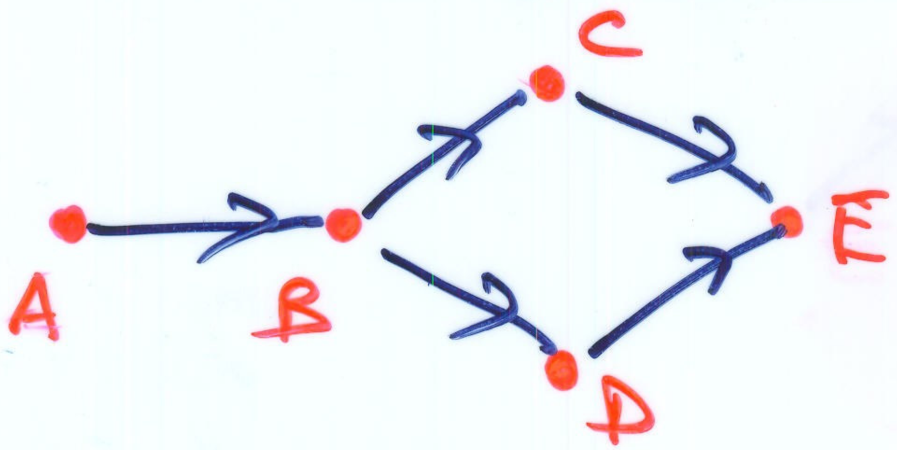
USING INFLUENCE DIAGRAMS FOR CAUSAL INFERENCE

PHILIP DAWID

UNIVERSITY
OF
CAMBRIDGE

INTERPRETATIONS OF DAGs



① . Conditional independence

$C \perp\!\!\!\perp A \mid B$ $\qquad$ $E \perp\!\!\!\perp (A, B) \mid C, D$

$D \perp\!\!\!\perp (A, C) \mid B$ $\qquad$ $F \perp\!\!\!\perp (A, B, C, D) \mid E$

② . Probabilistic causality (Suppes)

A causes B

B is a common cause of C + D

$A \to B \to C \to E$ is a causal pathway

[F is not relevant]

③ . Statistical causality (Pearl)

Manipulating C does not affect distribution of $E \mid C, D$

Under CI,

$A \longrightarrow B$ and $A \longleftarrow B$

are equivalent

— not so for causal interpretations.

— POTENTIAL CONFUSION!

Is the graph itself fundamental?

How should we interpret e.g. direction of an arrow?

Dangers of "reification"

Can achieve clarity, while sticking to CI semantics, by using

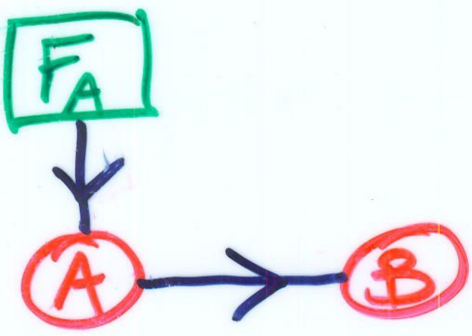
INFLUENCE DIAGRAMS

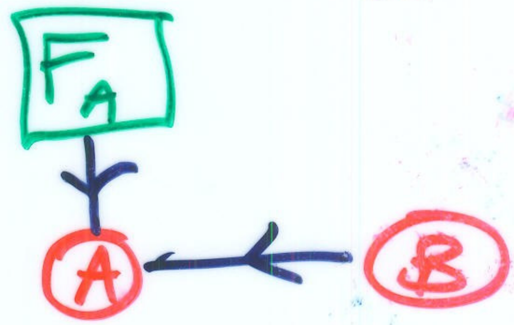
– including DECISION NODES

– e.g. an intervention node

$F_A = a$ "set A to a"

$F_A = \emptyset$ "hands off!"





$B \perp\!\!\!\perp F_A \mid A$ $\qquad$ $B \perp\!\!\!\perp F_A$

– now (graphically) inequivalent

WHICH (IF EITHER) IS APPROPRIATE?

