Title: Bounding counterfactual distributions in discrete structural causal models

Speakers: Jin Tian

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 19, 2023 - 3:30 PM

URL: https://pirsa.org/23040118

Abstract: We investigate the problem of bounding counterfactual queries from an arbitrary collection of observational and experimental distributions and qualitative knowledge about the underlying data-generating model represented in the form of a causal diagram. We show that all counterfactual distributions in an arbitrary structural causal model (SCM) with finite discrete endogenous variables could be generated by a family of SCMs with the same causal diagram where unobserved (exogenous) variables are discrete with a finite domain. Utilizing this family of SCMs, we translate the problem of bounding counterfactuals into that of polynomial programming whose solution provides optimal bounds for the counterfactual query.

Counterfactuals

- Goal: inferring counterfactual queries from
  - observational/experimental data
  - causal diagram: qualitative knowledge about the underlying data-generating model
- E.g., investigating the gender discrimination in admission: "Would admission outcome for female applicant change had she been a male?"

Counterfactuals

- Goal: inferring counterfactual queries from
  - observational/experimental data
  - causal diagram: qualitative knowledge about the underlying data-generating model
- E.g., investigating the gender discrimination in admission: "Would admission outcome for female applicant change had she been a male?"

Partial Counterfactual Identifica.

- Identifying counterfactual distributions from data and causal diagram
  - (Halpern, 1998, Shpitser and Pearl 2007, Correa et al. 2021)
- Often non-identifiable
- Partial identification: deriving informative bounds (Manski, 1990; Robins,1989; Balke & Pearl, 1994; 1997; Tian & Pearl, 2000; Evans,2012; Richardson et al., 2014; Zhang & Bareinboim, 2017;Kallus & Zhou, 2018; Finkelstein & Shpitser, 2020; Kilbertus et al., 2020; Zhang & Bareinboim, 2021; ...)

Structural Causal Models

A structural causal model (SCM) is a tuple $\langle \boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, P \rangle$ where

- $\boldsymbol{V}$ is a set of endogenous variables
- $\boldsymbol{U}$ is a set of mutually independent exogenous variables
- $\mathcal{F}$ is a set of functions where each $f_V \in \mathcal{F}$ decides values of an endogenous variable $V \in \boldsymbol{V}$

$$v \leftarrow f_V(pa_V, u_V), PA_V \subseteq \boldsymbol{V}, U_V \subseteq \boldsymbol{U}$$

- $P(\boldsymbol{U})$ is a distribution

We consider acyclic SCMs.

Pearl's Causal Hierarchy - L1&2

- $\mathcal{F}$ can be seen as a mapping from $\boldsymbol{U} \longrightarrow \boldsymbol{V}$
- An SCM M induces distribution $P(\boldsymbol{V})$, called the *observational distribution*
- An intervention do$(\boldsymbol{X} = \boldsymbol{x})$ induces a submodel $M_{\boldsymbol{x}}$

$$x \leftarrow f_X(pa_X, u_X) \quad \text{replaced by} \quad x \leftarrow x$$

- $M_{\boldsymbol{x}}$ induces an *interventional distribution* $P(\boldsymbol{V}|\text{do}(\boldsymbol{x}))$
- Causal effect $P(y|do(x))$: how the outcome Y will respond if we take an action $X = x$

Pearl's Causal Hierarchy - L3

- The *potential response* $\boldsymbol{Y}_{\boldsymbol{x}}(\boldsymbol{u})$ is defined as the solution of $\boldsymbol{Y}$ in the submodel $M_{\boldsymbol{x}}$ given $\boldsymbol{U} = \boldsymbol{u}$.
- $P(\boldsymbol{U})$ induces a *counterfactual variable* $\boldsymbol{Y}_{\boldsymbol{x}}$

$$P\left(\boldsymbol{Y}_{\boldsymbol{x}} = y\right) = P(\boldsymbol{y}|\mathsf{do}(\boldsymbol{x})) = \int_{\Omega_{\boldsymbol{U}}} \mathbb{1}_{\boldsymbol{Y}_{\boldsymbol{x}}(\boldsymbol{u})=\boldsymbol{y}} dP(\boldsymbol{u})$$

