Consistency of Neural Causal Partial Identification

Joint work with Jose Blanchet and Vasilis Syrgkanis

Presenter: Jiyuan Tan (Stanford University)

The Thirty-Eighth Annual Conference on Neural Information Processing Systems

Introduction

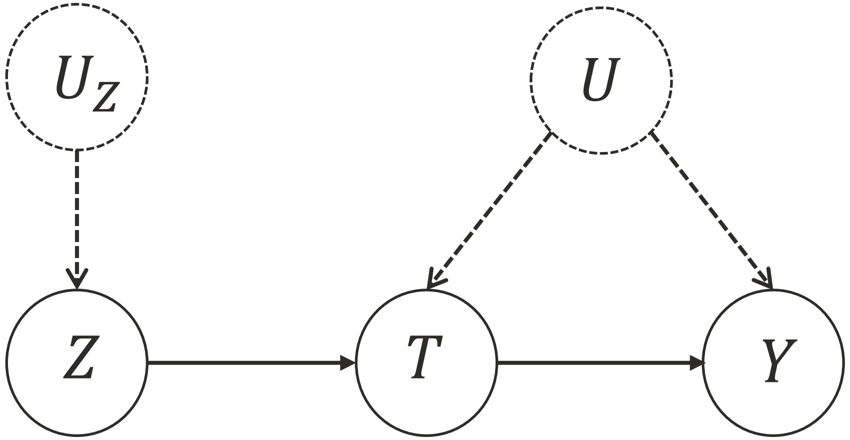
- We are interested in estimating counterfactual (aka interventional) quantities (e.g. ATE), assuming that the observed data is generated by a Structural Causal Model (SCM).
- In many situations, the counterfactual quantity is not point-identified.
- Partial Identification (PI) aims to obtain a tight bound to the quantity.
- For binary IV, bounds on ATE can be derived in closed-form [1,2]. For general SCMs with only discrete variables, PI can be encoded as a polynomial programming problem [3].
- Recent works in computer science provide a way to solve general PI problems with both discrete and continuous variables using generative models. However, these works are primarily empirical.
- **Our work:** Provide theoretical foundations of the neural causal approach for PI.

[1] Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In Uncertainty Proceedings 1994, pages 46–54. Elsevier, 1994.
[2] Manski, Charles F. “Nonparametric Bounds on Treatment Effects.” *The American Economic Review*, vol. 80, no. 2, 1990, pp. 319–23.
[3] Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings, September 2021.
[4] Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference, October 2022.

Background

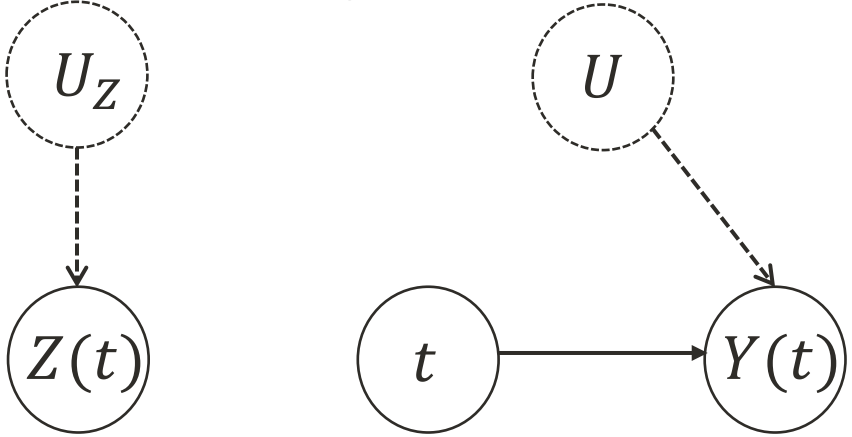
Observed data model:

$$Z = f_Z(U_Z),$$
$$T = f_T(Z, U),$$
$$Y = f_Y(T, U).$$



Interventional data model:
fix $T \to t$

$$Z(t) = f_Z(U_Z),$$
$$T(t) = t,$$
$$Y(t) = f_Y(t, U).$$



- Interventional variables $Y(t)$ are essentially potential outcomes in the potential outcomes framework
- **Goal.** Inference on interventional quantities from observation data.
- Example: Average Treatment Effect (ATE), $E[Y(t_1) - Y(t_0)]$