Moralization criterion

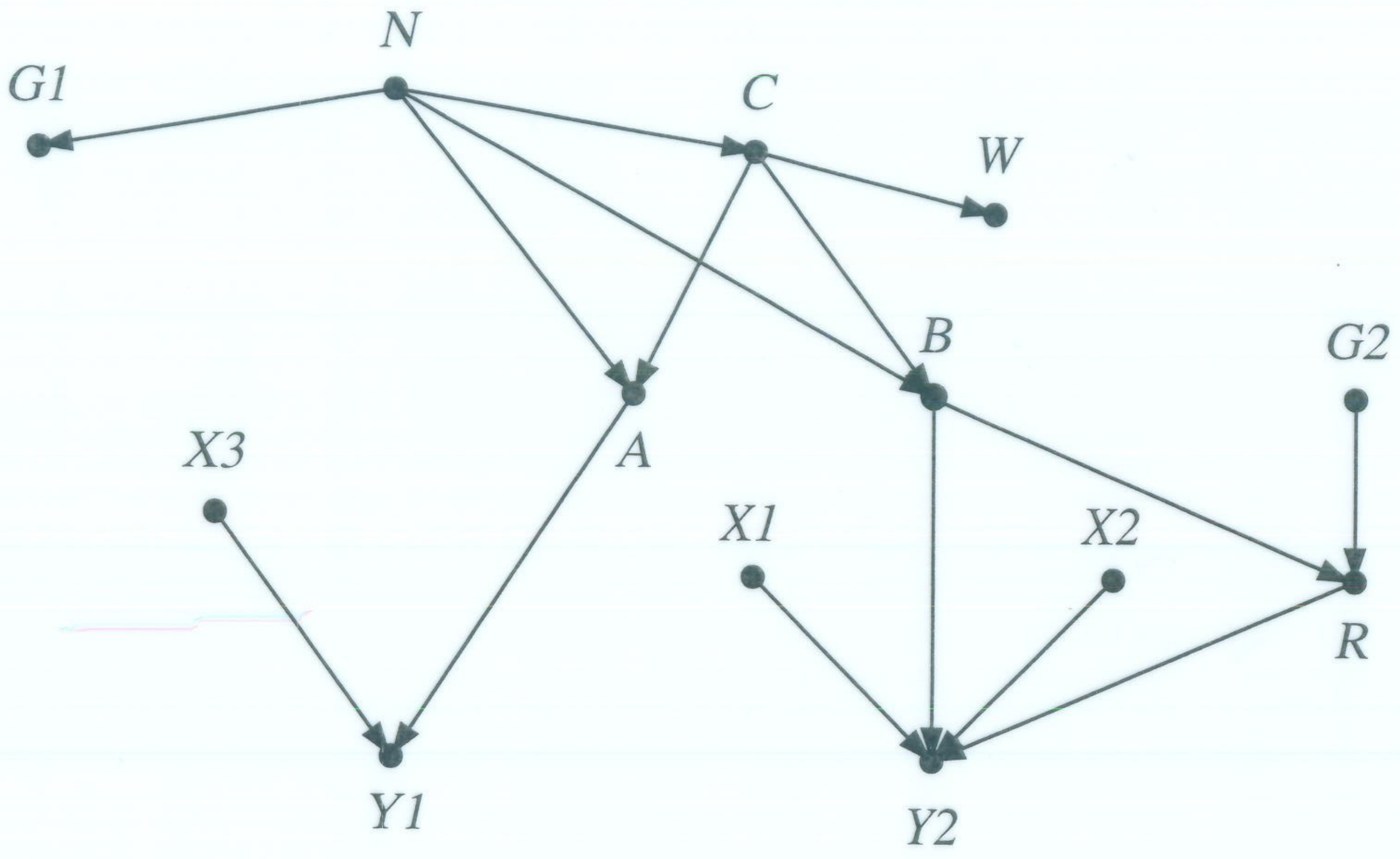
DAG $\mathcal{D}$ — or ID

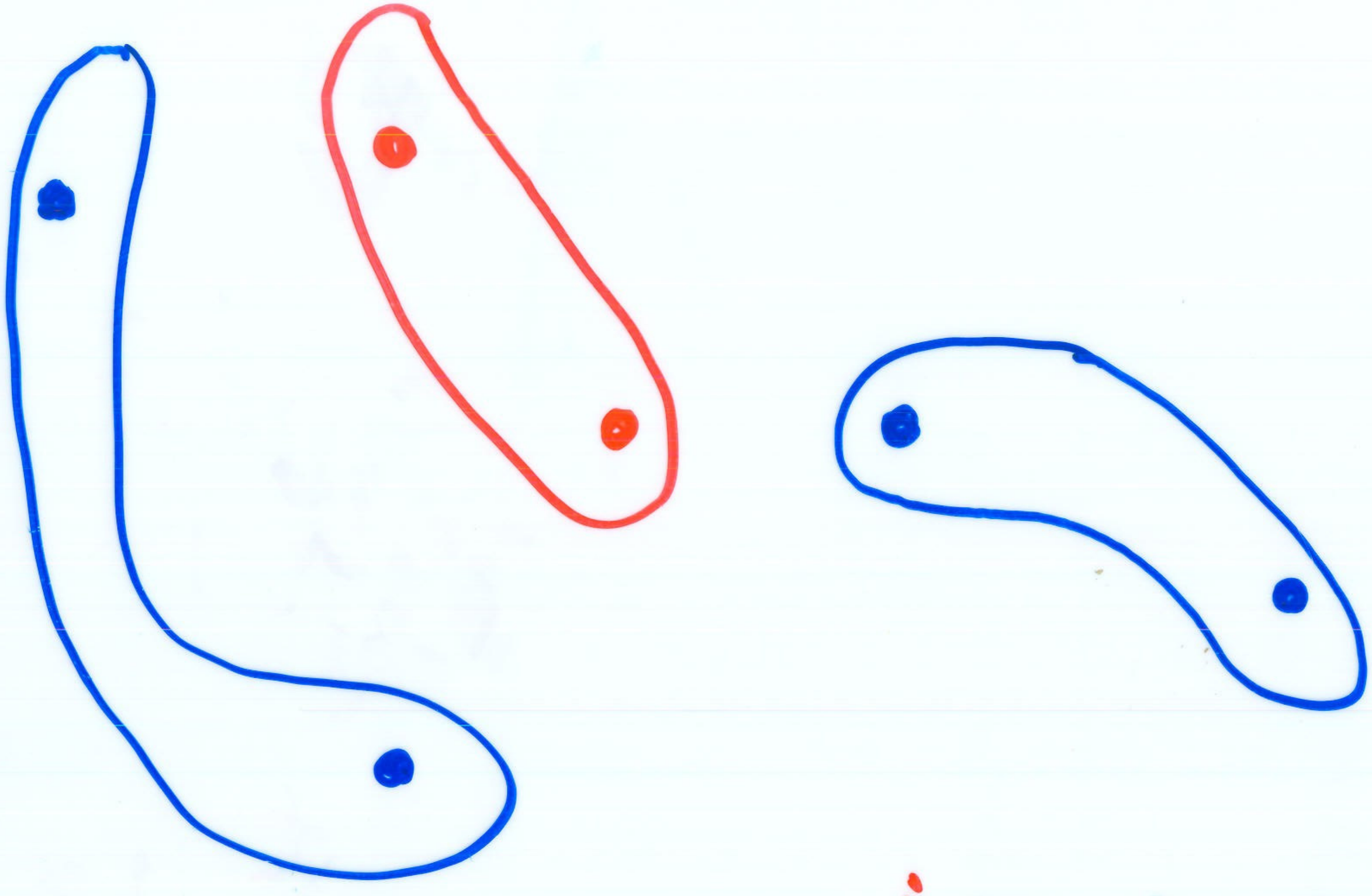
Node sets A, B, C

$A \perp\!\!\!\perp_{\mathcal{D}} B \mid C$ if, in the MORALIZED ANCESTRAL GRAPH of $A \cup B \cup C$, any path from A to B intersects C.

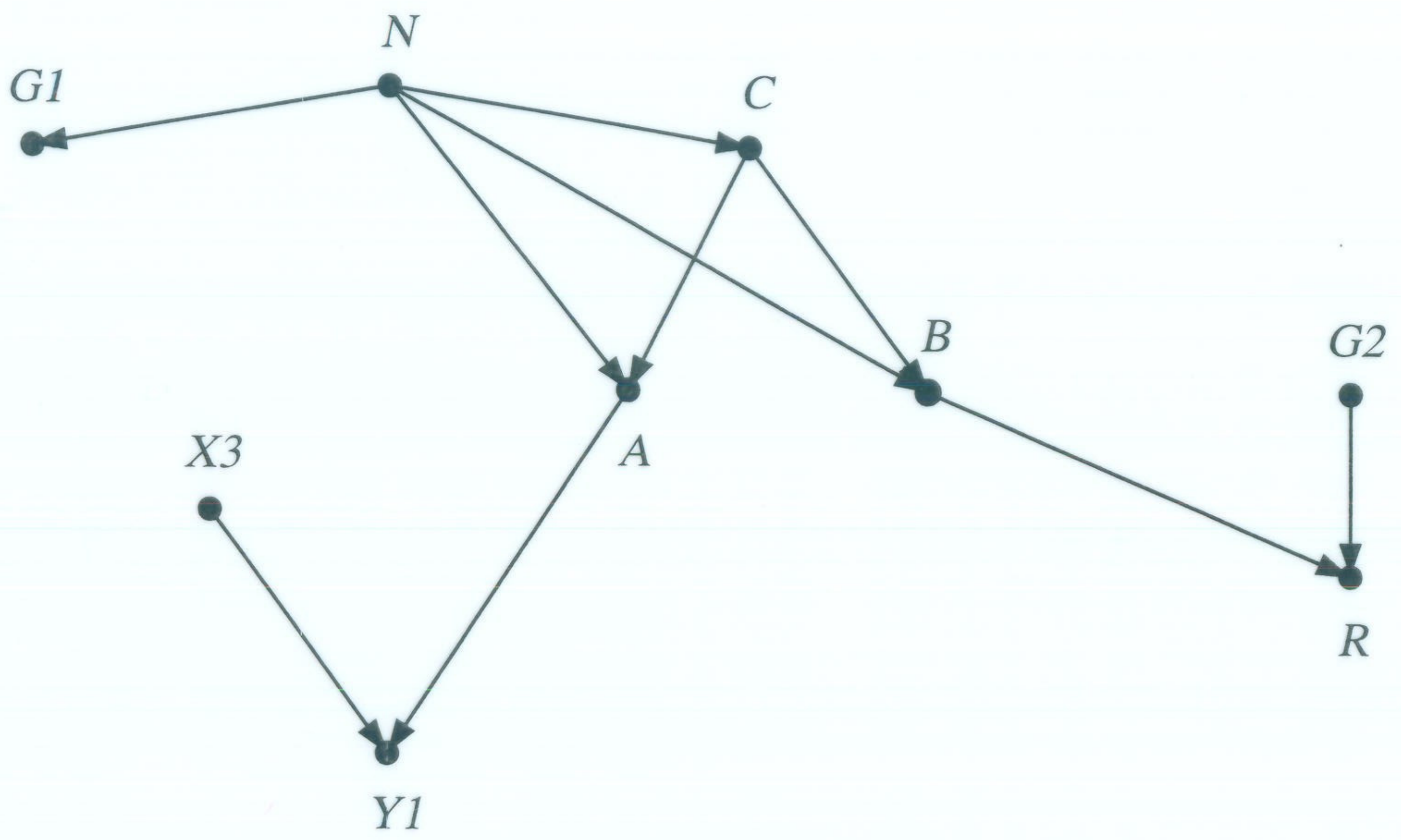
If P is directed Markov wrt $\mathcal{D}$, this $\Rightarrow$