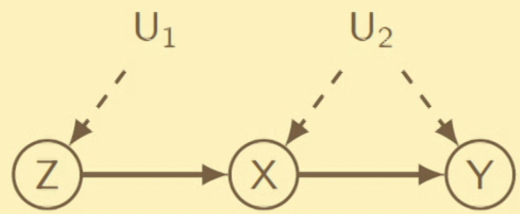
- A *counterfactual distribution*

$$P\left(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}}\right) = \int_{\Omega_{\boldsymbol{U}}} \mathbb{1}_{\boldsymbol{Y}_{\boldsymbol{x}}(\boldsymbol{u})=\boldsymbol{y}, \ldots, \boldsymbol{Z}_{\boldsymbol{w}}(\boldsymbol{u})=\boldsymbol{z}} dP(\boldsymbol{u})$$

Causal Diagram

- Every SCM M induces a causal diagram

$$z \leftarrow f_Z(u_1)$$
$$x \leftarrow f_X(z, u_2)$$
$$y \leftarrow f_Y(x, u_2)$$



- Researchers may know the scope of the functions, but not the details about the underlying mechanisms
- A graph is compatible with infinitely many SCMs

Causal Inference Tasks

- Model knowledge: causal diagram
- Data: $P(v)$, $P(v|do(z))$
- Query: $Q = P(y|do(x))$; $Q = P(y_x, z_w)$
- Develop CI algorithms:
  - Identifiable? $P(y|do(x)) = \sum_z P(y|x,z)P(z)$
  - Partially identifiable? $Q \in [a, b]$

CI by Optimization

- Task: given the observational distribution $P(v)$ in an arbitrary causal diagram $\mathcal{G}$, bound $P(y|do(x))$ or $P(Y_x, Z_w)$
- Assume endogenous variables $\boldsymbol{V}$ are discrete and finite
- let $\mathscr{M}(\mathcal{G})$ be the set of all SCMs associated with $\mathcal{G}$

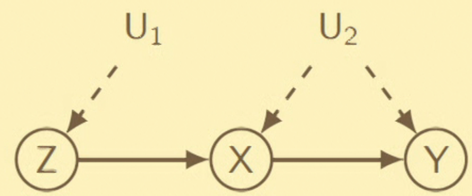
$$\min_{M \in \mathscr{M}(\mathcal{G})} / \max \quad P_M\left(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}}\right)$$
$$\text{s.t.} \quad P_M(\boldsymbol{V}) = P(\boldsymbol{V})$$

- Solving this optimization is difficult since we do not have access to the parametric forms of structural functions f_V nor $P(\boldsymbol{U})$

Canonical Partitioning (Balke&Pearl 94)

- IV model: binary X (treatment realized), Y (outcome), Z (treatment assigned)
- The *canonical functions* $x \leftarrow h_X^{(i)}(z)$

$$h_X^{(1)}(z) = 0, \quad \text{Never-taker}$$
$$h_X^{(2)}(z) = z, \quad \text{Complier}$$
$$h_X^{(3)}(z) = 1 - z, \quad \text{Defier}$$
$$h_X^{(4)}(z) = 1. \quad \text{Always-taker}$$

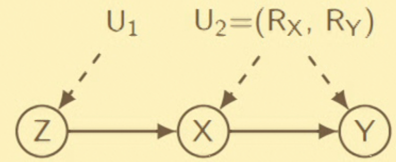


- For any $x \leftarrow f_X(z, u_2)$, there exists a *canonical partition* $\mathcal{U}_X^{(i)}, i = 1, 2, 3, 4$ over the domain of U_2 such that $u_2 \in \mathcal{U}_X^{(i)}$ if and only if $f_X(\cdot, u_2) = h_X^{(i)}$

Canonical IV Model (Balke&Pearl 94)

