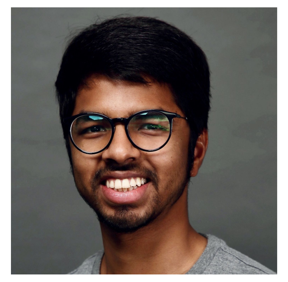
Front-door Adjustment Beyond Markov Equivalence with Limited Graph Knowledge

arXiv Link

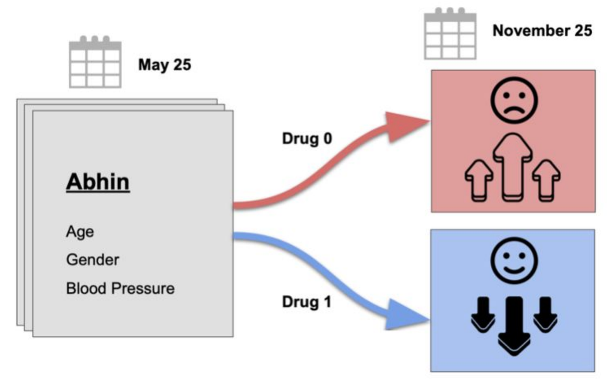

Abhin Shah
MIT


Karthikeyan Shanmugam
Google Research, India


Murat Kocaoglu
Purdue University

Causal Effect Estimation

Causal effect of a drug on cholesterol level from observational data



$\mathbb{P}(\text{cholesterol} \mid do(\text{drug}))$?

Observational Data

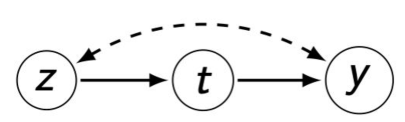
Age	Gender	Blood Pressure	Drug	Cholesterol (0)	Cholesterol (1)
22	Male	145/95	0	⇧⇧⇧	?
26	Female	135/80	0	⬇⬇⬇	?
58	Female	130/70	1	?	⇧⇧⇧
24	Female	150/85	1	?	⇧⇧⇧

Observed features **x** — Treatment t — Outcome y

Causal Estimands With Known Graph

Given the complete knowledge of the causal graph,
a complete algorithm was proposed for causal effect estimation [1,2]

Special Case 1 — Back-door Criterion

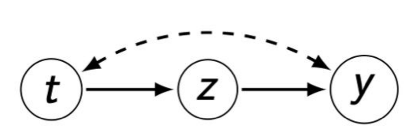


A set **z** satisfies the back-door criterion if

1. all paths between t and y containing an arrow into t (i.e., back-door paths) are blocked by **z**
2. **z** does not contain any descendant of t

$$\mathbb{P}(y \mid do(t=t)) = \sum_{\mathbf{z}} \mathbb{P}(y \mid t=t, \mathbf{z}=z)\mathbb{P}(\mathbf{z}=z)$$

Special Case 2 — Front-door Criterion



A set **z** satisfies the front-door criterion if

1. **z** blocks all directed paths from t to y
2. all back-door paths between t and **z** are blocked
3. all back-door paths between **z** and y are blocked by t

$$\mathbb{P}(y \mid do(t=t)) = \sum_{\mathbf{z}} \Big(\sum_{t'} \mathbb{P}(y \mid t=t', \mathbf{z}=z)\mathbb{P}(t=t')\Big)\mathbb{P}(\mathbf{z}=z \mid t=t)$$

Causal Estimands Without Graph

Given partial ancestral graphs (PAGs), which can be learned from observational data, complete algorithms were proposed for causal effect estimation [3,4]

PAGs — A collection of causal graphs that encode the same conditional independence relations