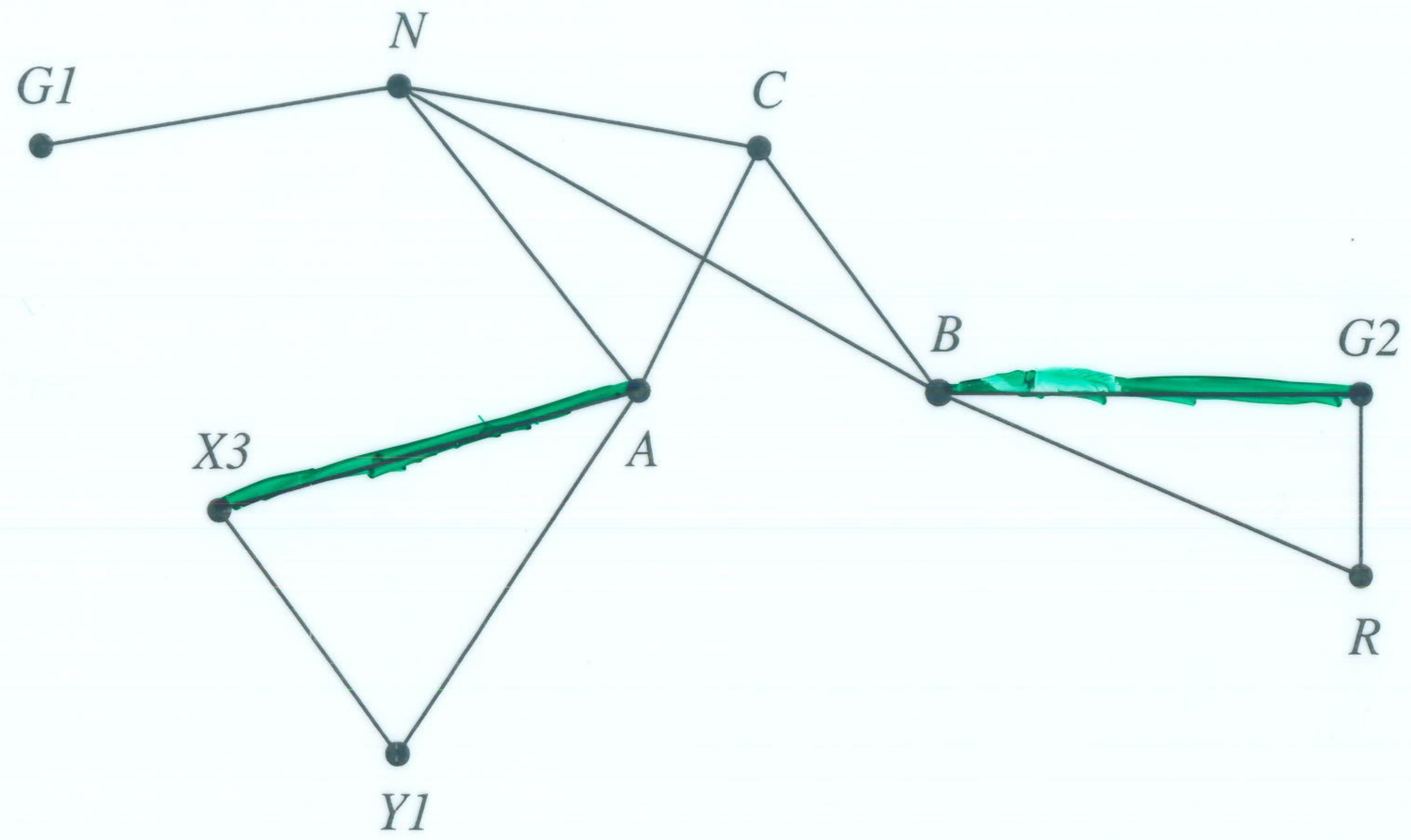
$A \perp\!\!\!\perp_{P} B \mid C$


N
G1
C
W
B
G2
X3
A
X1
X2
R
Y1
Y2



¿$(G1, Y_1) \perp\!\!\!\perp (B, R) \mid (N, A)$?


