Estimating Causal Effects Using Weighting Based Estimators

Yonghan Jung
Jin Tian
Elias Barenboim

AAAI 2020

Causal Model

G: DAG

CPT:

$\Pr[u \mid pa(u)]$

G encodes: 1) CI between variables

2) Also causal relationships

1)
$X \rightarrow Y \rightarrow Z$
$X \leftarrow Y \rightarrow Z$
$X \leftarrow Y \leftarrow Z$
} $X \perp\!\!\!\perp Z \mid Y$

$X \rightarrow Y \leftarrow Z$ } have no CI constraint

2) $P(Y=y \mid do(X=x))$

$\equiv$ remove incoming edges to X and fix it to x; others follow CPT

$$X \rightarrow Y \rightarrow Z \quad \Rightarrow \quad \uparrow X \rightarrow Y \rightarrow Z$$

$$X \leftarrow Y \rightarrow Z \quad \Rightarrow \quad x \quad Y \rightarrow Z$$

$$X \leftarrow Y \leftarrow Z \quad \Rightarrow \quad x \quad Y \leftarrow Z$$

$$P(y \mid do(x)) \equiv P(y \mid \hat{x}) \equiv P_x(y)$$

With hidden variables

G: Acyclic Directed Mixed Graph

(certain variables cannot be observed)

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left, dashed, leftrightarrow] & Z \arrow[r] & Y \end{tikzcd} \qquad \text{(ADMG)}$$

$$\equiv \quad \begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ X \arrow[r] & Z \arrow[r] & Y \end{tikzcd} \qquad \text{(DAG)}$$

Identifiability

$$\underbrace{P(S \mid do(T))}_{\text{interventional distr.}} \overset{?}{=} f(\underbrace{P}_{\text{observable distribt.}})$$

$$T = t$$

Given G (ADMG), Can be determined

in polytime.

Estimation: Suppose we know its identifiable.

Can we determine $P(S \mid do(T))$

from finitely many samples from $\mathbb{P}$?

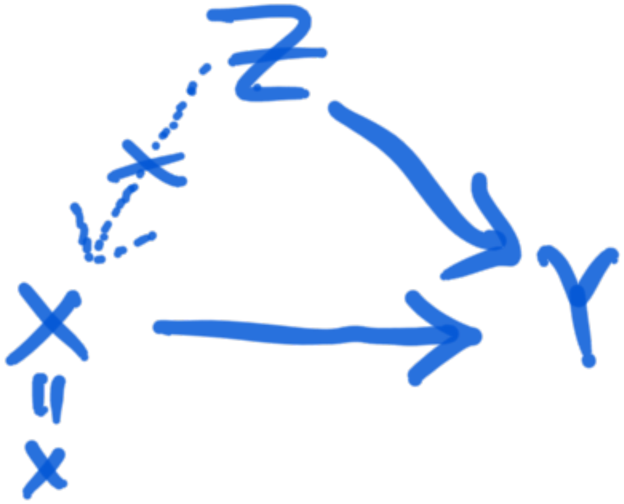
Example:

$$
\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}
$$

$$P(y \mid do(x)) = ?$$

$$P(y \mid do(x)) = \sum_{z} P(y, z \mid do(x))$$

$$= \sum_{z} P(y|z, do(x)) \cdot P(z|do(x))$$



$$= \underbrace{\sum_{z} P(y|z,x) \quad P(z)}_{\text{Back-door adjustment.}}$$

$$\therefore \mathbb{E}[Y|do(x)] = \sum_{z} \mathbb{E}[Y|z,x] \, P(z)$$

In practice, Z is often high dimensiona

How to estimate the above quantity?

