<데이터 사이언스를 위한 인과 추론>.

1. Introduction

1.1. Pearl's Causal Hierarchy

- ① Association $(y \mid \underline{x})$
- ② Intervention $(y \mid do(x), c)$
- ③ Counterfactual $P(y_x \mid x', y')$.

1.2. Review

- Chain Rule.

$$p(a,b,d,e) = p(a \mid b,d,e)\, p(b \mid d,e)\, P(d \mid e)\, P(e).$$

- Sum of Products

$$\sum_{a,c} f(a,b)\, g(b,c)\, h(c,d) = \sum_c h(c,d)\, g(b,c) \left(\sum_a f(a,b) \right)$$

$$\sum_{a,c} p(a \mid b)\, g(b,c)\, h(c,d) = \left(\sum_a p(a \mid b) \right) \left(\sum_c g(b,c)\, h(c,d) \right) = \sum_c g(b,c)\, h(c,d)$$

2. Causal Models and Graphs.

2.1. SCM

Def Structural causal model $\mathcal{M} = \langle V, U, F, P(U) \rangle$.

$$\begin{cases} V = \{V_1 \dots V_n\} & : \text{endogeneous}. \\ U & : \text{exogeneous} \\ F = \{f_1 \dots f_n\} & : V_i \leftarrow f_i(pa_i, u_i). \\ P(U) & : \text{dist. over } U. \end{cases}$$

2.2. Properties of SCM.

① SCM induces distribution $P(V)$.

- $F: U \to V$.
  - → U: r.v. 이므로 V도 r.v.
  - → $P(v)$: layer 1 of PCH.

② SCM induces causal diagram

- represented as DAG. * SCM $\rightleftarrows$ G. (one-way).
  - → $V_i \in V$: node (vertex)
  - → directed edge $B \to V_i$ for all $B \in Pa_i$.
  - → bidirected edge $V_i \leftrightarrow V_j$ whenever $U_i \not\perp U_j$.

- Terminology:

For $V \in \underline{V}$,
- Parents
- Children
- Ancestors
- Descendants
- Siblings

} 자기 자신도 포함.

- For $V \in \mathbf{V}$,
  - **Parents** of V: $\{W \mid W \to V \in \mathcal{G}\}$
  - **Children** of V: $\{W \mid W \leftarrow V \in \mathcal{G}\}$.
  - **Ancestors** of V: $\{W \mid W \to \ldots \to V \in \mathcal{G}\}$.
  - **Descendants** of V: $\{W \mid W \leftarrow \ldots \leftarrow V \in \mathcal{G}\}$.
  - **Siblings** of V: $\{W \mid W \leftrightarrow V \in \mathcal{G}\}$.

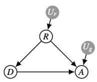
③ SCM implies PCH.

L [1]: $P(v) = \sum_u P(V=v, U=u) = \sum_u P(\forall_{V_i \in V}\ f_i(pa_i, u_i) = v_i,\ U=u)$.

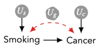
→ topological order: $a \prec b$. (모든 var 이 parents 다음에 오도록.).

→ $P(v) = \sum_u P(u) \prod_{V_i \in V} P(v_i \mid v_1 \cdots v_{i-1}, \underline{u})$. (∵ chain rule).

$= \sum_u P(u) \prod_{V_i \in V} P(v_i \mid pa_i, u_i)$.



$P(r,d,a) = \sum_{u_R, u_G} P(u_R, u_G) P(r \mid u_R) P(d \mid r) P(a \mid r, d, u_G)$



$P(s,c) = \sum_{u_S, u_G, u_C} P(u_S, u_G, u_C) P(s \mid u_G, u_S) P(c \mid s, u_G, u_C)$

2.3. Conditional Independence : G 가 P(v)에 대해 의미하는 바?

- $X \perp\!\!\!\perp Y \iff P(Y=y \mid X=x) = P(Y=y)$.

  $X \perp\!\!\!\perp Y \mid Z \iff P(X=x \mid Z=z, Y=y) = P(X=x \mid Z=z)$.

- $P(v_i \mid \underbrace{v_1 \cdots v_{i-1}}, u) = P(v_i \mid pa_i, u_i)$.

  ↳ maximal set = non-descendants of V_i. = nd_i.

⇒ $V_i \perp\!\!\!\perp Nd_i \setminus Pa_i,\ U \setminus U_i \mid Pa_i, U_i$.

2.4. Markovian Condition

- $U_1 \ldots U_n \in U$ are jointly independent. ($\iff$ no bidirected edge in G).

  ⇒ model is called Markovian

- Markovian factorization : $P(v) = \sum P(u) \prod P(v_i \mid pa_i, u_i) = \prod P(v_i \mid pa_i)$.

  ↓. parents 에만 영향.

$$
\begin{aligned}
P(\mathbf{v}) &= \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i) \\
&= \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i) P(\mathbf{u}_i) && (\because \mathbf{U}_i \text{ are independent}) \\
&= \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i) P(\mathbf{u}_i \mid \mathbf{pa}_i) && (\because \mathbf{U}_i \perp\!\!\!\perp \mathbf{Pa}_i) \\
&= \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P(v_i, \mathbf{u}_i \mid \mathbf{pa}_i) \\
&= \prod_{V_i \in \mathbf{V}} \sum_{\mathbf{u}_i} P(v_i, \mathbf{u}_i \mid \mathbf{pa}_i) = \boxed{\prod_{V_i \in \mathbf{V}} P(v_i \mid \mathbf{pa}_i)}.
\end{aligned}
$$

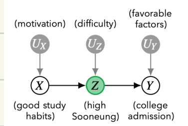
Bayesian Factorization

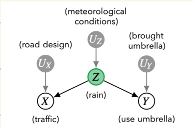
local Markov : $V_i \perp\!\!\!\perp Nd_i \setminus Pa_i \mid Pa_i$.

2.5. Graph Separation

- $X \not\perp\!\!\!\perp Y$: $P(X=1, Y=1) \neq P(X=1)\,P(Y=1)$.

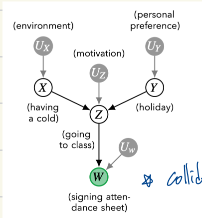
  $X \perp\!\!\!\perp Y \mid Z$: $P(x, y \mid z) = P(x \mid z)\,P(y \mid z)$.