Problem Formulation: the IV Example

- The IV model is parametrized by structural functions f_Z, f_T, f_Y and latent distributions $P(U), P(U_Z)$.
- The partial identification problem can be formulated as following [1].

$$\max_{\substack{\text{structural functions: } f_Z, f_T, f_Y \\ \text{latent distributions: } P(U), P(U_Z)}} / \min \ \text{Counterfactual quantity, e.g. } E[Y(1) - Y(0)] \qquad (P)$$

$$s.t. \text{ implied distribution of } (Z, T, Y) = \text{observed distribution of } (Z, T, Y).$$

$$P\big(f_Z(U_Z), f_T(Z, U), f_Y(T, U)\big) = D^{observe}$$

---

[1] Vahid Balazadeh, Vasilis Syrgkanis, and Rahul G. Krishnan. Partial identification of treatment effects with implicit generative models, October 2022.

What about PI with general variables?

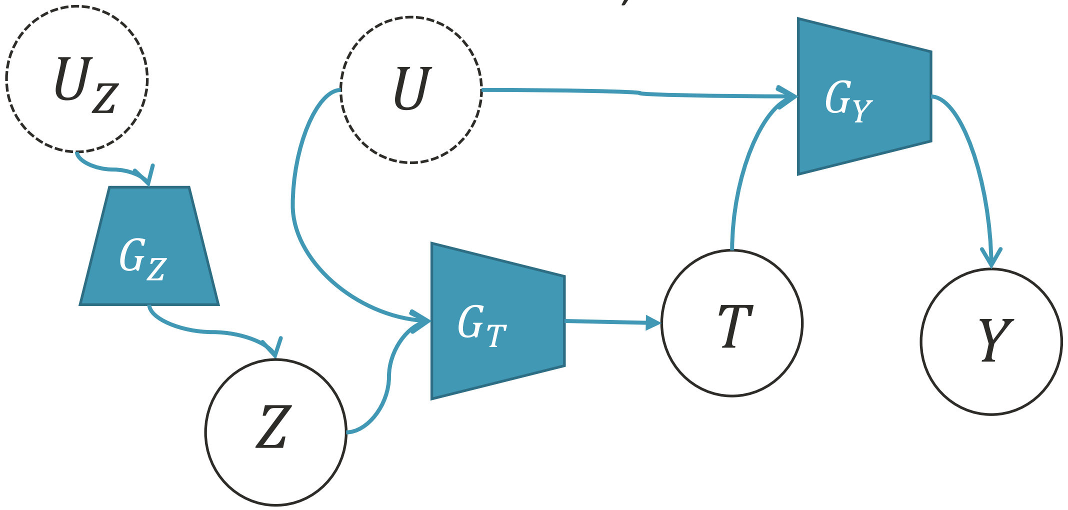
Neural-Causal PI: A candidate for General Purpose PI

- As pointed out by [1], an SCM can be viewed as a generative model.
- The intervention can be realized by changing the network structure.
- When all the functions in an SCM are neural networks and latent variables are uniform variables, it is called a Neural Causal Model (NCM).

[1]Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection:Expressiveness, learnability, and inference, October 2022.

