TUBAF

Die Ressourcenuniversität. Seit 1765.

Covariate Adjustment, Do-Calculus, and Equivalence

Research seminar on causality

Björn Sprungk

Institute of Stochastics, TU Bergakademie Freiberg

June 05th, 2024

Content

Calculating Intervention Distributions by Covariate Adjustment

Objective

- Consider structural causal model $\mathfrak{C}$ with associated DAG $\mathcal{G}$

$$X_j = f_j(\mathbf{PA}_j, N_j), \qquad j = 1, \dots, d$$

- Recall: An intervention is a change of assignments of (some) X_k

$$X_k = f_k(\mathbf{P}A_j, N_j) \qquad \Rightarrow \qquad X_k = \tilde{f}_k(\tilde{\mathbf{PA}}_j, \tilde{N}_j)$$

- **Goal:** Compute intervention distributions

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y), \qquad X, Y \in \{X_1, \dots, X_d\}, \; X \neq Y.$$

Identifiability

Definition

An intervention distribution $p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y)$ is identifiable if it can be computed from the observational distribution, e.g., $p^{\mathfrak{C}}(x_1, \dots, x_d)$, and the graph structure $\mathcal{G}$.

- The observational distribution involves also conditional distributions $p_{X_j}^{\mathfrak{C}}(x_j \mid X_k = x_k)$
- We will consider two different approaches to compute identifiable intervention distributions:
  1. by covariate adjustment
  2. by the front-door formula (special case of do-calculus)

A useful invariance and the manipulation formula

- From the definition of structural causal models it follows for an SCM $\mathfrak{C}$ that

$$p^{\mathfrak{C}}(x_j \mid \mathbf{PA}_j) = p^{\tilde{\mathfrak{C}}}(x_j \mid \mathbf{PA}_j)$$

for any SCM $\tilde{C}$ that is contructed from $\mathfrak{C}$ by intervening on (some) X_k, $k \neq j$.

- Using the Markov property and the above invariance we thus obtain

$$p^{\mathfrak{C};\mathrm{do}(X_k:=\tilde{N}_k)}(x_1, \ldots, x_d) = \prod_{j=1}^{d} p^{\mathfrak{C};\mathrm{do}(X_k:=\tilde{N}_k)}\left(x_j \mid \mathbf{pa}_j\right) = \tilde{p}(x_k) \prod_{j \neq k} p^{\mathfrak{C}}\left(x_j \mid \mathbf{pa}_j\right)$$

where $\tilde{p}$ denotes density of $\tilde{N}_k$

- **Special and important case:**

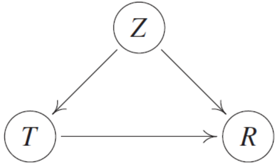
$$p^{\mathfrak{C};\mathrm{do}(X_k:=x)}(x_1,\ldots,x_d) = \begin{cases} \prod_{j\neq k} p^{\mathfrak{C}}\left(x_j \mid \mathbf{pa}_j\right), & \text{if } x_k = x, \\ 0, & \text{else.} \end{cases}$$

- For so-called “source nodes” ,i.e., nodes with parents we can now show that intervening is equal to conditioning: Let X_1 be a source node, then

$$\begin{aligned} p^{\mathfrak{C};\mathrm{do}(X_1:=x)}(x_1,\ldots,x_d) &= \mathbf{1}_{\{x\}}(x_1) \prod_{j\neq k} p^{\mathfrak{C}}\left(x_j \mid \mathbf{pa}_j\right) \\ &= \frac{p^{\mathfrak{C}}(x_1 \mid X_1 = x) \prod_{j\neq k} p^{\mathfrak{C}}\left(x_j \mid \mathbf{pa}_j\right)}{p^{\mathfrak{C}}(x_1 \mid X_1 = x)} \\ &= p^{\mathfrak{C}}(x_1,\ldots,x_d \mid X_1 = x) \end{aligned}$$

Example: Kidney stones

	Overall	Patients with small stones	Patients with large stones
Treatment a: Open surgery	78% (273/350)	**93%** (81/87)	**73%** (192/263)
Treatment b: Percutaneous nephrolithotomy	**83%** (289/350)	87% (234/270)	69% (55/80)



Source: J. Peters et al. *Elements of Causal Inference*. MIT Press, 2017.