N
G1
C
B
G2
X3
A
R
Y1

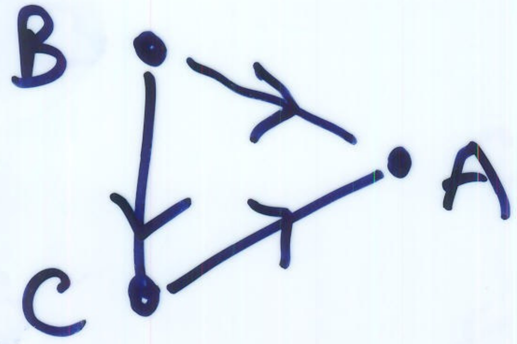

G1
N
C
B
G2
X3
A
R
Y1

CAUTION –

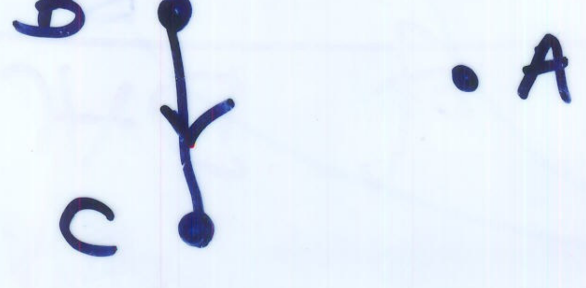
DAGs can lie!

Non-theorem

$$\left.\begin{array}{l} A \perp\!\!\!\perp B \mid C \\ A \perp\!\!\!\perp C \mid B \end{array}\right\} \Rightarrow A \perp\!\!\!\perp (B, C)$$



- Can remove $B \to A$
- Can remove $C \to A$



A *causal DAG* is a DAG in which:

(1) the lack of an arrow from V_j to V_m can be interpreted as the absence of a direct causal effect of V_j on V_m (relative to the other variables on the graph); and

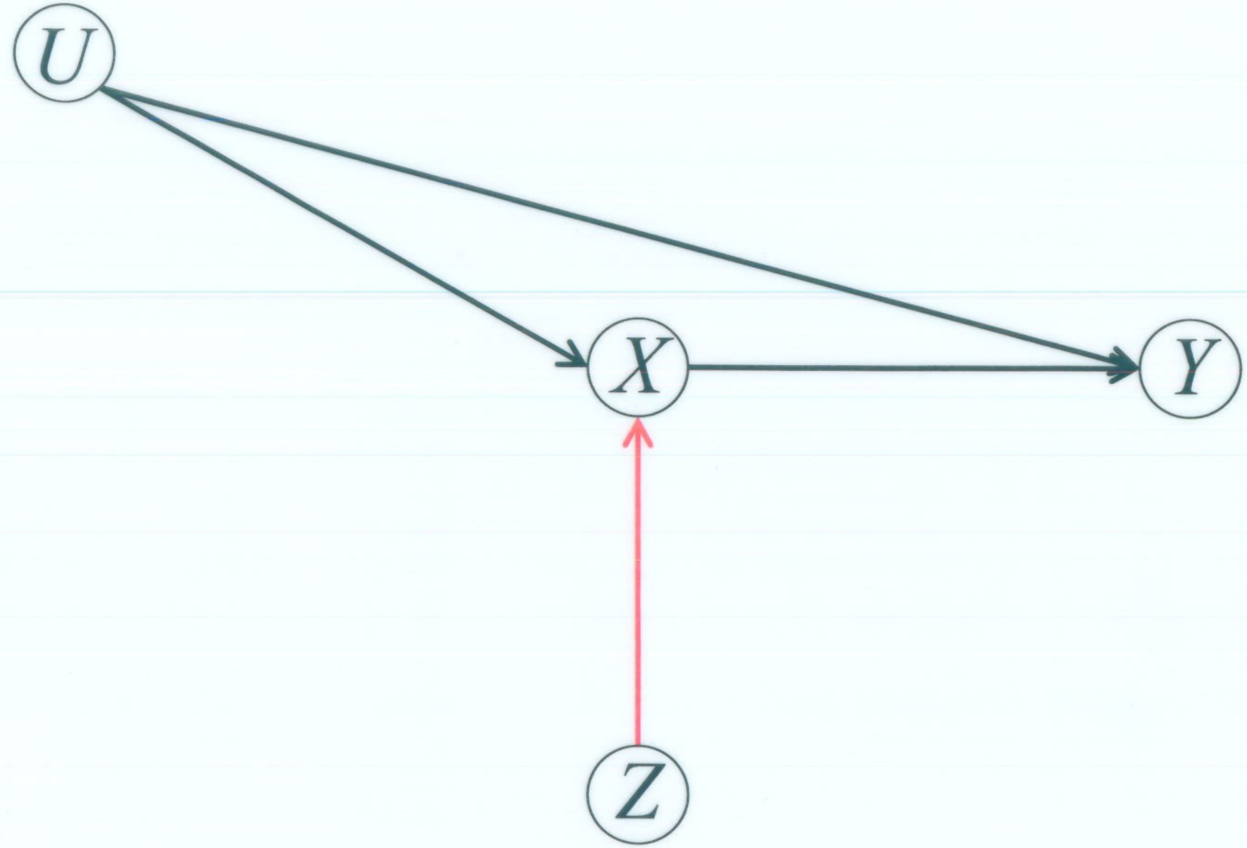
(2) all common causes, even if unmeasured, of any pair of variables on the graph are themselves on the graph...In Figure 2... the inclusion of the measured variables (Z, X, Y) implies that the causal DAG must also include their unmeasured common causes (U, U^*).