- $X \not\perp\!\!\!\perp Y$

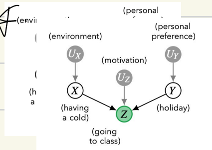
  $X \perp\!\!\!\perp Y \mid Z$



- $X \perp\!\!\!\perp Y$

  $X \not\perp\!\!\!\perp Y \mid W$*.

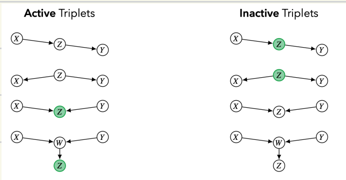
* collider 의 descendants 에 주의.



- $X \perp\!\!\!\perp Y$.

  $X \not\perp\!\!\!\perp Y \mid Z$.

  Z를 알면 X-Y 관계 발생.



- Active : $X \not\perp\!\!\!\perp Y$.

  Inactive : $X \perp\!\!\!\perp Y$. → d-separated.

- Active 한 경로* 하나라도 있으면 안됨.

3. Identification of Causal Effects

3.1. SCM induces L_2-Distribution

- Intervention : $do(\cdot)$ operator.

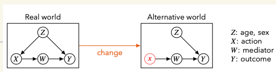
  → given $\mathcal{M}$, fixing $X \in V$ to constant x

  → transforms $\mathcal{M}$ to $\mathcal{M}_x$. ($G \to G_{\bar{X}}$).

  $=1$ in $\mathcal{M}_x$

  $P(v) = P(z)\,\boxed{P(x \mid z)}\,P(w \mid z)\,P(y \mid w, z)$

  $P_x(v \setminus \{x\}) = P(z)\,P(w \mid z)\,P(y \mid w, z)$

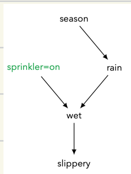


- Ex) $P(v) = P(Sn)\,P(Sp \mid Sn)\,P(Rn \mid Sn)\,P(Wt \mid Sp, Rn)\,P(Sl \mid Wt)$

  → $P(Wt \mid Sp = on)$ $= \dfrac{\sum_{sn,rn} P(sn)P(Sp = on \mid sn)P(rn \mid sn)P(wt \mid Sp = on, rn)}{\sum_{sn} P(Sp = on \mid sn)P(sn)}$

  $P(Wt \mid do(Sp = on))$ $= \dfrac{\sum_{sn,rn} P(sn)P(Sp = on)P(rn \mid sn)P(wt \mid Sp = on, rn)}{\sum_{sn} P(Sp = on)P(sn)}$

  $= \sum_{sn,rn} P(sn)P(rn \mid sn)P(wt \mid Sp = on, rn)$



3.2. Truncated Factorization Product

- $P(v \setminus x \mid do(x)) = \prod_{V_i \in V \setminus X} P(v_i \mid pa_i)$ (* in a Markovian model).

  $= \dfrac{P(v)}{\prod_{X \in X} P(x \mid pa_X)}$ $\left(\because P(v) = \prod_{V_i \in V \setminus X} P(v_i \mid pa_i) \prod_{X \in X} P(x \mid pa_X)\right)$

→ when $X = \{X\}$, $= \dfrac{P(v)}{P(x \mid pa_x)} = \dfrac{P(v)\,P(pa_x)}{P(x, pa_x)}$

$= P(v'' \mid x, pa_x)\ P(pa_x)$.

↳ "re-weighting process"

3.3. Identification Problem

<u>Def</u> "identifiable" $\Leftarrow$ $P(y \mid do(x))$ can be computed uniquely

$\Leftarrow$ $P_{M_1}(y \mid do(x)) = P_{M_2}(y \mid do(x))$.

→ M_1, M_2 가 같은 G, $P(v)$ induce 하면 $P(y \mid do(x))$ 가 유일해야 identifiable.

<u>Thm</u> (Identification in Markovian Models). → 항상 TFP로 identifiable.

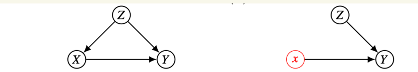
Since $P(v' \mid do(x)) = \prod_{v_i \in V \setminus X} P(v_i \mid pa_i)$, $P(y \mid do(x)) = \sum_{v' \setminus y} \prod_{v_i \in V \setminus X} P(v_i \mid pa_i)$.

↳ If X is singleton, $P(y \mid do(x)) = \sum_{pa_x} P(y \mid x, pa_x) P(pa_x)$

cf. $\sum_{v'' \setminus Y} P(v'' \mid x, pa_x)\ P(pa_x)$.

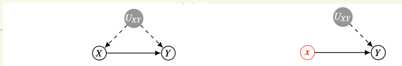
4. Confounding and Backdoor.

4.1. (Recap) Identifiable vs Non-identifiable



$P(\mathbf{v}) = P(z)P(x \mid z)P(y \mid x,z)$ $\quad$ $P(\mathbf{v}' \mid do(x)) = P(z)P(y \mid x,z)$

vs.



$P(\mathbf{v}) = \sum_{u_{xy}} P(y \mid x, u_{xy}) P(x \mid u_{xy}) P(u_{xy})$ $\quad$ $P(\mathbf{v} \mid do(x)) = \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy})$

$P^{(1)}(y \mid do(x)) \neq P^{(2)}(y \mid do(x))$.

4.2. Backdoor Criterion.

— Confounding bias → removable?

<u>Def</u> Z : back door criterion w.r.t. X, Y.

if (i) no node in Z is descendant of X.

(ii) Z blocks every path between $X^{\in X}$ and $Y^{\in Y}$ that contains arrow into X.

<u>Thm</u> If Z satisfies back-door criterion, effect on Y is identifiable

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

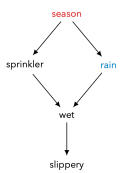
→ pa_x는 backdoor의 예.
bidirected 존재 → parents로 충분 X.

Ex)

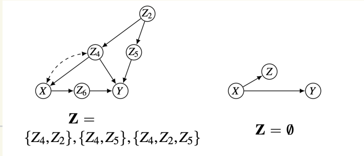
Rain satisfies the backdoor criterion relative to Sprinkler and Wet:
(i) Rain is not a descendant of Sprinkler, and
(ii) Rain blocks the only backdoor path from Sprinkler to Wet.

