb{E}[Y \mid X_S, T=1] - \mathbb{E}[Y \mid X_S, T=0]\right]$$

- **Problem:** Adjusting for all available covariates often introduces **bias**, especially when **unobserved** and **post-treatment variables** are present.



$\{X_1, X_2\}$ good controls



X_2 good, X_1 bad control

We propose **R**obust **A**TE identification from **M**ultiple **EN**vironments (**RAMEN**), an algorithm that identifies the ATE by leveraging the **invariance** of either **treatment or outcome** mechanisms across multiple environments.

☑ In the presence of bad controls

☑ Without the need to learn the causal graph

☑ Via heterogeneity of multi-environment observational data

Prior work

Able to find a valid adjustment set when:		
Full causal **graph is known** • Vander Weele & Shipster 2011	Structural knowledge given such as **anchor variables** • Cheng et al. 2022; • Shah et al. 2022	Multiple **heterogeneous datasets** available • Shi et al. 2021 (leverages Y-invariance only)

Methodology

Goal: Identify treatment effects for different environments e:

$$\theta^e = \mathbb{E}_e[Y^{\mathrm{do}(T=1)} - Y^{\mathrm{do}(T=0)}]$$

Key assumption: existence of an invariant node V_{inv}. For all $e \in \mathscr{E}$ either

(a) All parents $\mathrm{Pa}(T)$ of T are observed, $\mathbb{E}_e[T \mid X_{\mathrm{Pa}(T)}]$ invariant;

(b) All parents $\mathrm{Pa}(Y)$ of Y are observed, $\mathbb{E}_e[Y \mid X_{\mathrm{Pa}(Y)}]$ invariant.

Population-level pipeline:

1. **Identify a subset** of covariates S and **invariant node** $V_{\mathrm{inv}} \in \{T, Y\}$ s.t. the regression function is invariant across environments, that is

$$\mathbb{E}_{pool}[V_{inv} \mid X_S] = \mathbb{E}_e[V_{inv} \mid X_S], \quad \mathbb{P}^e - \text{a.s.}$$

Enforce it via minimizing the **RAMEN** loss function:

$$S \in \operatorname*{argmin}_{V \in \{T,Y\}} \min_S J_S(X; V), \text{ where}$$

$$J_S(X;V) := \max_{e \in \mathscr{E}} \left[\sup_{h \in L^0(\mathbb{R}^{|S|})} \mathbb{E}_e\left[(V_{\mathrm{inv}} - \mathbb{E}_{pool}[V_{\mathrm{inv}} \mid X_S]) h(X_S)\right]\right]^2$$

is the **invariance penalty** over one of the nodes $V \in \{T, Y\}$.

2. **Estimate the ATE** via a classical doubly robust estimator.

$$\mathbb{E}^e\left[\mu_1(X_S) - \mu_0(X_S) + \frac{(Y - \mu_1(X_S))T}{\pi(X_S)} - \frac{(Y - \mu_0(X_S))(1-T)}{1 - \pi(X_S)}\right]$$

Where $\mu_t(X_S) = \mathbb{E}[Y \mid X_S, T=t]$ and $\pi(X_S) = \mathbb{E}[T \mid X_S]$.

Computational approach

Problem: Searching over all subsets of covariates introduces exponential complexity.

Solution: a **fully differentiable continuous relaxation of the subset search** via the **Gumbel trick**.

We select the covariates via $X_w = B(w) \circ X$,

where $B_j(w)$ is sampled randomly from a Bernoulli distribution with probability $\approx \mathrm{sigmoid}(w_j)$.

- The **Gumbel approximation** of Bernoulli sampling makes the loss differentiable w.r.t. the weights w_j.

Experiments

1. **Three synthetic scenarios** in which the unobserved variable U may break the invariances between T, Y and X.

(a) invariances preserved	(b) $X_p - T$ invariance broken	(c) $X_p - Y$ invariance broken