- A canonical IV model with $X, Y, Z \in \{0, 1\}$:
  $U_2 = (R_X, R_Y)$ where $R_X, R_Y \in \{1, 2, 3, 4\}$

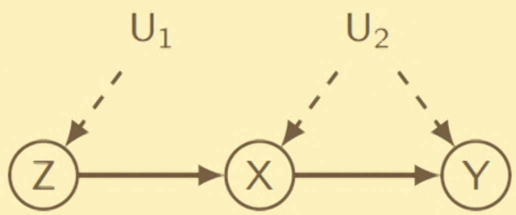
$$x \leftarrow f_X(z, R_X) = h_X^{(R_X)}(z),$$
$$y \leftarrow f_Y(x, R_Y) = h_Y^{(R_Y)}(x).$$



- For any IV model M there exists a canonical IV model N such that
  $P_M(x, y, z) = P_N(x, y, z)$, $P_M(y|do(x)) = P_N(y|do(x))$
  $P_M(Z, X_{z_0}, X_{z_1}, Y_{x_0}, Y_{x_0}) = P_N(Z, X_{z_0}, X_{z_1}, Y_{x_0}, Y_{x_0})$

CI by Optimization (Balke&Pearl 94)

- Task: given the observational distribution $P(x, y, z)$, bound the causal effect $P(y|do(x))$



- Parameters in a canonical IV model: $P(R_X, R_Y)$
- $P(x, y, z)$ imposes constraints on $P(R_X, R_Y)$
- Optimize $P(y|do(x))$ expressed in terms of $P(R_X, R_Y)$

LP Formulation (Balke&Pearl 94)

- 16 parameters

$$q_{jk} = P(R_X = j, R_Y = k)$$

- Express $p_{ij.k} = P(X = i, Y = j | Z = k)$ as linear functions of q_{jk}

$$p_{00.0} = P(Y = 0, X = 0 | Z = 0) = q_{00} + q_{01} + q_{10} + q_{11}$$

- Express objective function $P(y|do(x))$ as linear functions of q_{jk}

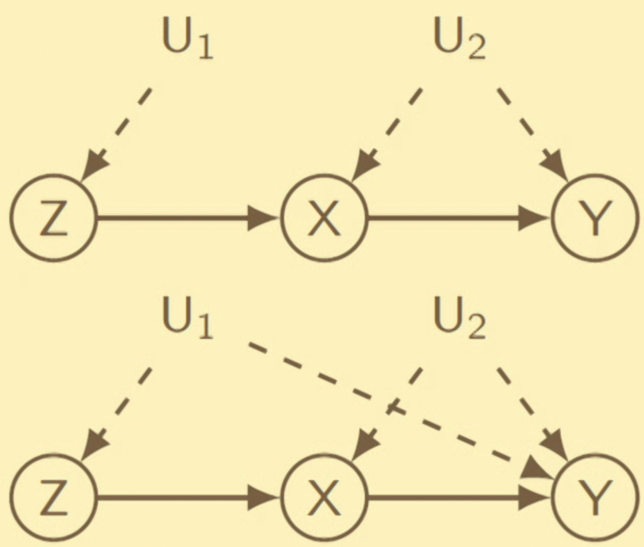
$$P(Y = 1 | do(X = 0))$$
$$= q_{02} + q_{12} + q_{22} + q_{32} + q_{03} + q_{13} + q_{23} + q_{33}$$

Beyond IV Model

- Goal: extend the canonical IV model to general causal diagrams
- Assume endogenous variables $\boldsymbol{V}$ are discrete and finite
- Evans et al. (2018) showed observational distributions in geared graphs could be generated by a model with exogenous variables of discrete domains
- Rosset et al. (2018), Fraser (2020) showed observational distribution in an arbitrary causal diagram can be generated by a model with finite-state exogenous variables