Neural-Causal PI: A candidate for General Purpose PI

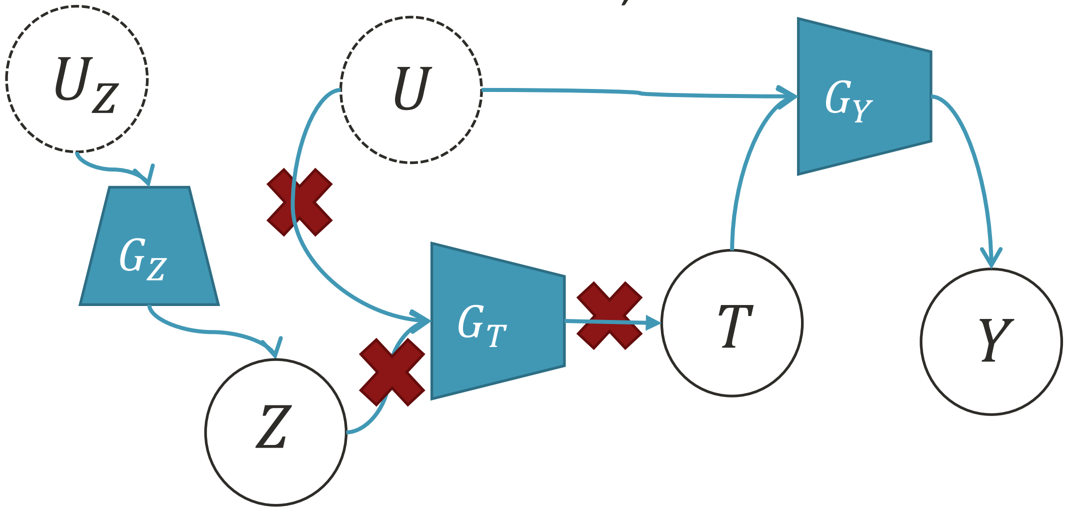
- As pointed out by [1], an SCM can be viewed as a generative model.
- The intervention can be realized by changing the network structure.
- When all the functions in an SCM are neural networks and latent variables are uniform variables, it is called a Neural Causal Model (NCM).



[1]Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection:Expressiveness, learnability, and inference, October 2022.

Neural-Causal PI: A candidate for General Purpose PI

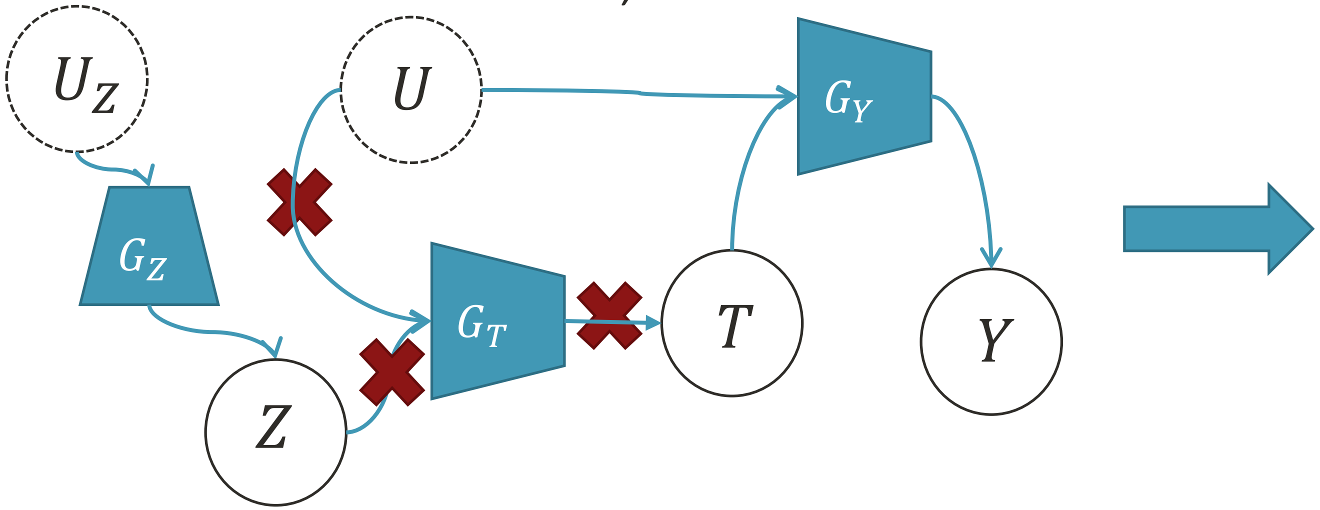
- As pointed out by [1], an SCM can be viewed as a generative model.
- The intervention can be realized by changing the network structure.
- When all the functions in an SCM are neural networks and latent variables are uniform variables, it is called a Neural Causal Model (NCM).



[1]Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection:Expressiveness, learnability, and inference, October 2022.