Drawbacks —

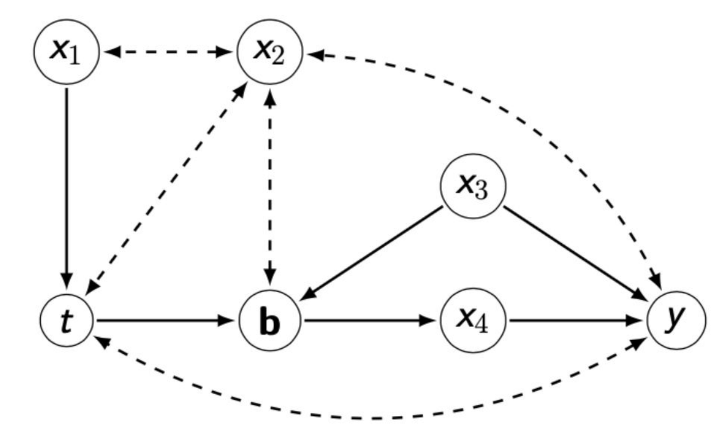
A. Only applicable if every causal graph in the collection has the same causal estimand
B. Sequentiality in learning PAGs propagates errors between tests.
C. PAG learning cannot incorporate side information about the graph.

Causal Estimands With Limited Graph Knowledge?

How much of the causal graph $\mathcal{G}$ do we need to know?

[5,6,7,8] — How to identify back-door-like sets using only partial structural information?

This work — How to identify front-door-like sets using only partial structural information?



The knowledge of all the children of the treatment is sufficient!

e.g., $t \longrightarrow \mathbf{b}$ in the above graph

Assumptions — Semi-Markovian Model

1. y is a descendant of t.
2. There is an unobserved confounder (denoted by a bi-directed arrow) between y and t.

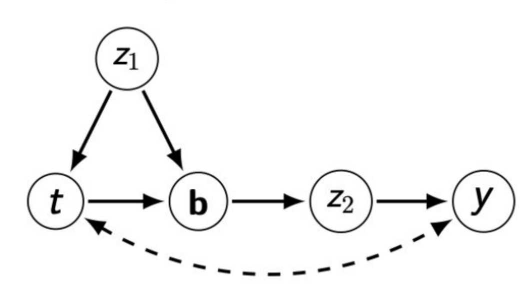
Sufficient Conditions For Causal Identifiability

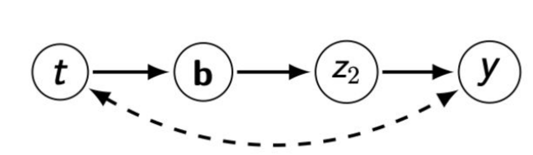
- $\mathbf{b}$: the set of all children of t.
- Consider any subset of the remaining observed features i.e., $\mathbf{z} \subseteq \mathbf{x} \backslash \{\mathbf{b}\}$.
- $\exists\, \mathbf{z}$ such that $\mathbf{b} \perp y \mid t, \mathbf{z} \implies \mathbb{P}(y \mid do(t=t))$ is identifiable from observational data

e.g., $\mathbf{z} = (x_3, x_4)$ in $\mathcal{G}_{toy}$

Proof Idea — Show that there is no bi-directed path between t and $\mathbf{b}$ in $\mathcal{G}^{\text{Ancestor}(y)}$ + [9]

However, $\mathbf{b} \perp y \mid t, \mathbf{z}$ alone is insufficient to establish a unique causal formula.





These graphs satisfy $\mathbf{b} \perp y \mid t, \mathbf{z}$ but have different causal formula.

Generalized Front-door Criterion

- Two additional conditional independencies imply a unique causal formula!
- If **z** satisfying $\mathbf{b} \perp y \mid t, \mathbf{z}$ can be decomposed into $\mathbf{z}^{(o)} \subseteq \mathbf{z}$ and $\mathbf{z}^{(i)} = \mathbf{z} \backslash \mathbf{z}^{(o)}$ such that

$$\mathbf{z}^{(i)} \perp t \qquad \text{and} \qquad \mathbf{z}^{(o)} \perp t \mid \mathbf{z}^{(i)}, \mathbf{b}$$