Adjusting for the direct parents of Sprinkler, we have:

$$\begin{aligned} P(\text{wt} \mid do(\text{sp})) &= \sum_{\text{sn}} P(\text{wt} \mid \text{sp}, \text{sn}) P(\text{sn}) \\ &= \sum_{\text{sn},\text{rn}} P(\text{wt} \mid \text{sp}, \text{sn}, \text{rn}) P(\text{rn} \mid \text{sp}, \text{sn}) P(\text{sn}) \\ &= \sum_{\text{sn},\text{rn}} P(\text{wt} \mid \text{sp}, \text{rn}) P(\text{rn} \mid \text{sn}) P(\text{sn}) \\ &= \sum_{\text{rn}} P(\text{wt} \mid \text{sp}, \text{rn}) \sum_{\text{sn}} P(\text{rn}, \text{sn}) \\ &= \sum_{\text{rn}} P(\text{wt} \mid \text{sp}, \text{rn}) P(\text{rn}) \end{aligned}$$



— Graphical condition : d-sep X from Y in $G_{\underline{X}}$



4.3. Evaluation

— Inverse Probability weighting (IPW)

$$P(y|do(x)) = \sum_z P(y|x,z)P(z) = \sum_z \frac{P(y,x,z)}{P(x|z)}$$ ← P(v)에서
← fit with $g(z)$.

$$= \sum_z \frac{\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(y_i=y,\ x_i=x,\ z_i=z)}{g(z)} = \frac{1}{N}\sum_i \frac{\mathbb{1}(y_i=y, x_i=x)}{g(z_i)}$$

$g(z)$. ↳ $P(X=x|Z_i)$

$$\mathbb{E}[y|do(x)] = \frac{1}{N}\sum y_i \frac{\mathbb{1}(x_i=x)}{g(z_i)}$$

$\Rightarrow \hat{ATE} = \hat{E}[Y|do(X=1)] - \hat{E}[Y|do(X=0)]$

└ "pseudo causal samples" 에서 $\mathbb{E}[y|do(x)]$ estimate.

5. Adjustment Criterion

5.1. Conditional Backdoor. Tightness of Backdoor.

Thm $P(y|do(x), w)$ is identified if $W \cup Z$ satisfies Backdoor criterion.

$\rightarrow P(y|do(x),w) = \sum_z P(y|x,z,w)P(z|w)$.

— Backdoor criterion은 충분조건 (BD ⇒ identifiable).

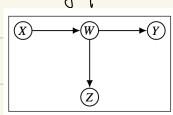
↳ X의 descendant 제외 ⇒ tight.

5.2. Adjustment Criterion

Def (Proper Causal Paths). path from X to Y is causal if it is directed

proper if it only contains X at the starting point.

→ Disturbed by adjusting variable on path (+ its descendant).



↳ $P(y|do(x)) = P(y|x)$

$\neq \sum_z P(y|x,z)P(z)$. (혹은 W를 adjust)

Def Z is admissible for adjustment if

(i) Z does not disturb any PCP (⇔ no element in Z is descendant in $G_{\overline{X}}$ of any $w \notin X$ on a PCP).

(ii) all proper non-causal paths in G are blocked by Z

$\Rightarrow Z$ satisfies A.C. $\Leftrightarrow P(y|do(x)) = \sum_z P(y|x,z)P(z)$.

— 찾는 순서 :

① Find variables on PCPs. : $W = De(X)_{G_{\bar{X}}} \cap An(Y)_{G_{\bar{X}}}$.

bidirected edge 제거한 뒤 찾기.

② Determine forbidden variables : $F = De(W \setminus X)_{G_{\bar{X}}}$.

↳ if $Z \cap F \neq \phi$, Z is not admissible.

③ Check Z blocks all proper non-causal paths : $G' = G \setminus \{X \to W \in H \mid w \in W, x \in X\}$.

↳ non-causal path 만 남음.

↳ if $(X \perp\!\!\!\perp Y \mid Z)_{G'}$ → Z is admissible.

5.3. Explicit Construction

$An^{-}(W) = An(W) \setminus W$.

Lemma If there exists admissible set Z, $Z_0 = An^{-}(X \cup Y) \setminus F$ is also admissible.

⇒ If Z_0 is not admissible, there exists no possible separator.

↳ Motivation of Algorithm.

5.4. Conditional Ignorability in PO Framework

Def $Y_x \perp\!\!\!\perp X \mid Z$

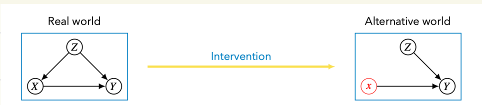
↳ $P(y \mid do(x), z) = P(y \mid x, z)$ → A.C.는 Z의 존재성에 대한 정당화를 제공.

6. The Causal Calculus.

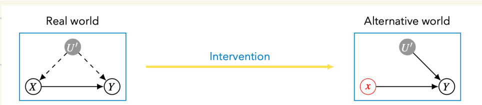
6.1. Three Examples : TFP in Semi-Markovian.

— $P(v' \mid do(x)) = \sum_{u} \prod_{V_i \in V \setminus X} P(v_i \mid pa_i, u_i) P(u)$

⇒ $P(y \mid do(x)) = \sum_{v \setminus x \setminus y} \sum_{u} \sum_{V_i \in V \setminus X} P(v_i \mid pa_i, u_i) P(u)$

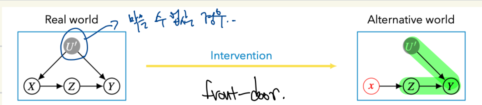


→ $P(y \mid do(x)) = \sum_{z} P(y \mid x, z) P(z)$



→ $P(y \mid do(x)) = \sum_{u'} P(y \mid x, u') P(u')$.

< non-identifiable! >



→ $P(z, y \mid do(x)) = P(z \mid x) \sum_{u'} P(y \mid z, u') P(u')$.

$P(y \mid do(z))$.

↳ ⓧ를 block 하면 backdoor 불요.

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

< identifiable! >.

⇒ TFP로 나타낸 뒤, U-free expression 유도.

6.2. Causal Calculus

- Prelude : $(A \perp\!\!\!\perp B \mid C)$ under $do(x)$ $\Rightarrow$ $(A \perp\!\!\!\perp B \mid C, \underset{\text{constant.}}{X})$ in $G_{\overline{X}}$.

  $\Leftrightarrow$ $(A \perp\!\!\!\perp B \mid C)$ in $G \setminus X$

- $(Z \perp\!\!\!\perp Y \mid X)_{G_{\overline{X}}}$ $\Leftrightarrow$ $P(y \mid do(x), z) = P(y \mid do(x))$.