Hernan and Robins, Epidemiology (2006)

Instrumental variable

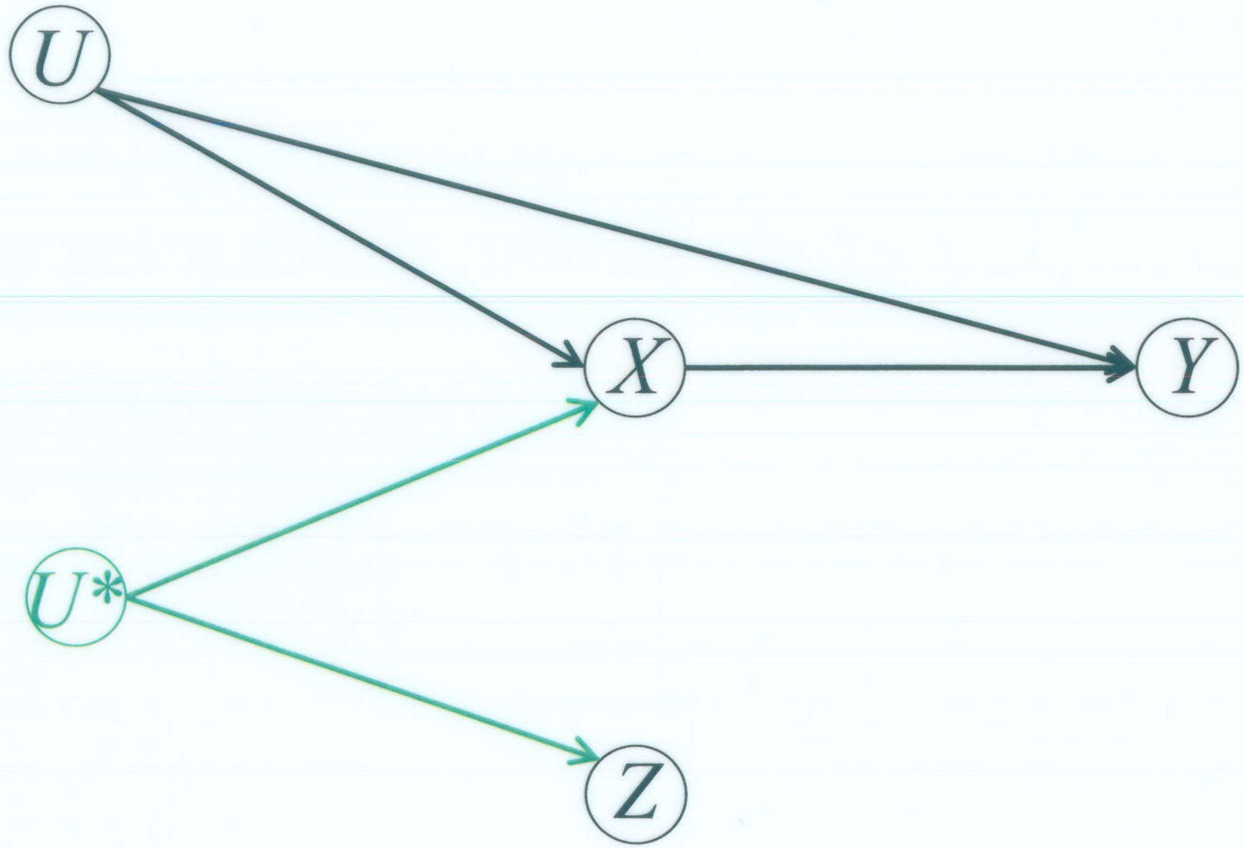
Hernan and Robins, Epidemiology (2006)

Figure 1



Instrumental variable?

Figure 2



$$
\begin{aligned}
Z &\perp\!\!\!\perp U \\
Y &\perp\!\!\!\perp Z \mid (X, U)
\end{aligned}
$$