<u>Solution</u>: Define $R(x,y,z) \propto P(x,y,z) \cdot \frac{P(x)}{P(x|z)}$

$P[x|z]$

$$\propto P(y|x,z)\, P(x)\, P(z)$$

then $$\boxed{R(y|x) = P(y|do(x))}$$

proof:

$$R(x,y) = \sum_z \underbrace{P(y|x,z)\, P(x)\, P(z)}$$

$$R(x) = P(x).$$

Estimating $E[Y|x]$ from samples of a distribution is known (using least squares regression or its generalizations).

$$R(x,y,z) = P(x,y,z) \frac{P(x)}{P(x|z)}$$

multi step process:

$w(x,y,z)$

1) estimate $\frac{\hat{P}(x)}{\hat{P}(x|z)} \leftarrow \hat{w}(x,y,z)$

2) make $\underline{\hat{w}(x,y,z)}$ copies of each

$P \rightarrow R$

Sample

this creates a pseudo-population from R

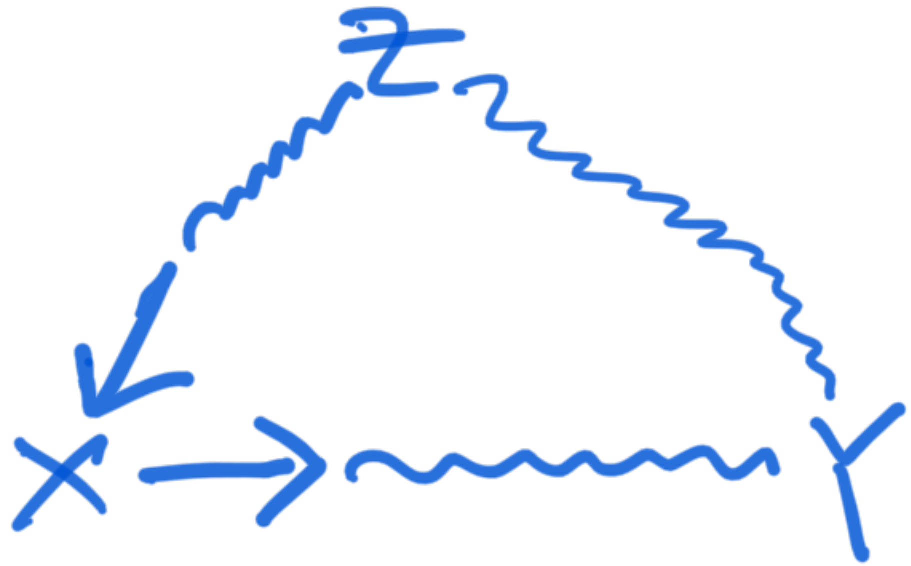
3) use least square regression over pseudo-population

"Inverse Probability of Treatment Weighting"

This paper generalizes IPTW for other

graphs. ← gives such weights we

<u>Generalization 0</u>: BD Criterion.



- any path (undirected) between X Y that goes into X is a "<u>back door</u>" path.
- A variable set <u>Z blocks</u> all paths between sets X & Y if

fork: $X \rightsquigarrow \leftarrow Z \rightarrow \rightsquigarrow Y$

chain: $X \rightsquigarrow \rightarrow Z \rightarrow \rightsquigarrow Y$

chain: $X \rightsquigarrow \leftarrow Z \leftarrow \rightsquigarrow Y$

Z <u>includes</u> such a z or,

collider: $X \rightsquigarrow \rightarrow Z \leftarrow \rightsquigarrow Y$

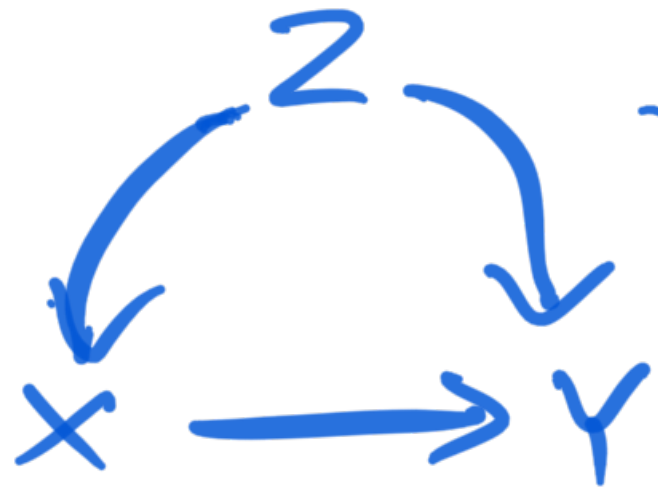
Z <u>excludes</u> such a z

All paths between X, Y must be blocked by Z in this manner.