  $(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}}$ $\Leftrightarrow$ $P(y \mid do(x), z) = P(y \mid x, z)$.

  $(Z \perp\!\!\!\perp X)_{G_{\overline{X}}}$ $\Leftrightarrow$ $P(z \mid do(x)) = P(z)$

Rule 1: *Adding/removing Observations*
$P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} \mid do(\mathbf{x}))$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.

Rule 2: *Action/observation Exchange*
$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z})) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$.

Rule 3: *Adding/removing Actions*
$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z})) = P(\mathbf{y} \mid do(\mathbf{x}))$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}\mathbf{Z}}}}$

Rule 1: *Adding/removing Observations*
$P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.

Rule 2: *Action/observation Exchange*
$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$.

Rule 3: *Adding/removing Actions*
$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w})$ if $(\mathbf{Z} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}\overline{\mathbf{Z}(\mathbf{W})}}}$
where $\mathbf{Z}(\mathbf{W}) = \mathbf{Z} \setminus An(\mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.

- Rule 2) model – submodel generates same prob. $\to$ F_Z 로 선택, $(F_Z \perp\!\!\!\perp Y \mid Z, W)_{G'}$

  Z가 given $\Rightarrow$ d-connection은 $\to Z \leftarrow$. $G_{\underline{Z}}$ 만 check.

- Rule 3) 'change to do' $\to$ $f_Z = 1$ 이면 $do(Z)$. $F_Z \perp\!\!\!\perp Y \mid W, X$ in G'.

  $F_Z \to Z \to \cdots$ 항상 ok.

  $F_Z \to Z \leftarrow \cdots$ $Z \in An(W)$ in $G_{\overline{X}}$. $\Rightarrow$ $(Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$.

6.3. Do-Calculus and Identification

Thm $Q = P(y \mid do(x))$ is identifiable $\Leftrightarrow$ $\exists$ seq. of rules of do-calculus $\times$ prob. axioms that reduces Q into a do-free expression.

- 예) $P(c \mid do(s)) = \sum_t P(t \mid s) \sum_{s'} P(c \mid s', t) P(s')$. $S \to T \to C$

Lemma Q is not identifiable in G $\Rightarrow$ G^+ 에서도 not identifiable.

identifiable in G $\Rightarrow$ G^- 에서도 identifiable.

Thm $\exists$ bidirected path X to its children, $P(v' \mid do(x))$ is not identifiable.

6.4. Simpson's Paradox

$P(Y \mid F, X) < P(Y \mid F, \neg X)$
$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$
$P(Y \mid X) > P(Y \mid \neg X)$

$\Rightarrow$ If
$P(Y \mid F, do(X)) < P(Y \mid F, do(\neg X))$
$P(Y \mid \neg F, do(X)) < P(Y \mid \neg F, do(\neg X))$,

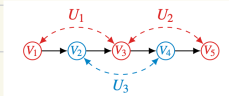
$P(Y \mid do(X)) < P(Y \mid do(\neg X))$. always hold.

pf)
$$P(y \mid do(x)) = \sum_f P(y \mid do(x), f) P(f \mid do(x))$$
$$= \sum_f P(y \mid do(x), f) P(f)$$

7. Algorithmic ID.

7.1. C-Factorization

<u>Def</u> (C-Component) V_i, V_j share common U.



$\rightarrow$ (V_1, V_3, V_5), (V_2, V_4). $\quad$ ↳ bidirected edge 만 고려. $(G[H])$.

– Decomposition (Markovian). $\rightarrow$ $\cancel{\sum_u P(u)} \prod p(v_i \mid pa_i, \cancel{u_i})$. $\Rightarrow$ computable

(Semi-Markovian) $\rightarrow$ $P(\mathbf{v}) = \left(\sum_{u_3} P(v_2 \mid v_1, u_3) P(v_4 \mid v_3, u_3) P(u_3)\right)\left(\sum_{u_1,u_2} P(u_1,u_2) P(v_1 \mid u_1) P(v_3 \mid v_2, u_1, u_2) P(v_5 \mid v_4, u_2)\right).$

$\Rightarrow$ $Q[C]\ (c, pa_c) = \sum_{u(\mathbf{C})} P(u(\mathbf{C})) \prod_{V_i \in \mathbf{C}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i)$ where $U(\mathbf{C}) = \bigcup_{V_i \in \mathbf{C}} \mathbf{U}_i$. $\rightarrow Q[C]$.

$\Rightarrow$ $P(\mathbf{v}) = Q[V_2, V_4](v_2, v_4, v_1, v_3) \cdot Q[V_1, V_3, V_5](v_1, v_3, v_5, v_2, v_4).$

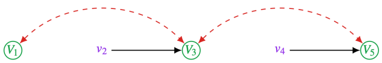
$= Q[V_1, V_3, V_5]\ Q[V_2, V_4].$

$*$ $\begin{cases} Q[V] = P(V) \\ Q[\emptyset] = 1. \end{cases}$

– C-factors are causal effects :

$$P(c \mid do(v \setminus c)) = \sum_u P(u) \prod_{V_i \in C} P(v_i \mid pa_i, u_i)$$
$$= \sum_{u(C)} \prod_{V_i \in C} P(v_i \mid pa_i, u_i)\, P(u(C)) = Q[C].$$

$$Q[V_1, V_3, V_5](V_1, V_3, V_5, v_2, v_4) = \sum_{u_1,u_2} P(u_1,u_2) P(V_1 \mid u_1) P(V_3 \mid v_2, u_1, u_2) P(V_5 \mid v_4, u_2)$$



– Marginalizing variables $\begin{cases} \text{not the parent of other components} \rightarrow \text{가능.} \\ \text{o.w.} \rightarrow \text{불가능.} \end{cases}$

$$\sum_{v_3} Q[V_1, V_3, V_5] = \sum_{v_3} \sum_{u_1,u_2} P(u_1,u_2) P(v_1 \mid u_1) P(v_3 \mid v_2, u_1, u_2) P(v_5 \mid v_4, u_2)$$
$$= \sum_{u_1,u_2} P(u_1,u_2) P(v_1 \mid u_1) P(v_5 \mid v_4, u_2)$$
$$= Q[V_1, V_5].$$

<u>Lemma</u> If W is ancestral, $Q[W] = \sum_{C \setminus W} Q[C]$. $\quad$ $C(\cdot)$: (maximal) C-components

<u>Lemma</u> Let $H_1 \ldots H_k$ be C-components of $G[H]$. $Q[H] = \prod_{H_j \in C(G[H])} Q[H_j]$.