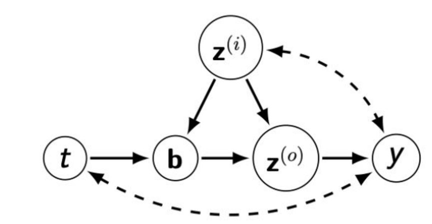
then, **z** and $\mathbf{s} \triangleq (\mathbf{b}, \mathbf{z}^{(i)})$ are generalized front-door sets, i.e.,

$$\mathbb{P}(y \mid do(t=t)) = \sum_{\mathbf{z}} \Big(\sum_{t'} \mathbb{P}(y \mid t=t', \mathbf{z}=z)\mathbb{P}(t=t')\Big)\mathbb{P}(\mathbf{z}=z \mid t=t)$$

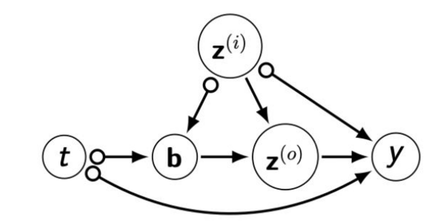
$$\mathbb{P}(y \mid do(t=t)) = \sum_{\mathbf{s}} \Big(\sum_{t'} \mathbb{P}(y \mid t=t', \mathbf{s}=s)\mathbb{P}(t=t')\Big)\mathbb{P}(\mathbf{s}=s \mid t=t)$$

e.g., $\mathbf{z}^{(i)} = x_3$ and $\mathbf{z}^{(o)} = x_4$ in $\mathcal{G}_{toy}$

Relation To PAG-based Algorithms

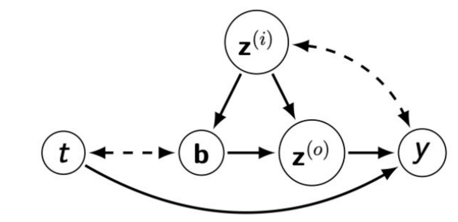


Graph $\mathcal{G}_1$ where our method applies



PAG corresponding to graphs $\mathcal{G}_1$ and $\mathcal{G}_2$

IDP algorithm fails here!



Graph $\mathcal{G}_2$ which is Markov equivalent to $\mathcal{G}_1$

Algorithm For ATE Estimation

Initialization: $\text{ATE}_z = 0, \text{ATE}_s = 0, c_1 = 0$
for $r = 1, \cdots, n_r$ **do** // Use a different train-test split in each run
  $\text{ATE}_z^r = 0,\ \text{ATE}_s^r = 0,\ c_2 = 0;$
  for $\mathbf{z} \in \mathcal{Z}$ **do**
    if $CI(\mathbf{b} \perp_p y \mid \mathbf{z}, t) > p_v$ **then** // where CI stands for conditional independence
      for $\mathbf{z}^{(o)} \subseteq \mathbf{z}$ **do**
        $\mathbf{z}^{(i)} = \mathbf{z} \setminus \mathbf{z}^{(o)};$
        if $\min\{CI(\mathbf{z}^{(i)} \perp_p t), CI(\mathbf{z}^{(o)} \perp_p t \mid \mathbf{b}, \mathbf{z}^{(i)})\} > p_v)$ **then**
          $c_2 = c_2 + 1,\ \mathbf{s} = (\mathbf{b}, \mathbf{z}^{(i)});$
          $\text{ATE}_z^r = \text{ATE}_z^r + \frac{\sum_{j:t_j=1}\sum_{t'} \mathbb{E}[y \mid \mathbf{z}_j, t']\mathbb{P}(t')}{|\{j:t_j=1\}|} - \frac{\sum_{j:t_j=0}\sum_{t'} \mathbb{E}[y \mid \mathbf{z}_j, t']\mathbb{P}(t')}{|\{j:t_j=0\}|};$
          $\text{ATE}_s^r = \text{ATE}_s^r + \frac{\sum_{j:t_j=1}\sum_{t'} \mathbb{E}[y \mid \mathbf{s}_j, t']\mathbb{P}(t')}{|\{j:t_j=1\}|} - \frac{\sum_{j:t_j=0}\sum_{t'} \mathbb{E}[y \mid \mathbf{s}_j, t']\mathbb{P}(t')}{|\{j:t_j=0\}|};$
  if $c_2 > 0$ **then**
    $\text{ATE}_z = \text{ATE}_z + \text{ATE}_z^r/c_2,\ \text{ATE}_s = \text{ATE}_s + \text{ATE}_s^r/c_2,\ c_1 = c_1 + 1;$
if $c_1 > 0$ **then**
  $\text{ATE}_z = \text{ATE}_z/c_1, \text{ATE}_s = \text{ATE}_s/c_1;$
else
  Failed to find $\mathbf{z} = (\mathbf{z}^{(i)}, \mathbf{z}^{(o)})$ satisfying (4) and (5);