- Let Z be size of stone, T kind of treatment, and R the recovery (all binary).
- Let us compute

$$P^{\mathfrak{C};\mathrm{do}(T=a)}(R=1), \qquad P^{\mathfrak{C};\mathrm{do}(T=b)}(R=1)$$

- We have with $\mathfrak{C}_a := \mathfrak{C}; \mathrm{do}\,(T = a)$

$$P^{\mathfrak{C}_a}(R = 1) = \sum_{t=a,b;z=0,1} P^{\mathfrak{C}_a}(R = 1, T = t, Z = z)$$
$$= \sum_{z=0,1} P^{\mathfrak{C}_a}(R = 1, T = a, Z = z)$$

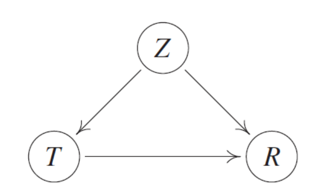
- And by the manipulation theorem

$$P^{\mathfrak{C}_a}(R = 1, T = a, Z = z) = P^{\mathfrak{C}}(R = 1 \mid T = a, Z = z)\, P^{\mathfrak{C}}(Z = z)$$

- In summary

$$P^{\mathfrak{C}_a}(R = 1) = \sum_{z=0,1} P^{\mathfrak{C}}(R = 1 \mid T = a, Z = z)\, P^{\mathfrak{C}}(Z = z)$$

	Overall	Patients with small stones	Patients with large stones
Treatment a: Open surgery	78% (273/350)	**93%** (81/87)	**73%** (192/263)
Treatment b: Percutaneous nephrolithotomy	**83%** (289/350)	87% (234/270)	69% (55/80)



- We can then estimate

$$P^{\mathfrak{C}_a}(R=1) \approx 0.93 \cdot \frac{357}{700} + 0.73 \cdot \frac{343}{700} = 0.832, \qquad P^{\mathfrak{C}_b}(R=1) \approx 0.782$$

- Average causal effect

$$P^{\mathfrak{C};\mathrm{do}(T=a)}(R=1) - P^{\mathfrak{C};\mathrm{do}(T=b)}(R=1) \approx 0.05$$

significantly different from

$$P^{\mathfrak{C}}(R=1 \mid T=a) - P^{\mathfrak{C}}(R=1 \mid T=b) = 0.78 - 0.83 = -0.05$$

Remark

- This simple three-node example illustrates nicely the difference between intervention and conditioning:

$$
\begin{aligned}
p_R^{\mathfrak{C};\mathrm{do}(X:=x)}(r) &= \sum_z p_R^{\mathfrak{C}}(r \mid X = x, Z = z)\, p_Z^{\mathfrak{C}}(z) \\
&\neq \sum_z p_R^{\mathfrak{C}}(r \mid X = x, Z = z)\, p_Z^{\mathfrak{C}}(z \mid X = x) \\
&= p_R^{\mathfrak{C}}(r \mid X = x)
\end{aligned}
$$

- We now generalize the observation from the example

TU BERGAKADEMIE FREIBERG

Adjustment formula

Definition

Let $\mathfrak{C}$ be an SCM over nodes $\mathbf{V}$ with a directed path from X to Y, $X, Y \in \mathbf{V}$. The causal effect from X to Y is called confounded if

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) \neq p_Y^{\mathfrak{C}}(y \mid X = x) \qquad \forall x, y.$$

Otherwise, it is called unconfounded.