Instrumental variable

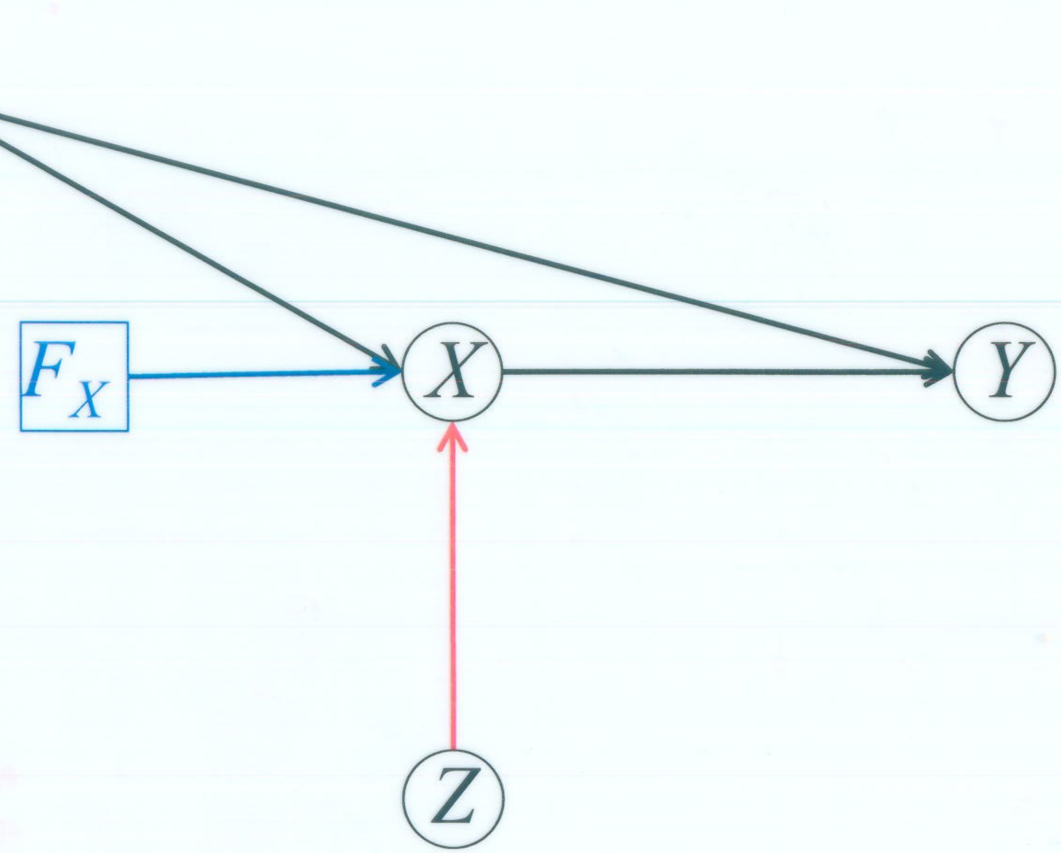
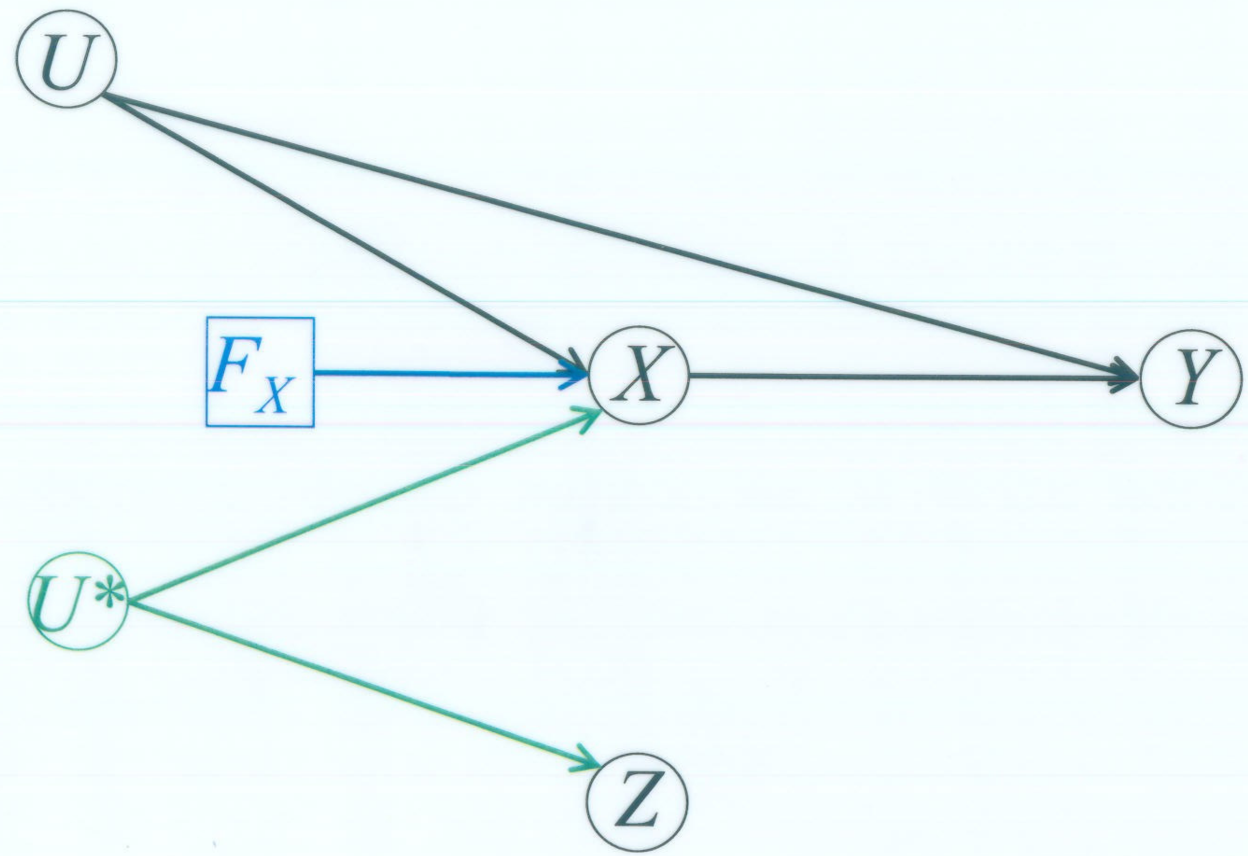


Figure 1

Instrumental variable?