$\Rightarrow$ <u>Lemma</u> Let $V^{(1)} \prec V^{(2)} \prec \cdots \prec V^{(|H|)}$

$$Q[H_j] = \prod_{V^{(i)} \in H_j} \frac{Q[H^{\preceq i}]}{Q[H^{\preceq i-1}]}, \quad Q[H^{\preceq i}] = \sum_{h^{>i}} Q[H].$$

$$Q[V_1, V_3, V_5] = \frac{Q[V_1]}{Q[\ \ ]} \cdot \frac{Q[V_3, V_2, V_1]}{Q[\ \ V_2, V_1]} \cdot \frac{Q[V_5, V_4, V_3, V_2, V_1]}{Q[\ \ V_4, V_3, V_2, V_1]},$$

$$Q[V_2, V_4] = \frac{Q[V_2, V_1]}{Q[\ \ V_1]} \cdot \frac{Q[V_4, V_3, V_2, V_1]}{Q[\ \ V_3, V_2, V_1]}.$$

$$Q[V_1, V_3, V_5] = \frac{Q[V_1]}{Q[\ \ ]} \cdot \frac{Q[V_3, V_2, V_1]}{Q[\ \ V_2, V_1]} \cdot \frac{Q[V_5, V_4, V_3, V_2, V_1]}{Q[\ \ V_4, V_3, V_2, V_1]}$$
$$= \frac{\sum_{v_2,v_3,v_4,v_5} Q[\mathbf{V}]}{\sum_{\mathbf{v}} Q[\mathbf{V}]} \cdot \frac{\sum_{v_4,v_5} Q[\mathbf{V}]}{\sum_{v_3,v_4,v_5} Q[\mathbf{V}]} \cdot \frac{\sum_{\emptyset} Q[\mathbf{V}]}{\sum_{v_5} Q[\mathbf{V}]}$$
$$= \frac{\sum_{v_2,v_3,v_4,v_5} P(\mathbf{v})}{\sum_{\mathbf{v}} P(\mathbf{v})} \cdot \frac{\sum_{v_4,v_5} P(\mathbf{v})}{\sum_{v_3,v_4,v_5} P(\mathbf{v})} \cdot \frac{\sum_{\emptyset} P(\mathbf{v})}{\sum_{v_5} P(\mathbf{v})}$$
$$= \frac{P(v_1)}{1} \cdot \frac{P(v_1, v_2, v_3)}{P(v_1, v_2)} \cdot \frac{P(\mathbf{v})}{P(v_1, v_2, v_3, v_4)} = P(v_1) P(v_3 \mid v_1, v_2) P(v_5 \mid v_1, v_2, v_3, v_4).$$

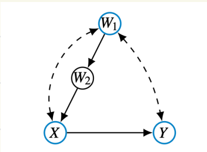
7.2. Examples

— ① P(v) 를 c-factorization

② intervention $P(v \setminus x \mid do(x))$ c-factorization $\xrightarrow{sum}$ $P(y \mid do(x))$

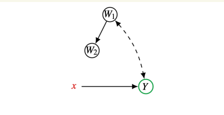
③ ①의 Q[C] 를 이용하여 ②의 Q[C] 계산. (marg / fact)

④ do(x) 그래프 → x를 x'로 대체.



→ adjustment 불가 : $\begin{cases} X \leftarrow W_2 \leftarrow W_1 \leftrightarrow Y \\ X \leftrightarrow W_1 \leftrightarrow Y. \end{cases}$

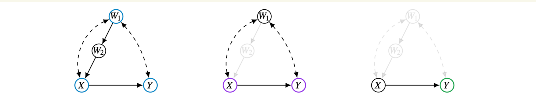
$\begin{cases} W_1 \text{ 막으면 collider.} \to \text{open.} \\ W_2 \text{ 막아도 path B open.} \end{cases}$, $P(v) = Q[W_1, X, Y]\, Q[W_2]$ → Q[Y] ?



$P(y \mid do(x)) = \sum_{\mathbf{w}} Q[W_1, W_2, Y]$

→ $Q[W_1, W_2, Y]$ 에서 Y가 ancestral

⇒ 쪼개지 않고 Q[Y]로



$$Q[W_1,X,Y] = \frac{Q[W_1]}{Q[\emptyset]}\frac{Q[W_1,W_2,X]}{Q[W_1,W_2]}\frac{Q[W_1,W_2,X,Y]}{Q[W_1,W_2,X]} = P(w_1)P(x \mid w_1,w_2)P(y \mid w_1,w_2,x).$$

$$Q[X,Y] = \sum_{w_1} Q[W_1,X,Y] = \sum_{w_1} P(w_1)P(x \mid w_1,w_2)P(y \mid w_1,w_2,x).$$

$$Q[X,Y] = Q[X]Q[Y].$$

$$Q[Y] = \frac{Q[X,Y]}{Q[X]} = \frac{Q[X,Y]}{\sum_y Q[X,Y]} = \frac{\sum_{w_1} P(w_1)P(x \mid w_1,w_2)P(y \mid w_1,w_2,x)}{\sum_{w_1} P(w_1)P(x \mid w_1,w_2)} = P(y \mid do(x)$$

① $Q[W_1, X, Y]$ 에서 W_1 marg. (ancestral 확인).

② $Q[X, Y]$ X → Y. $= Q[X]\,Q[Y]$.

③ Q[Y] : factorization.

7.3. General Approach

① $P(y \mid do(x))$ 나타내기: $An(Y)_{G_{\underline{X}}}$ 고려.

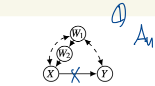
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}} \setminus \mathbf{Y}} P(An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}} \mid do(\mathbf{V} \setminus An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}})) = \sum_{c \setminus y} P(c \mid do(v \setminus c)) = \sum_{c \setminus y} Q[C].$$

→ $Q[C] = \prod_i Q[C_i]$, $Q[V] = \prod Q[T_i]$: $Q[C_i]$ 를 $Q[T_i]$로 구하기

② Identify Algorithm

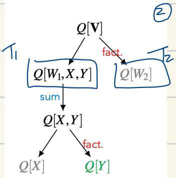
A function determining if $Q[\mathbf{C}]$ is computable from $Q[\mathbf{T}]$.