Definition

Let $\mathfrak{C}$ be an SCM over nodes $\mathbf{V}$ and let $X, Y \in \mathbf{V}$ where $Y \notin \mathbf{PA}_X$. We call $\mathbf{Z} \subset \mathbf{V} \setminus \{X, Y\}$ a valid adjustment set for the ordered pair (X, Y) if

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) = \sum_{\mathbf{z}} p_Y^{\mathfrak{C}}(y \mid X = x, \mathbf{Z} = \mathbf{z})\, p_{\mathbf{Z}}^{\mathfrak{C}}(\mathbf{z}) \qquad \forall x, y.$$

When is an adjustment set valid?

- For any $\mathbf{Z} \subset \mathbf{V} \setminus \{X, Y\}$ we have

$$
\begin{aligned}
p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) &= \sum_{\mathbf{z}} p_{(Y,\mathbf{Z})}^{\mathfrak{C};\mathrm{do}(X=x)}(y, \mathbf{z}) \\
&= \sum_{\mathbf{z}} p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y \mid X = x, \mathbf{Z} = \mathbf{z})\, p_{(X,\mathbf{Z})}^{\mathfrak{C};\mathrm{do}(X=x)}(x, \mathbf{z}) \\
&= \sum_{\mathbf{z}} p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y \mid X = x, \mathbf{Z} = \mathbf{z})\, p_{\mathbf{Z}}^{\mathfrak{C};\mathrm{do}(X=x)}(\mathbf{z})
\end{aligned}
$$

- Thus, we require

$$
\begin{aligned}
p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y \mid X = x, \mathbf{Z} = \mathbf{z}) &= p_Y^{\mathfrak{C}}(y \mid X = x, \mathbf{Z} = \mathbf{z}), \\
p^{\mathfrak{C};\mathrm{do}(X=x)}{}_{\mathbf{Z}}(\mathbf{z}) &= p_{\mathbf{Z}}^{\mathfrak{C}}(\mathbf{z})
\end{aligned}
$$

A sufficient graphical condition

- Recall the augmentation of an SCM $\mathfrak{C}$ and DAG $\mathcal{G}$ by binary "intervention variables" I_k denoting that an intervention $X_k = x_k$ takes place, i.e.,

$$I_k = N_{I_k}, \qquad\qquad N_{I_k} \sim \mathrm{Bernoulli}(0.5),$$
$$X_k = \begin{cases} f_k(\mathbf{PA}_k, N_k), & \text{if } I_k = 0 \\ x_k, & \text{if } I_k = 1. \end{cases}$$

- In the augmented DAG $\mathcal{G}^*$ the I_k are parentless nodes pointing directly to X_k
- Then, recall Markov property

$$Y \perp\!\!\!\perp_{\mathcal{G}^*} I_k \mid \mathbf{Z} \implies Y \perp\!\!\!\perp I_k \mid \mathbf{Z}$$
$$\implies p_Y^{\mathfrak{C}^*}(y \mid Z = z) = p_Y^{\mathfrak{C}^*}(y \mid Z = z, I_k = 1) = p_Y^{\mathfrak{C}^*;\mathrm{do}(X_k = x_k)}(y \mid Z = z)$$

- Thus, let I denote the intervention variable for $\mathrm{do}\,(X = x)$. Then, if

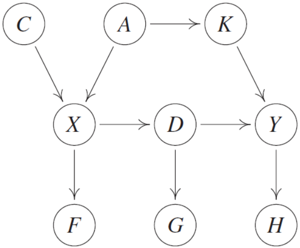
$$Y \perp\!\!\!\perp_{\mathcal{G}^*} I \mid X, \mathbf{Z} \qquad \text{and} \qquad \mathbf{Z} \perp\!\!\!\perp_{\mathcal{G}^*} I$$

we have the desired properties

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y \mid X = x, \mathbf{Z} = \mathbf{z}) = p_Y^{\mathfrak{C}}(y \mid X = x, \mathbf{Z} = \mathbf{z}), \qquad p^{\mathfrak{C};\mathrm{do}(X=x)_{\mathbf{Z}}(\mathbf{z})} = p_{\mathbf{Z}}^{\mathfrak{C}}(\mathbf{z}).$$