<u>**Theorem**</u>: If Z <u>blocks</u> all <u>back-door</u>

paths between X & Y then

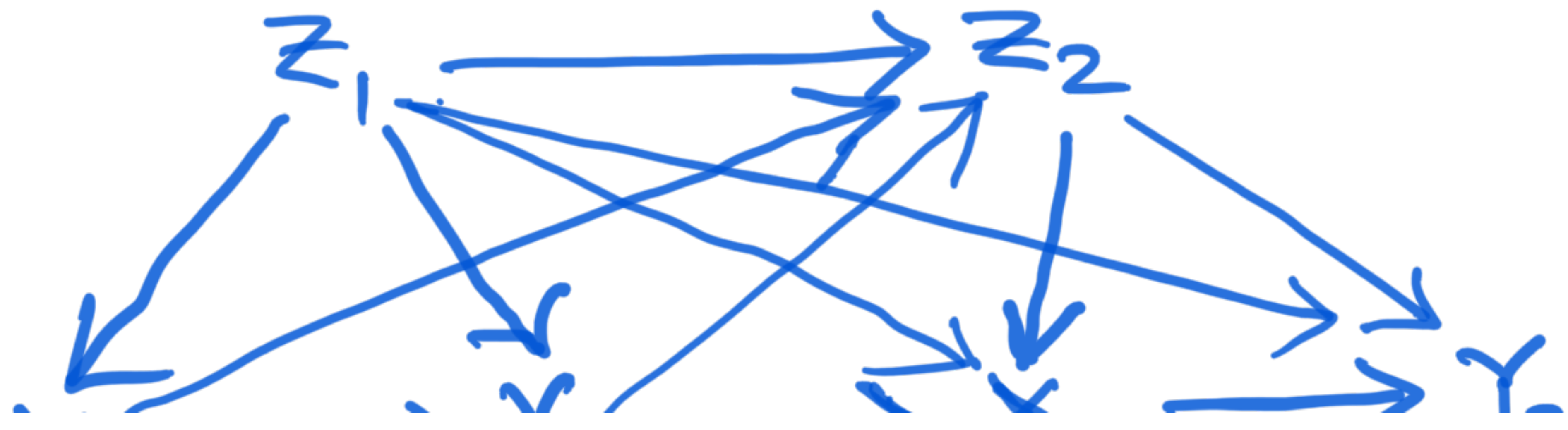
$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z)\, P(z)$$



Generalization 1: MSBD

M → Multi-outcome

S → sequential



$$\begin{tikzcd}
X_1 \arrow[r] \arrow[rr, bend right] \arrow[rrr, bend right] & Y_1 \arrow[r] \arrow[rr, bend right] & X_2 \arrow[r] & Y_2
\end{tikzcd}$$

$$P(y_1, y_2 \mid do(x_1), do(x_2)) = ?$$

Theorem: Consider sequence of

treatments: $X_1, X_2, \ldots, X_n$

Outcomes: $Y_1, Y_2, \ldots, Y_n$

confounders: $Z_1, Z_2, \ldots, Z_n$.