Identify($\mathbf{C}$, $\mathbf{T}$, $Q[\mathbf{T}]$, $\mathcal{G}$):
  Let $\mathbf{A} = An(\mathbf{C})_{\mathcal{G}[\mathbf{T}]}$
  If $\mathbf{A} = \mathbf{C}$, ... C is ancestral in T
    return $Q[\mathbf{C}] = \sum_{\mathbf{t} \setminus \mathbf{c}} Q[\mathbf{T}]$
  If $\mathbf{A} = \mathbf{T}$, ... nothing to marginalize, unable to factorize
    return Fail
  Let $\mathbf{A}_i$ be the c-comp of $\mathcal{G}[\mathbf{A}]$ that contains $\mathbf{C}$
  Get $Q[\mathbf{A}_i]$ from $Q[\mathbf{T}]$. .... marginalize, then factorize
  Return Identify($\mathbf{C}, \mathbf{A}_i, Q[\mathbf{A}_i], \mathcal{G}$).



$P(y \mid do(x)) = Q[Y]$

① $An(Y)_{G_{\underline{X}}} = \{Y\}$. → $C = \{Y\}$.



② Q[T] 에서 Q[C] 구하기.

— $A = An(C)_{G[T_1]} = \{X, Y\}$

i) $A \neq C$.

ii) $A = C$ 도 아님.

iii) Q[A]로 marginalize.

→ fact: Q[X]. Q[Y].

8. Linear SCM.

8.1. Introduction.

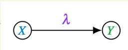
— linear SCMs : defined as system of linear equations.

: $V_i \leftarrow \sum_{V_j \in pa_i} \lambda_{ij} V_j + \varepsilon_i$.

— latent confounding : bidirected edge 로 표현. $\varepsilon_i, \varepsilon_j$ 간의 covariance (ε_{ij}).

— Ex) $X \xrightarrow{\lambda} Y$: $\mathbb{E}[Y \mid do(X=x)] = \mathbb{E}[\lambda x + \varepsilon_y] = \lambda x$.

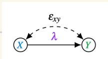
— identification : structural coef. 를 $\Sigma(\sigma_{ij})$ 로 표현.



$\Rightarrow$

$$\begin{aligned}\sigma_{xy} &= \mathbb{E}[XY]\\ &= \mathbb{E}[X(\lambda X + \varepsilon_y)]\\ &= \mathbb{E}[\lambda XX + X\varepsilon_y]\\ &= \lambda \underbrace{\mathbb{E}[XX]}_{1} + \underbrace{\cancel{\mathbb{E}[X\varepsilon_y]}}_{0}\\ &= \lambda\end{aligned}$$

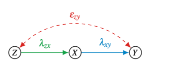
vs.



$\Rightarrow$

$$\begin{aligned}\sigma_{xy} &= \mathbb{E}[XY]\\ &= \mathbb{E}[X(\lambda X + \varepsilon_y)]\\ &= \lambda\mathbb{E}[XX] + \mathbb{E}[X\varepsilon_y]\\ &= \lambda + \mathbb{E}[\varepsilon_x\varepsilon_y]\\ &= \lambda + \varepsilon_{xy}.\end{aligned}$$

— Trek & Wright's Rule : $\sigma_{xy} = \sum$ (paths between X, Y).
↳ w/o self-intersecting. colliders.

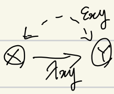


$$\begin{aligned}\sigma_{xy} &= (X \xrightarrow{\lambda_{xy}} Y) + (X \xleftarrow{\lambda_{zx}} Z \xleftrightarrow{\varepsilon_{zy}} Y)\\ &= \lambda_{xy} + \lambda_{zx}\varepsilon_{zy}.\end{aligned}$$

$\Rightarrow$ Let $X \xrightarrow{\alpha} Y$, $\sigma_{xy} = \alpha$ if $X \perp Y$ in G_α (G that edge α removed).

8.2. Regressions : when can/cannot be used to find causal effects?

— $\begin{cases} X \leftarrow \varepsilon_x. \\ Y \leftarrow \lambda_{xy} X + \varepsilon_y. \\ \varepsilon_x, \varepsilon_y \text{ correlated.}\end{cases}$ $\Rightarrow$ $Y = \beta X + e$ gives biased answer. ($^*\beta = \sigma_{xy}$ r_{yx}).

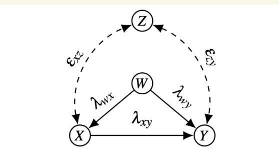


$$\begin{aligned}\sigma_{xy} &= \lambda_{xy}\mathbb{E}[XX] + \mathbb{E}[\varepsilon_x\varepsilon_y]\\ &= \lambda_{xy} + \varepsilon_{xy}\end{aligned}$$

< λ_{xy} is not-identifiable >.

— $(X \perp Y \mid Z) \Leftrightarrow r_{yx \cdot z} = 0$. $(Y = \beta X + \alpha Z)$

— Causal 하려면 구조 고려해야



$y = \beta X + e \longrightarrow \beta = \lambda_{xy} + \lambda_{wx}\lambda_{wy}$

$y = \beta X + \alpha W + \gamma Z + e \longrightarrow \beta = \lambda_{xy} - \dfrac{\varepsilon_{xz}\varepsilon_{zy}}{1 - \lambda_{wx}^2 - \varepsilon_{xz}^2}$

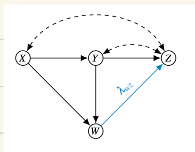
$y = \beta X + \alpha W + e. \longrightarrow \beta = \lambda_{xy}$

Thm (Single-Door) $X \xrightarrow{\lambda} Y$. λ is identifiable if $\exists Z$ s.t.

(i) Z contains no descendant of Y.

(ii) Z d-separates X from Y in G_λ.

$\rightarrow \lambda = r_{yx \cdot z}$



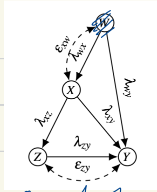
$\lambda_{xy} = r_{yx}$
$\lambda_{wz} = r_{zwyx}$

Thm (Back-Door) Total effect ($X \to Y$) is identifiable if $\exists Z$ s.t.

(i) no member of Z is descendant of X.