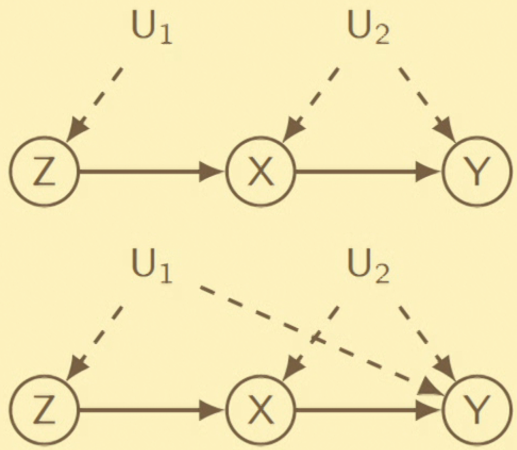
C-Components

Definition (C-component): Two endogenous variables are in the same c-component if and only if they are connected by a bidirected path, a path composed entirely of bi-directed edges.



C-Components

- We denote by $\boldsymbol{C}(U)$ the maximal c-component covering U in $\mathcal{G}$, i.e., $U \in \bigcup_{V \in \boldsymbol{C}(U)} U_V$.



$$\boldsymbol{C}(U_1) = \{Z\}$$
$$\boldsymbol{C}(U_2) = \{X, Y\}$$

$$\boldsymbol{C}(U_1) = \{X, Y, Z\}$$
$$\boldsymbol{C}(U_2) = \{X, Y, Z\}$$

Canonical SCMs

Definition: An SCM $M = \langle \boldsymbol{V}, \boldsymbol{U}, \mathscr{F}, P \rangle$ is said to be a canonical SCM over discrete endogenous $V \in \boldsymbol{V}$ with finite domain Ω_V if

- Every exogenous $U \in \boldsymbol{U}$ has a finite domain Ω_U with cardinality

$$|\Omega_U| = \prod_{V \in \boldsymbol{C}(U)} |\Omega_{PA_V} \mapsto \Omega_V|,$$

$|\Omega_{PA_V} \mapsto \Omega_V| = |\Omega_V|^{|\Omega_{PA_V}|}$ is the number of possible (canonical) functions mapping domain Ω_{PA_V} to Ω_V.

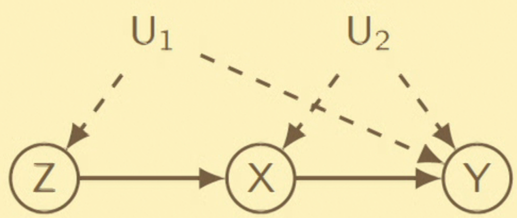
$$v \leftarrow f_V(pa_V, u_V)$$

Canonical SCMs

Theorem For an arbitrary SCM $M = \langle \boldsymbol{V}, \boldsymbol{U}, \mathscr{F}, P \rangle$ over discrete and finite endogenous variables $\boldsymbol{V}$, there exists a canonical SCM N such that

1. M and N are associated with the same causal diagram, i.e., $\mathcal{G}_M = \mathcal{G}_N$.
2. For any set of counterfactual variables $\boldsymbol{Y}_{\boldsymbol{x}}, \ldots, \boldsymbol{Z}_{\boldsymbol{w}}$, $P_M(\boldsymbol{Y}_{\boldsymbol{x}}, \ldots, \boldsymbol{Z}_{\boldsymbol{w}}) = P_N(\boldsymbol{Y}_{\boldsymbol{x}}, \ldots, \boldsymbol{Z}_{\boldsymbol{w}})$.

Example



$$d = |\Omega_{U_1}| = |\Omega_{U_2}| = 32$$

$$P(x_{z'}, y_{x'}) = \sum_{u_1,u_2=1}^{d} \mathbb{1}_{f_X(z',u_2)=x} \mathbb{1}_{f_Y(x',u_1,u_2)=y} P(u_1)P(u_2).$$

$$P(\boldsymbol{y_x}, \ldots, \boldsymbol{z_w}) = \sum_{U \in \boldsymbol{U}: u=1}^{d_U} \mathbb{1}_{\boldsymbol{Y_x}(\boldsymbol{u})=\boldsymbol{y},\ldots,\boldsymbol{Z_w}(\boldsymbol{u})=\boldsymbol{z}} \prod_{U \in \boldsymbol{U}} P(u).$$

Canonical SCMs - Int. Dist.