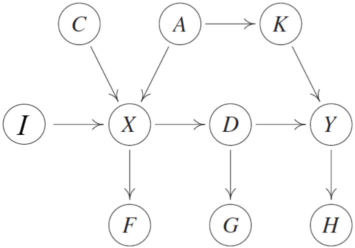
- **Example:**

$\mathcal{G}$



$\Rightarrow$

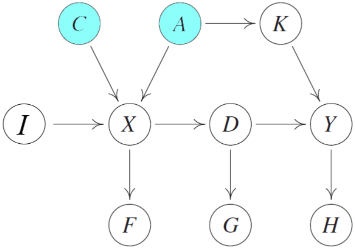
$\mathcal{G}^*$



Special valid adjustments

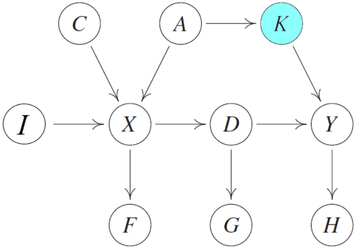
Parent adjustment

$$\mathbf{Z} = \mathbf{PA}_X$$



Back-door adjustment

- **Z** contains no descendant of X and
- **Z** blocks all "back-door" paths from X to Y, i.e., paths that start with an incoming arrow.



General adjustment criterion

Proposition (Shpitser et al., 2010)

Let $\mathfrak{C}$ be an SCM over nodes $\mathbf{V}$ and let $X, Y \in \mathbf{V}$ where $Y \notin \mathbf{PA}_X$. Any $\mathbf{Z} \subset \mathbf{V} \setminus \{X, Y\}$ with

- $\mathbf{Z}$ contains no descendant of any node $W \neq X$ on a directed path from X to Y
- and $\mathbf{Z}$ blocks all non-directed paths from X to Y

is a valid adjustment set for (X, Y).

- This graphical criterion is sufficient **and** necessary: if $\mathbf{Z}$ does not satisfy this criterion, then there exists SCM with same DAG $\mathcal{G}$ where $\mathbf{Z}$ is not valid for (X, Y).
- Parent is a special back-door adjustment and back-door is also special case of this general criterion.

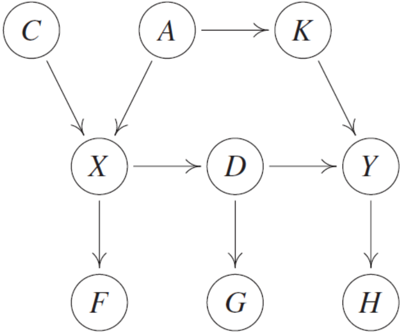


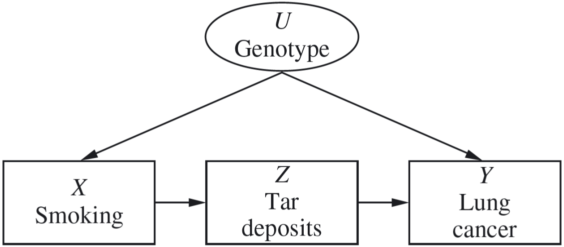
Figure 6.5: Only the path $X \leftarrow A \rightarrow K \rightarrow Y$ is a "backdoor path" from X to Y. The set $\mathbf{Z} = \{K\}$ satisfies the backdoor criterion (see Proposition 6.41 (ii)); but $\mathbf{Z} = \{F, C, K\}$ is also a valid adjustment set for (X, Y); see Proposition 6.41 (iii).

Source: J. Peters et al. *Elements of Causal Inference*. MIT Press, 2017.