(ii) Z d-separates X from Y in $G_{\underline{X}}$

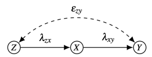
$\Rightarrow$ total effect $= r_{yx.Z}$



↳ total effect: $\lambda_{xz}\lambda_{zy} + \lambda_{xy}$

back door: $r_{yx.W}$

8.3. Algorithmic ID.



$$\Sigma = \underbrace{\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}}_{\text{computable}} = \underbrace{\begin{bmatrix} 1 & \lambda_{xy} + \lambda_{zx}\varepsilon_{zy} & \lambda_{zx} \\ \lambda_{xy} + \lambda_{zx}\varepsilon_{zy} & 1 & \lambda_{zx}\lambda_{xy} + \varepsilon_{zy} \\ \lambda_{zx} & \lambda_{zx}\lambda_{xy} + \varepsilon_{zy} & 1 \end{bmatrix}}_{\text{expressions with unobservables}}$$

$$\sigma_{xz} = \lambda_{zx}$$
$$\sigma_{xy} = \lambda_{xy} + \lambda_{zx}\varepsilon_{zy}$$
$$\sigma_{zy} = \lambda_{zx}\lambda_{xy} + \varepsilon_{zy}$$

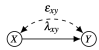
→ σ are known.

λ, ε unknown.

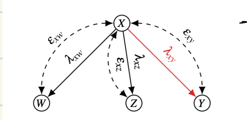
$\Rightarrow$ after substitution, full-rank equations.

— Non-identifiable,

Finite-identifiable

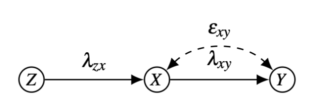


$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_{xy} + \varepsilon_{xy} \\ \lambda_{xy} + \varepsilon_{xy} & 1 \end{bmatrix}$$



→ quadratic in λ_{xy}.

— Efficiently search? : IV.



→ λ_{xy} : non-parametrically identify 불가.

but) $\sigma_{xy} = \lambda_{xy} + \varepsilon_{xy}$

$\sigma_{zy} = \lambda_{zx} \cdot \lambda_{xy}$

$\sigma_{xz} = \lambda_{xz}$

$\Rightarrow \lambda_{xy} = \dfrac{\sigma_{zy}}{\sigma_{zx}}$

Definition (Instrumental Variable (p. 248 [8]))

A variable Z is an **IV** for λ_{xy} from X to Y if

- Z is d-separated from Y in the subgraph $\mathcal{G}_{\lambda_{xy}}$,
- Z is not d-separated from X in $\mathcal{G}_{\lambda_{xy}}$.

(다른 방법: 2-SLS.

↳ Z가 여러개인 경우...

- Do the regression of X on Z:

$$X = \alpha Z + e \implies \alpha = \sigma_{zx}.$$

- Let $\hat{X} = \sigma_{zx} Z$ and regress Y on $\hat{X}$:

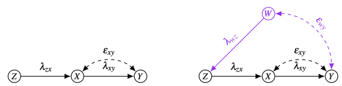
$$Y = \beta \hat{X} + e \implies \beta = \frac{\sigma_{zy}}{\sigma_{zx}}.$$

— Conditional IV: given W 일 상태에서 IV.

Definition (Conditional IV [3])

A variable Z qualifies as a conditional IV given a set $\mathbf{W}$ for structural coefficient λ_{xy} from X to Y if

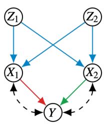
- $\mathbf{W}$ contains only non-descendants of Y
- $\mathbf{W}$ d-separates Z from Y in the subgraph $\mathcal{G}_{\lambda_{xy}}$,
- $\mathbf{W}$ does not d-separate Z from X in $\mathcal{G}_{\lambda_{xy}}$.



8.4. IV to Instrumental Sets

~ IV는 존재 X.

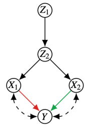
but 쉽게 해결 가능.



$$\sigma_{z_i x_j} = \lambda_{z_i x_j} \text{ for } i, j \in \{1, 2\}$$

$$\sigma_{z_1 y} = \lambda_{z_1 x_1}\lambda_{x_1 y} + \lambda_{z_1 x_2}\lambda_{x_2 y}$$
$$\sigma_{z_2 y} = \lambda_{z_2 x_1}\lambda_{x_1 y} + \lambda_{z_2 x_2}\lambda_{x_2 y}$$

- 실패하는 경우 : equations 가 collinear.



$$\sigma_{z_1y} = \sigma_{z_1x_1}\lambda_{x_1y} + \sigma_{z_1x_2}\lambda_{x_2y}$$
$$\sigma_{z_2y} = \sigma_{z_2x_1}\lambda_{x_1y} + \sigma_{z_2x_2}\lambda_{x_2y}$$

$\Rightarrow$

$$\lambda_{z_1z_2}\sigma_{z_2y} = \lambda_{z_1z_2}(\sigma_{z_2x_1}\lambda_{x_1y} + \sigma_{z_2x_2}\lambda_{x_2y})$$
$$\sigma_{z_2y} = \sigma_{z_2x_1}\lambda_{x_1y} + \sigma_{z_2x_2}\lambda_{x_2y}.$$

- Non-Degenerate $\Leftrightarrow$ Full matching btw $\{Z_1, Z_2\}$ and $\{X_1, X_2\}$ with non-intersecting paths.
  $\rightarrow$ can be found using a max-flow

  해석 가능.

9. Partial ID

9.1. Partial ID Problem

Causal Effect Partial-Identifiability

The causal effect of X on Y is said to be *partially-identifiable* given a causal diagram $\mathcal{G}$, if
- $P(\mathbf{V})$ determines a bound $P(y \mid do(x)) \in [a,b]$;
- the bound $[a,b]$ is a proper subset of $[0,1]$.

- $P(y \mid do(x))$ is identifiable $\Longrightarrow P(y \mid do(x))$ is partially identifiable.
- $P(y \mid do(x))$ is partially identifiable $\not\Longrightarrow P(y \mid do(x))$ is identifiable.

9.2. Natural Bounds

Theorem

Given the observational distribution $P(x,y)$, the causal effect $P(y \mid do(x))$ is bounded in $[a(y;x), b(y;x)]$ where

$$a(y;x) = P(y,x),$$
$$b(y;x) = P(y,x) + 1 - P(x).$$

$x \rightarrow do(x)$

- 하한 : 먹은 사람 당연히 ok.
- 상한 : 약 먹지 않은 사람 모두 효과.

- Tightness

Theorem

For binary $X, Y \in \{0,1\}$, given any distribution $P(x,y)$, there exist SCMs $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}$ such that

$$P^{(1)}(x,y) = P^{(2)}(x,y) = P(x,y)$$

while $P^{(1)}(y \mid do(x)) = a(y;x)$, $P^{(2)}(y \mid do(x)) = b(y;x)$.