Theorem For an arbitrary SCM $M = \langle \boldsymbol{V}, \boldsymbol{U}, \mathscr{F}, P \rangle$ over discrete and finite endogenous variables $\boldsymbol{V}$, there exists a SCM N over $\boldsymbol{V}$ with discrete exogenous $\boldsymbol{U}$ having cardinality

$$|\Omega_U| = \prod_{V \in \boldsymbol{C}(U)} |\Omega_{PA_V}| \times |\Omega_V|$$

such that

1. M and N are associated with the same causal diagram, i.e., $\mathcal{G}_M = \mathcal{G}_N$.
2. M and N generate the same set of interventional distributions, i.e., for any subsets $\boldsymbol{X}, \boldsymbol{Y} \subseteq \boldsymbol{V}$, $P_M(\boldsymbol{Y}_{\boldsymbol{x}}) = P_N(\boldsymbol{Y}_{\boldsymbol{x}})$.

Canonical SCMs - Obs. Dist.

Theorem For an arbitrary SCM $M = \langle \boldsymbol{V}, \boldsymbol{U}, \mathscr{F}, P \rangle$ over discrete and finite endogenous variables $\boldsymbol{V}$, there exists a SCM N over $\boldsymbol{V}$ with discrete exogenous $\boldsymbol{U}$ having cardinality

$$|\Omega_U| = \prod_{V \in Pa(\boldsymbol{C}(U))} |\Omega_V|$$

such that

1. M and N are associated with the same causal diagram, i.e., $\mathcal{G}_M = \mathcal{G}_N$.
2. M and N generate the same observational distribution, i.e., $P_M(\boldsymbol{V}) = P_N(\boldsymbol{V})$.

This has been shown in (Rosset et al., 2018, Fraser 2020).

Bounding Counterfactuals

- Given a collection of observational and interventional distributions $\{P(\boldsymbol{V}|do(\boldsymbol{z})) \mid \boldsymbol{z} \in \mathbb{Z}\}$
- Qualitative assumption: causal diagram $\mathcal{G}$
- Assume endogenous variables $\boldsymbol{V}$ are discrete and finite
- Query: $P\left(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}}\right)$
- let $\mathscr{N}(\mathcal{G})$ be the set of all canonical SCMs associated with $\mathcal{G}$

$$\min_{N \in \mathscr{N}(\mathcal{G})} / \max \quad P_N\left(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}}\right)$$

$$\text{s.t.} \quad P_N(\boldsymbol{V}_{\boldsymbol{z}}) = P(\boldsymbol{V}_{\boldsymbol{z}}) \ \ \forall \boldsymbol{z} \in \mathbb{Z}$$

Polynomial Optimization

- For every $U \in \boldsymbol{U}$, let parameters $\theta_u = P(U = u)$

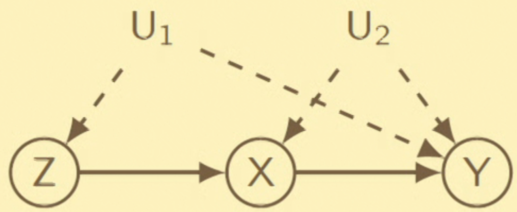
$$\theta_u \in [0, 1], \qquad \sum_{u \in \Omega_U} \theta_u = 1.$$

- For every $V \in \boldsymbol{V}$, we represent the output of function $f_V(pa_V, u_V)$ given input pa_V, u_V using an indicator vector $\mu_V^{(pa_V, u_V)} = \left(\mu_v^{(pa_V, u_V)} \mid \forall v \in \Omega_V\right)$ such that

$$\mu_v^{(pa_V, u_V)} \in \{0, 1\}, \qquad \sum_{v \in \Omega_V} \mu_v^{(pa_V, u_V)} = 1.$$