Front-door adjustment and do-calculus

Motivation

- Often not all variables in a SCM are observable. This limits the application of covariate adjustment.
- **Example:** Assume U is not observable, then $P_Y^{\mathfrak{C};\mathrm{do}(X=x)}$ is not computable by (observable) covariate adjustment (back-door adjustment: $\mathbf{Z} = \{U\}$)



Source: Pearl et al. *Causal Inference in Statistics*. Wiley, 2016.

Front-door adjusment

- Based on the manipulation formula we have

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) = \sum_{u,\tilde{x},z} p^{\mathfrak{C};\mathrm{do}(X=x)}(u,\tilde{x},z,y) = \sum_{u,z} p_Y^{\mathfrak{C}}(y \mid U=u, Z=z)\, p_Z^{\mathfrak{C}}(z \mid X=x)\, p_U^{\mathfrak{C}}(u)$$

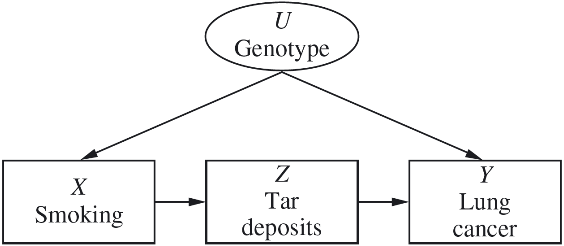
$$= \sum_{z} p_Z^{\mathfrak{C}}(z \mid X=x) \sum_{u} p_Y^{\mathfrak{C}}(y \mid U=u, Z=z)\, p_U^{\mathfrak{C}}(u)$$

- Now, we apply the back-door adjustment to rewrite

$$p_Y^{\mathfrak{C};\mathrm{do}(Z=z)}(z) = \sum_{u} p_Y^{\mathfrak{C}}(y \mid U=u, Z=z)\, p_U^{\mathfrak{C}}(u)$$

- But applying by applying the other back-door adjustment we also have

$$p_Y^{\mathfrak{C};\mathrm{do}(Z=z)}(z) = \sum_{\tilde{x}} p_Y^{\mathfrak{C}}(y \mid X=\tilde{x}, Z=z)\, p_X^{\mathfrak{C}}(\tilde{x})$$



Source: Pearl et al. *Causal Inference in Statistics*. Wiley, 2016.

- We, thus, obtain the front-door adjustment

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) = \sum_{z,\tilde{x}} p_Y^{\mathfrak{C}}(y \mid X = \tilde{x}, Z = z)\, p_X^{\mathfrak{C}}(\tilde{x})\, p_Z^{\mathfrak{C}}(z \mid X = x)$$

- This can be computed from observable variables

Definition

Let $\mathfrak{C}$ be an SCM over nodes $\mathbf{V}$ and let $X, Y \in \mathbf{V}$ where $Y \notin \mathbf{PA}_X$. A set $\mathbf{Z} \subset \mathbf{V} \setminus \{X, Y\}$ satisfies the front-door criterion relative to (X, Y) if

1. $\mathbf{Z}$ intercepts all directed paths from X to Y.
2. There is no back-door path from X to $\mathbf{Z}$.
3. All back-door paths from $\mathbf{Z}$ to Y are blocked by X.

Proposition

If $\mathbf{Z}$ satisfies the front-door criterion relative to (X, Y) and if $p^{\mathfrak{C}}_{(X,\mathbf{Z})}(x, \mathbf{z}) > 0$ for all x, z, then the causal effect of X on Y is identifiable and is given by

$$p_Y^{\mathfrak{C};\mathrm{do}(X=x)}(y) = \sum_{\mathbf{z}} p^{\mathfrak{C}}_{\mathbf{Z}}(\mathbf{z} \mid X = x) \sum_{\tilde{x}} p^{\mathfrak{C}}_Y(y \mid X = \tilde{x}, \mathbf{Z} = \mathbf{z})\, p^{\mathfrak{C}}_X(\tilde{x})$$

Example: Smoking

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar 400		No tar 400		All subjects 800	
	Smokers	Nonsmokers	Smokers	Nonsmokers	Smokers	Nonsmokers
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	1 (5%)	18 (90%)	38 (10%)	341 (85%)	39 (9.75%)
Cancer	57 (15%)	19 (95%)	2 (10%)	342 (90%)	59 (15%)	361 (90.25%)

Source: Pearl et al. *Causal Inference in Statistics*. Wiley, 2016.