$\rightarrow$ natural bounds are not tight 한 경우. (e.g. IV model w/ Z).

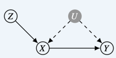
9.3. Canonical Type Models & LP Bounds

Theorem

Given the observational distribution $P(X,Y \mid Z)$, the causal effect $P(y \mid do(x))$ is bounded in $[a(y;x), b(y;x)]$ where

$$a(y;x) = \max_z P(y,x \mid z)$$
$$b(y;x) = \min_z P(y,x \mid z) + 1 - P(x \mid z).$$

$\rightarrow$ 이건 natural bound 보다 더 tight하니?

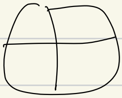


Applying natural bounds to $\mathcal{M}_z$

$$P(y \mid do(x,z)) \leq P(y,x \mid do(z)) + 1 - P(x \mid do(z))$$

However, natural bounds are not tight in IV models.

- $P(y \mid do(x))$ 에 대한 최적화 문제로 생각.
  - → 가능한 모든 SCM ?
  - → 적절하게 parameterization. (사람들의 특성을 ...).
- U를 selector로 생각: $Z \to X$를 선택.
  - ↳ $|\mathfrak{X}_X|, |\mathfrak{X}_Z| = 2$ 이므로 $f = 4$가지 ftn. ($f_X(Z)$)

$\mathfrak{X}_U =$  : 4가지 group. $\begin{pmatrix} 0, 0 \\ 0, 1 \\ 1, 0 \\ 1, 1 \end{pmatrix}$.

↳ R_X 로 정의.

$R_X = 0 \iff \{u \in \mathfrak{X}_U \mid f_X(Z{=}0,u) = 0,\ f_X(Z{=}1,u) = 0\}$ (never-taker)
$R_X = 1 \iff \{u \in \mathfrak{X}_U \mid f_X(Z{=}0,u) = 0,\ f_X(Z{=}1,u) = 1\}$ (complier)
$R_X = 2 \iff \{u \in \mathfrak{X}_U \mid f_X(Z{=}0,u) = 1,\ f_X(Z{=}1,u) = 0\}$ (defier)
$R_X = 3 \iff \{u \in \mathfrak{X}_U \mid f_X(Z{=}0,u) = 1,\ f_X(Z{=}1,u) = 1\}$ (always-taker)

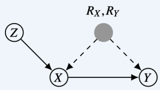
→ Y에 대해서도 4 종류. (R_Y).

⇒ $q_{j,k} = P(R_X = j, R_Y = k)$.

- Canonical SCM.

Definition

Consider an SCM $\mathcal{M}'$ with $X, Y, Z \in \{0,1\}$ and $U = \{R_X, R_Y, R_Z\}$ where $R_X, R_Y \in \{0,1,2,3\}$, $R_Z \in \{0,1\}$.



$$f_X(z, r_X) = \begin{cases} 0 & \text{if } r_X = 0 \quad \text{never-taker} \\ z & \text{if } r_X = 1 \quad \text{complier} \\ 1 - z & \text{if } r_X = 2 \quad \text{defier} \\ 1 & \text{if } r_X = 3 \quad \text{always-taker} \end{cases}$$

$$\mathcal{M}' = \begin{cases} X \leftarrow f_X(z, r_X) \\ Y \leftarrow f_Y(x, r_Y) \\ Z \leftarrow r_Z \end{cases} \qquad f_Y(x, r_Y) = \begin{cases} 0 & \text{if } r_Y = 0 \quad \text{never-recover} \\ x & \text{if } r_Y = 1 \quad \text{helped} \\ 1 - x & \text{if } r_Y = 2 \quad \text{hurt} \\ 1 & \text{if } r_Y = 3 \quad \text{always-recover} \end{cases}$$

⇒ Bound 설정 가능.

- $P(Y=y \mid do(X=x))$를 $q_{j,k}$의 linear ftn으로 생각.
  - ⇒ Bound는 다음과 같은 선형 program 최적화로 구할 수 있음. $[a_{y,x}, b_{y,x}]$

$$a_{10} = \min_{\mathbf{q}} q_{0,2} + q_{1,2} + q_{2,2} + q_{3,2} + q_{0,3} + q_{1,3} + q_{2,3} + q_{3,3}$$

$$b_{10} = \max_{\mathbf{q}} q_{0,2} + q_{1,2} + q_{2,2} + q_{3,2} + q_{0,3} + q_{1,3} + q_{2,3} + q_{3,3}$$

such that $\sum_{j=0}^{3}\sum_{k=0}^{3} q_{j,k} = 1$, $\forall_{j,k \in \{0,1,2,3\}} q_{j,k} \geq 0$, and

$$p_{00.0} = q_{0,0} + q_{0,1} + q_{1,0} + q_{1,1}, \qquad p_{00.1} = q_{0,0} + q_{0,1} + q_{2,0} + q_{2,1},$$
$$p_{01.0} = q_{2,0} + q_{2,2} + q_{3,0} + q_{3,2}, \qquad p_{01.1} = q_{1,0} + q_{1,2} + q_{3,0} + q_{3,2},$$
$$p_{10.0} = q_{0,2} + q_{0,3} + q_{1,2} + q_{1,3}, \qquad p_{10.1} = q_{0,2} + q_{0,3} + q_{2,2} + q_{2,3},$$
$$p_{11.0} = q_{2,1} + q_{2,3} + q_{3,1} + q_{3,3}, \qquad p_{11.1} = q_{1,1} + q_{1,3} + q_{3,1} + q_{3,3}.$$

$\xrightarrow{\text{closed-form sol.}}$

$$P(Y=1 \mid do(X=0)) \geq \max \begin{cases} p_{10.0} \\ p_{10.1} \\ p_{10.0} + p_{11.0} - p_{00.1} - p_{11.1} \\ p_{01.0} + p_{10.0} - p_{00.1} - p_{01.1} \end{cases}$$

$$P(Y=1 \mid do(X=0)) \leq \min \begin{cases} 1 - p_{00.0} \\ 1 - p_{00.1} \\ p_{10.0} + p_{11.0} + p_{01.1} + p_{10.1} \\ p_{01.0} + p_{10.0} + p_{10.1} + p_{11.1} \end{cases}$$

↳ $P(y, x \mid z) = p_{yx.z}$