Applicability To Random Graphs

- p — #observed variables
- d — expected in-degree
- q — controls #unobserved variables

Let $y = v_p$; t = any non-parent and non-grand parent ancestor of y; $\mathbf{b}$ = all children of t

	$p=10$			$p=15$		
	$d=2$	$d=3$	$d=4$	$d=2$	$d=3$	$d=4$
$q=0.0$	(43, 43)	(20, 20)	(21, 21)	(27, 26)	(9, 9)	(4, 2)
$q=0.5$	(23, 23)	(16, 16)	(7, 7)	(18, 17)	(4, 3)	(0, 0)
$q=1.0$	(6, 6)	(4, 4)	(5, 5)	(9, 9)	(10, 9)	(0, 0)

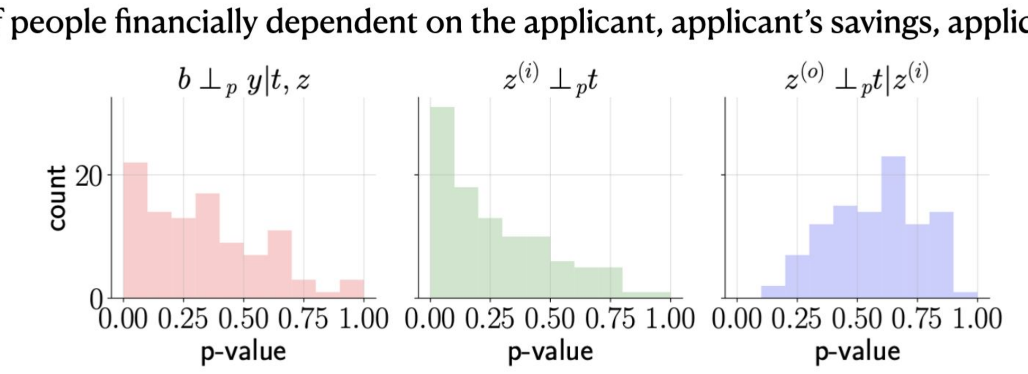
IDP gives 0 success out of 100 across for various p, d, and q

Fairness Application — German Credit Dataset

Compute the total effect of the sensitive attribute t on the outcome y

y = binary credit risk; t = applicant's age (binarized);

$\mathbf{b}$ = # of people financially dependent on the applicant, applicant's savings, applicant's job



Total effect$_z$ $= 0.0125 \pm 0.0011$, and Total effect$_s$ $= 0.0105 \pm 0.0018$

1. Shpitser and Pearl, "Identification of joint interventional distributions in recursive semi-Markovian causal models," *AAAI 2006*
2. Huang and Valtorta, "Pearl's calculus of intervention is complete," *UAI 2006*
3. Perkovic et al., "Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs," *JMLR 2018*
4. Jaber et al., "Causal identification under Markov equivalence: Completeness results," *ICML 2019*
5. Entner et al., "Data-driven covariate selection for nonparametric estimation of causal effects," *AISTATS 2013*
6. Gultchin et al., " Differentiable causal backdoor discovery," *AISTATS 2020*
7. Shah et al., "Finding valid adjustments under non-ignorability with minimal dag knowledge," *AISTATS 2022*
8. Cheng et al., "Toward Unique and Unbiased Causal Effect Estimation From Data With Hidden Variables", *IEEE Transactions on Neural Networks and Learning Systems 2023*
9. Tian and Pearl, "A general identification condition for causal effects," *AAAI 2002*