- **Tobacco industry:** The table proves the beneficial effect of smoking!
- **Antismoking lobbyists:** Smoking would actually increase your risk of lung cancer, since smoking obviously is building up the chance of tar deposits.
- Who’s right?

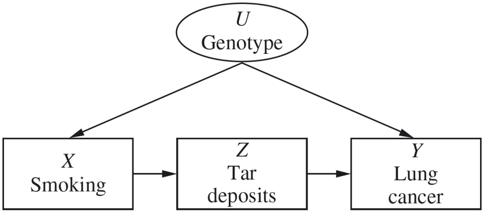


Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	18	1	38	324	56
	(85%)	(90%)	(5%)	(10%)	(81%)	(19%)
Cancer	57	2	19	342	76	344
	(15%)	(10%)	(95%)	(90%)	(19%)	(81%)

Source: Pearl et al. *Causal Inference in Statistics*. Wiley, 2016.

- Let us compute the (average) causal effect of smoking on getting lung cancer!
- By front-door adjustment we have

$$P^{\mathfrak{C};\mathrm{do}(X=x)}(Y=1) = \sum_{z=0}^{1} P^{\mathfrak{C}}(Z=z \mid X=x) \sum_{\tilde{x}=0}^{1} P^{\mathfrak{C}}(Y=1 \mid X=\tilde{x}, Z=z)\, P^{\mathfrak{C}}(X=\tilde{x})$$

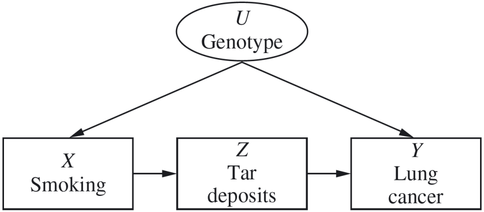


Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

Source: Pearl et al. *Causal Inference in Statistics*. Wiley, 2016.

$$P^{\mathfrak{C};\mathrm{do}(X=0)}(Y=1) = 0.95\,[0.9\cdot 0.5 + 0.1\cdot 0.5] + 0.05\,[0.95\cdot 0.5 + 0.15\cdot 0.5] = 0.5025$$

$$P^{\mathfrak{C};\mathrm{do}(X=1)}(Y=1) = 0.05\,[0.9\cdot 0.5 + 0.1\cdot 0.5] + 0.95\,[0.95\cdot 0.5 + 0.15\cdot 0.5] = 0.5475$$

$$0.045 = P^{\mathfrak{C};\mathrm{do}(X=1)}(Y=1) - P^{\mathfrak{C};\mathrm{do}(X=0)}(Y=1)$$

$$-0.755 = P^{\mathfrak{C}}(Y=1 \mid X=1) - P^{\mathfrak{C}}(Y=1 \mid X=0)$$

Do-Calculus

1. "Insertion/deletion of observations":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{z},\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X}, \mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ have been removed.

2. "Action/observation exchange":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x},\mathbf{Z}=\mathbf{z})}(\mathbf{y}\,|\,\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{z},\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X}, \mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ and outgoing edges from $\mathbf{Z}$ have been removed.