Such that (1) Z_i's are not descendents of $X^{\geq i}$

(2) $x^{\geq i}$ ... $(i-1)$ (i) $(i-1)$

② $Y \perp\!\!\!\perp X_i \mid Y, Z, X$

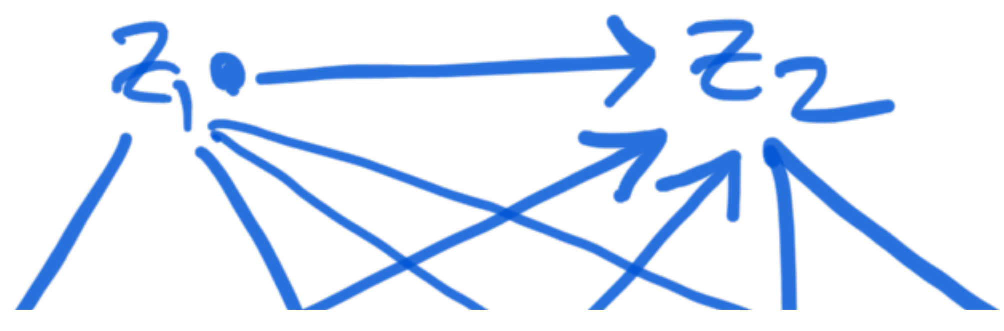
in $G_{\underline{X_i}\,\overline{X^{\geq i+1}}}$

then $P(y \mid do(x))$ is identifiable.

and is given by

$$P_x(y) = \sum_{z} \prod_k P(y_k \mid x^{(k)}, z^{(k)}, y^{(k-1)}) \times \prod_k P(z_j \mid x^{(j-1)}, z^{(j-1)}, y^{(j-1)})$$

vector.



$$P[y_2 \mid \hat{x}_1, \hat{x}_2] = \sum_{z_1, z_2, y_1} P(y_1 \mid x_1, z_1)\, P(y_2 \mid x_1, x_2, y_1, z_1, z_2)\, P(z_1) \cdot P(z_2 \mid x_1, y_1, z_1)$$

$$\mathbb{E}[Y_2 \mid \hat{x}_1, \hat{x}_2] = \sum_{z_1, z_2, y_1} \mathbb{E}[Y_2 \mid x_1, x_2, y_1, z_1, z_2]\, P(y_1 \mid x_1, z_1)\, P(z_1)\, P(z_2 \mid x_1, y_1, z_1$$

Again, z_1, z_2 can be high-dimensional

$$\boxed{w(x, y, z)} = \frac{P(x)}{\prod_k P(x_k \mid x^{(k-1)}, y^{(k-1)}, z^{(k)}}$$

$$\mathbb{E}_{P(Y|\hat{X})}\left[h(Y)\right] = \mathbb{E}_{R^W}\left[h(Y) \mid X\right]_{y_2|x}$$

expectation w.r.t. intervention

expectation w.r.t. pseudo-population.

$$R^W \propto P \cdot \underbrace{W}_{w_{mSBD}}$$

Generalization 2: front-door criterion

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left, dashed, leftrightarrow] & Z \arrow[r] & Y
\end{tikzcd}
$$

$$P(y \mid do(x)) = ?$$

$$= \sum_{z} P(y, z \mid do(x))$$

$$= \sum_{z} P(z \mid do(x))\, P(y \mid z, do(x))$$

$$\equiv$$

$$\sum_{z} P(z \mid do(x))\, P(y \mid do(z))$$

$$=$$

BD w.r.t. ϕ BD w.r.t. X

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left, dashed, leftrightarrow] & Z \arrow[r] & Y
\end{tikzcd}
\qquad = \sum_{z} P(z \mid x) \left(\sum_{x'} P(y \mid z, x')\, P(x') \right)
$$

$$\therefore\ \mathbb{E}[Y \mid do(x)] = \sum_{z} P(z \mid x) \left(\sum_{x'} \mathbb{E}[Y \mid z, x']\, P(x') \right)$$

z $\qquad$ x'

$f(z)$

we can use weighted regression for BD discussed before, to learn $f(z)$.

weights : $P(z) / P(z|x')$.

$$= \sum_z \underbrace{P(z|x)\, f(z)}.$$

again use weighted estimator from BD

$$wt = \frac{P(x)}{P(x|\phi)} = 1.$$

Theorem: (informal) If individual weighted estimators are consistent, then compositions are also consistent.

Generalization 3: front-door like graphs.