Neural-Causal PI: A candidate for General Purpose PI

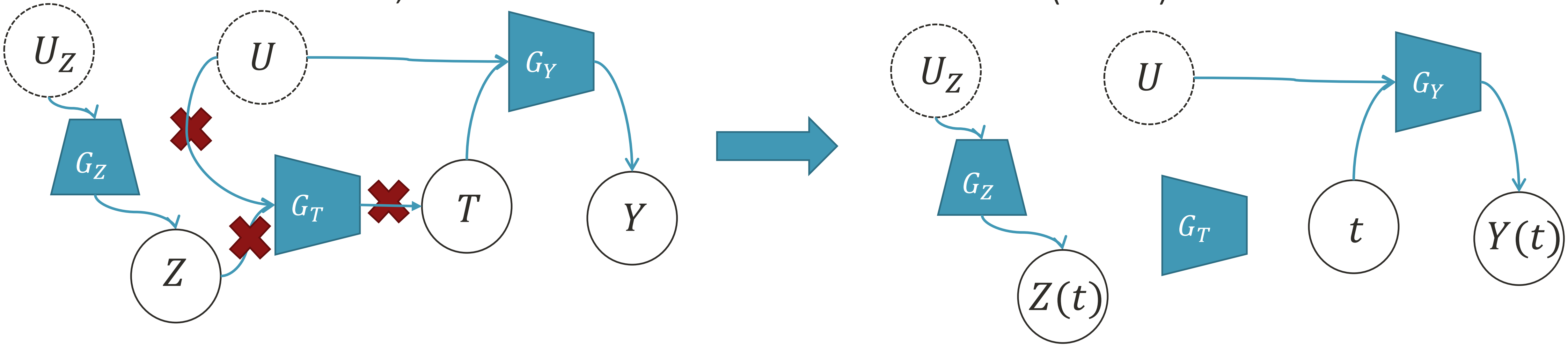
- As pointed out by [1], an SCM can be viewed as a generative model.
- The intervention can be realized by changing the network structure.
- When all the functions in an SCM are neural networks and latent variables are uniform variables, it is called a Neural Causal Model (NCM).



[1]Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection:Expressiveness, learnability, and inference, October 2022.

Neural-Causal PI: A candidate for General Purpose PI

- As pointed out by [1], an SCM can be viewed as a generative model.
- The intervention can be realized by changing the network structure.
- When all the functions in an SCM are neural networks and latent variables are uniform variables, it is called a Neural Causal Model (NCM).



[1]Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference, October 2022.

Problem Formulation II

- It is hard to search over all SCMs. As an alternative, we search over all NCMs.
- We consider the following empirical version of the problem

$$\max/\min_{\substack{\mathcal{M}: f_Z, f_T, f_Y,\\ P(U_Z), P(U)}} ATE_{\mathcal{M}}\,, \quad (P)$$

$$s.t. P^{\mathcal{M}}(Z,T,Y) = D^{observe}$$

Observation distribution of $\mathcal{M}$

$$\Rightarrow$$

$$\max/\min_{\substack{\mathcal{M}^{\theta}: f_Z^{\theta}, f_T^{\theta}, f_Y^{\theta},\\ P(\hat{U}_Z), P(\hat{U})}} ATE_{\mathcal{M}^{\theta}}\,, \quad (P_n)$$

$$s.t.\ d\left(P_m^{\mathcal{M}^{\theta}}(Z,T,Y), D_n^{observe}\right) \leq \alpha_n$$

Distance on the space of distributions

Empirical distribution of $P^{\mathcal{M}^{\theta}}$

Empirical distribution of $D^{observe}$

where the maximum/minimum is taken over all NCMs, θ is the parameter of the neural network and n is the sample size.

Is Neural-Causal PI consistent?

Two Open Problems

1. (Representation) Is the class of NCMs expressive enough to approximate general SCMs?

2. (Consistency) As sample size increases to infinity, do the optimal values of (P_n) converge to (P)?

Informal Results

- Theorem 2 (Informal). Under some regularity assumptions, NCMs can approximate Lipschitz SCMs arbitrarily well.

- Theorem 3 (Informal). Under some regularity assumptions, the neural causal partial identification method is consistent if we use Lipschitz regularization during training.

Paper

Thank You

Stanford University