- Write any counterfactual probability as a polynomial function of parameters $\mu_v^{(pa_V, u_V)}$ and θ_u

Example



$$d = |\Omega_{U_1}| = |\Omega_{U_2}| = 32$$

- Objective function

$$P(x_{z'}, y_{x'}) = \sum_{u_1,u_2=1}^{d} \mu_x^{(z',u_2)} \mu_y^{(x',u_1,u_2)} \theta_{u_1} \theta_{u_2}$$

- Observational constraints

$$P(x, y, z) = \sum_{u_1,u_2=1}^{d} \mu_z^{(u_1)} \mu_x^{(z,u_2)} \mu_y^{(x,u_1,u_2)} \theta_{u_1} \theta_{u_2}$$

Quasi-Markovian Models

- Given an interventional distribution $P_{\boldsymbol{z}}(\boldsymbol{V}) = P(\boldsymbol{V}|do(\boldsymbol{z}))$
- Consider $G_{\overline{\boldsymbol{Z}}}$

$$\sum_{u_i=1}^{d_{u_i}} \prod_{V_j \in C_i} \mathbb{1}_{f_{V_j}(pa_j, u_i)=v_j} \theta_{u_i} = \prod_{V_j \in C_i} P_{\boldsymbol{z}}(v_j|pa_j^+)$$

- If the target query also factorizes, e.g, $\boldsymbol{W} = \{Y, \ldots, Z\}$

$$P(y_{pa_y}, \ldots, z_{pa_z}) = \prod_i \sum_{u_i=1}^{d_{u_i}} \prod_{V_j \in \boldsymbol{W} \cap C_i} \mathbb{1}_{f_{V_j}(pa_j, u_i)=v_j} \theta_{u_i}$$

Bayesian Approach

- Solving polynomial optimization problems is generally hard.
- Duarte et al. (2021) presented an algorithm for bounding causal effects given data
- We develop a MCMC algorithm to approximate the bound given finite data

Bayesian Approach

- Given i.i.d. samples $\bar{\boldsymbol{v}} = \left\{ \boldsymbol{V}^{(n)} \right\}_{n=1}^{N}$ from $\{P(\boldsymbol{V}|do(\boldsymbol{z})) \mid \boldsymbol{z} \in \mathbb{Z}\}$
- Query $\theta_{\text{ctf}} = P\left(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}}\right)$
- Prior: For every V, $\forall pa_V, u_V$, $\mu_V^{(pa_V, u_V)}$ are drawn uniformly over domain Ω_V. For every U, θ_u are drawn from a Dirichlet distribution
- Sample the posterior distribution $P\left(\theta_{\text{ctf}} \mid \bar{\boldsymbol{v}}\right)$ given data $\bar{\boldsymbol{v}}$
  1. Draw $(\boldsymbol{\mu}, \boldsymbol{\theta}) \sim P\left(\boldsymbol{\mu}, \boldsymbol{\theta} \mid \bar{\boldsymbol{v}}\right)$ by Gibbs sampling
  2. Given parameters $\boldsymbol{\theta}, \boldsymbol{\mu}$, compute the counterfactual probability $\theta_{\text{ctf}} = P(\boldsymbol{y}_{\boldsymbol{x}}, \ldots, \boldsymbol{z}_{\boldsymbol{w}})$

Estimating Bounds

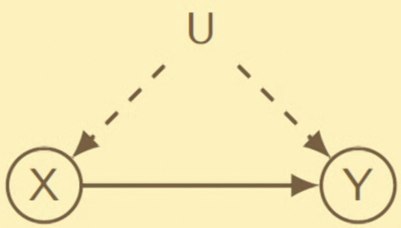
A $100(1-\alpha)\%$ *credible interval* $[l_\alpha, r_\alpha]$: any counterfactual probability θ_{ctf} that is compatible with observational data $\bar{\boldsymbol{v}}$ lies between the interval l_α and r_α with probability $1-\alpha$.