Suppose G satisfies the following:

1) $Y \perp\!\!\!\perp Z \mid X$ in $G_{\overline{X}\underline{Z}}$

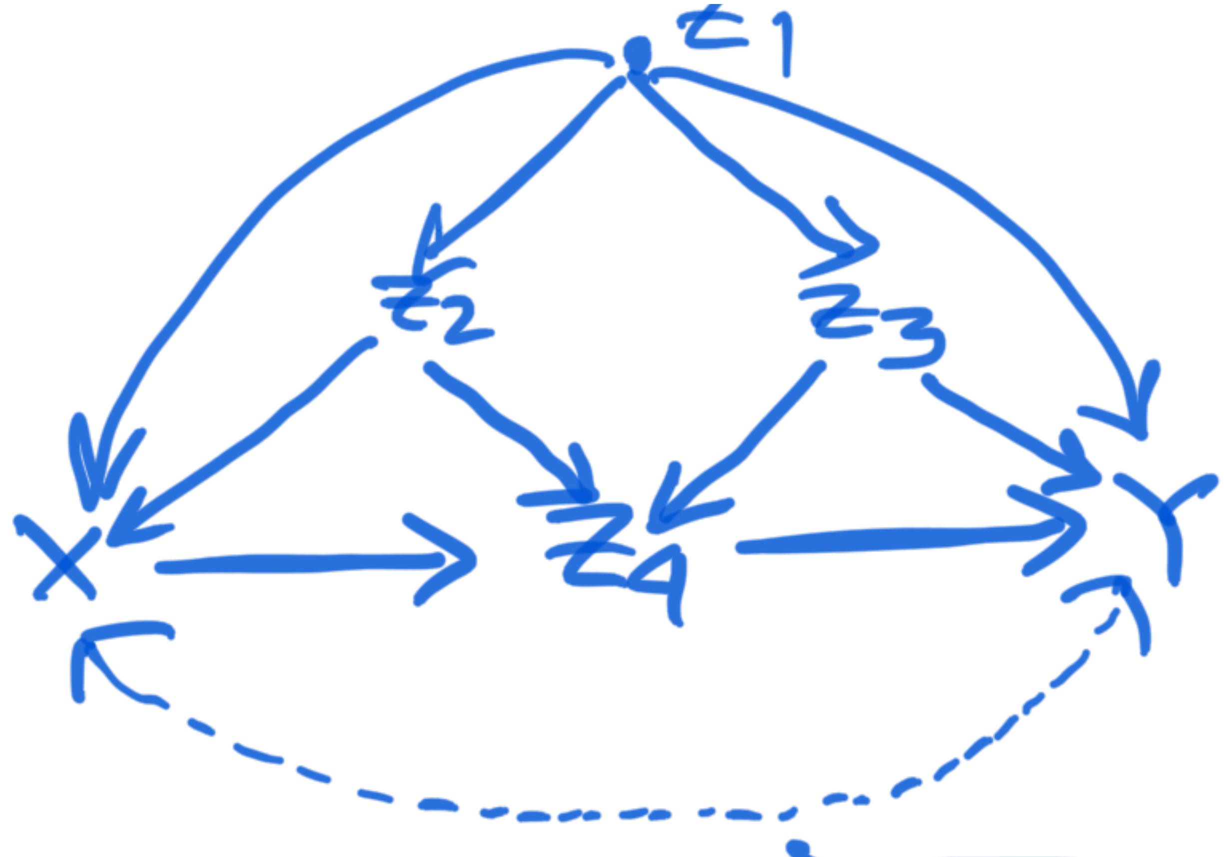
2) $Y \perp\!\!\!\perp X \mid Z$ in $G_{\overline{Z}\,\overline{X}}$

then also $\underline{P(y \mid z, do(x)) = P(y \mid do(z))}$

so, from previous analysis

$$P(y \mid do(x)) = \sum_z P(z \mid do(x))\, P(y \mid do(z))$$

then these two can be estimated by weighting if possible, eg: MSBD



$f(z_1 \ldots z_4)$

$$P(y|do(x)) = \sum_{z_1,z_2,z_3,z_4} P(z_1,z_2,z_3,z_4|do(x)) \cdot P(y|do(z_1,z_2,z_3,z_4))$$

$\mathbb{E}[f(Y)|do(x)]$ SBD w.r.t. ϕ

MSBD w.r.t. (ϕ,ϕ,ϕ,x)