3. "Insertion/deletion of actions":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x},\mathbf{Z}=\mathbf{z})}(\mathbf{y}\,|\,\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X}, \mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ and $\mathbf{Z}(\mathbf{W})$ have been removed. Here, $\mathbf{Z}(\mathbf{W})$ is the subset of nodes in $\mathbf{Z}$ that are not ancestors of any node in $\mathbf{W}$ in a graph that is obtained from $\mathcal{G}$ after removing all edges into $\mathbf{X}$.

Source: J. Peters et al. *Elements of Causal Inference*. MIT Press, 2017.

Theorem 6.45 (Do-calculus) *The following statements hold.*

(i) The rules are complete; that is, all identifiable intervention distributions can be computed by an iterative application of these three rules [Huang and Valtorta, 2006, Shpitser and Pearl, 2006].

(ii) In fact, there is an algorithm, proposed by Tian [2002] that is guaranteed [Huang and Valtorta, 2006, Shpitser and Pearl, 2006] to find all identifiable intervention distributions.

(iii) There is a necessary and sufficient graphical criterion for identifiability of intervention distributions [Shpitser and Pearl, 2006, Corollary 3], based on so-called hedges [see also Huang and Valtorta, 2006].

Source: J. Peters et al. *Elements of Causal Inference*. MIT Press, 2017.

Equivalence and Falsifiability of Causal Models

Equivalence

- **Probabilistic models:** able to predict observational distribution of $\mathbf{X} = (X_1, \ldots, X_d)$
- **Interventional models:** able to predict any interventional distribution of $\mathbf{X} = (X_1, \ldots, X_d)$ (e.g., causal graphical models)
- **Counterfactual models:** able to predict any counterfactual distribution of $\mathbf{X} = (X_1, \ldots, X_d)$ (e.g., SCM)

Definition

We consider two SCM $\mathfrak{C}_1$ and $\mathfrak{C}_2$ as

probabilistically / interventionally / counterfactually equivalent

if the entail the same observational / observational + intervention / observational + intervention + counterfactual distributions.

Proposition

Assume that two SCM $\mathfrak{C}_1$ and $\mathfrak{C}_2$ for $\mathbf{X} = (X_1, \ldots, X_d)$ induce strictly positive, continuous conditional densities

$$p_{X_j}^{\mathfrak{C}_i}(x_j \mid \mathbf{pa}_j) > 0 \qquad \forall x_j, \mathbf{pa}_j \; \forall j = 1, \ldots, d \; \forall i = 1, 2$$

and satisfy causal minimality. If

$$P_{\mathbf{X}}^{\mathfrak{C}_1;\mathrm{do}(X_j = \tilde{N}_j)} = P_{\mathbf{X}}^{\mathfrak{C}_2;\mathrm{do}(X_j = \tilde{N}_j)} \qquad \forall j = 1, \ldots, d \; \forall \tilde{N}_j \text{ with full support}$$

then, $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are interventionally equivalent.

$\Rightarrow$ Equality of single-node intervention distributions yield interventionally equivalence

Proposition

Assume that two SCM $\mathfrak{C}$ and $\tilde{\mathfrak{C}}$ share the same noise distributions $P_{\mathbf{N}}$ and differ only in the kth structural assignment

$$f_k(\mathbf{pa}_k, n_k) = \tilde{f}_k(\tilde{\mathbf{pa}}_k, n_k) \qquad \forall \mathbf{pa}_k \forall n_k \text{ with } p(n_k) > 0$$

where $\tilde{\mathbf{PA}}_k \subset \mathbf{PA}_k$. Then both SCMs are counterfactually equivalent.

$\Rightarrow$ It suffices to consider (counterfactually) equivalent SCM which satisfies causal minimality.

Falsifiability of SCM

- We view SCMs as models for real-world data-generating processes.

- We can then falsify probab. / intervent. models in the following way:
  - if the induced observational distribution differs from given data (distribution)
  - if some induced interventional distribution differs from the results of a corresponding randomized experiment.

- Falsification of SCM via counterfactual distributions is, in general, hard in practice.