1. **Input:** Credible level α, tolerance level δ, ϵ.
2. **Output:** A credible interval $[l_\alpha, h_\alpha]$ for θ_{ctf}.
3. Draw $T = \lceil 2\epsilon^{-2} \ln(4/\delta) \rceil$ samples $\left\{\theta^{(1)}, \ldots, \theta^{(T)}\right\}$ from the posterior distribution $P\left(\theta_{\text{ctf}} \mid \bar{\boldsymbol{v}}\right)$.
4. Return interval $\left[\hat{l}_\alpha(T), \hat{r}_\alpha(T)\right]$.

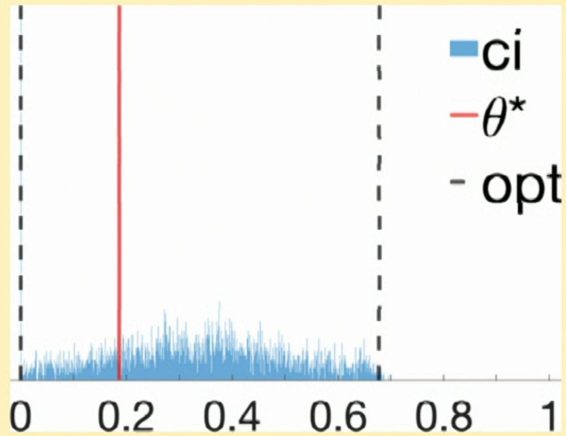
$$l_0 = \min_t \theta^{(t)}, \quad r_0 = \max_t \theta^{(t)}$$

$[l_0, r_0]$ converges almost surely to the optimal bound as $N, T \to \infty$

Simulation Example

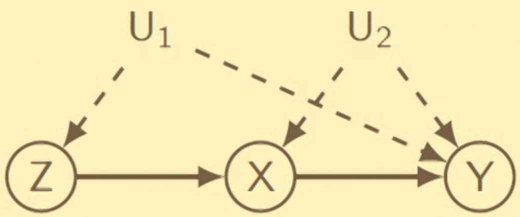


$X, Y \in \{0, 1\}$
$d = |\Omega_U| = 8$
Data: $P(X, Y)$
Query: PNS
$PNS = P(Y_{x=1} = 1, Y_{x=0} = 0)$



$N = 10^3$
$T = 4 \times 10^3$
opt: sharp bound(Tian&Pearl 2000)

Simulation Example



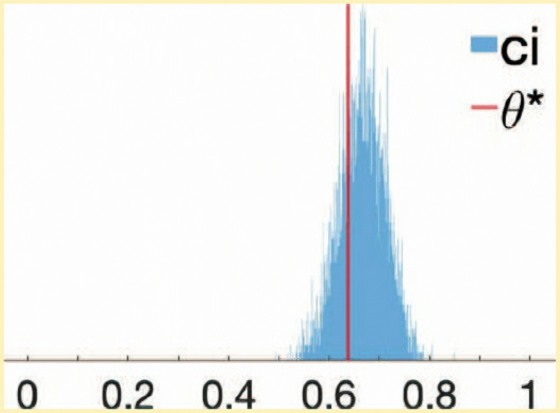
$X, Y, Z \in \{0, \ldots, 9\}$

$d = |\Omega_{U_1}| = |\Omega_{U_2}| = 10^{21}$

Data: $P(X, Y, Z)$,

$P(X_z, Y_z)$ for $z = 0, \ldots, 9$

Query $P(z, x_{z'}, y_{x'})$



$N = 10^3$

$T = 4 \times 10^3$

$\theta = P(Z + X_{z=0} + Y_{x=0} \geq 14)$

Conclusion

- We study the problem of bounding counterfactual probabilities from observational and experimental data given a causal diagram
- We introduce a family of canonical SCMs over discrete endogenous variables with discrete exogenous variables
- We show canonical SCMs could represent *all* counterfactual distributions over discrete observed variables in *any* causal diagram.
- We reduce the partial identification problem into a polynomial program
- We develop an MCMC algorithm to approximate the optimal bounds from finite samples.