Figure 2



$$\begin{aligned} (Z, U) &\perp\!\!\!\perp F_X \\ Z &\perp\!\!\!\perp U \mid F_X \\ Y &\perp\!\!\!\perp (Z, F_X) \mid (X, U) \end{aligned}$$

$Y \perp\!\!\!\perp Z \mid F_X = x$

CAUSAL EFFECT

of A on B

is a contrast between distributions of

B given $F_A = a$

as a varies

- e.g.

$$E(B \mid F_A = 1) - E(B \mid F_A = 0)$$

- average causal effect (ACE)

Observational study

SUPPOSE

$$Y \perp\!\!\!\perp F_T \mid T \qquad \circledast$$

$$\boxed{F_T} \longrightarrow T \longrightarrow Y$$

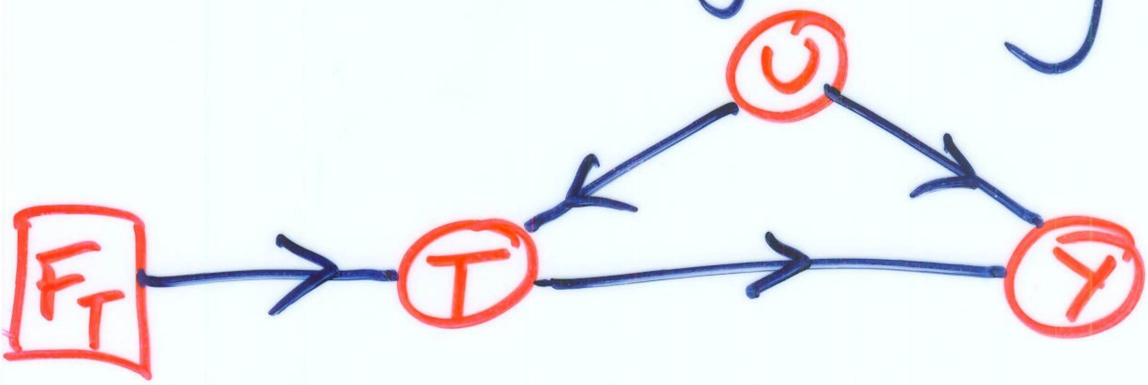
Then

$$\begin{aligned} Y \mid F_T = t &= Y \mid F_T = t,\ T = t \\ &= Y \mid F_T = \emptyset,\ T = t \end{aligned}$$

– observationally estimable.

What if $\circledast$ fails?

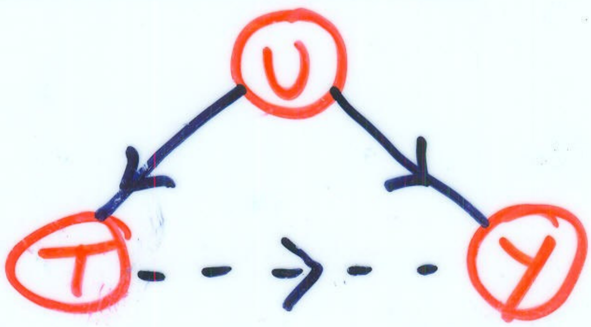
(Un-) Confounding



$$U \perp\!\!\!\perp F_T$$
$$Y \perp\!\!\!\perp F_T \mid (T, U)$$

U is a sufficient covariate for effect of T on Y

Pearlian version:



"Back-door" criterion

Causal Effect

$$ACE := E(Y \mid \underset{(T=1)}{F_T = 1}) - E(Y \mid \underset{(T=0)}{F_T = 0})$$

– not $E(Y \mid T=1, F_T = \emptyset) - E(Y \mid T=0, F_T = \emptyset)$

<u>unless</u> $Y \perp\!\!\!\perp F_T \mid T$

But if U is suff. covt.,

$E(Y \mid F_T = 1) =$

$E[\,E(Y \mid F_T = 1, U) \mid F_T = 1\,]$

$= E[\,E(Y \mid F_T = 1, T = 1, U) \mid F_T = 1\,]$

$(F_T = 1 \Rightarrow T = 1)$

$= E[\,E(Y \mid F_T = \emptyset, T = 1, U) \mid F_T = 1\,]$

$(Y \perp\!\!\!\perp F_T \mid T, U)$

$= E[\,E(Y \mid F_T = \emptyset, T = 1, U) \mid F_T = \emptyset\,]$

$(U \perp\!\!\!\perp F_T)$ – obs. est.

Pearl's "do-calculus"

Write $p(y \mid x, \check{z})$ for

$$\Pr(Y=y \mid \underset{(F_X=\phi)}{X=x}, F_Z=z)$$

Have CI relations between random & intervention variables

① $Y \perp\!\!\!\perp Z \mid X, F_X \neq \phi, W$

$\Rightarrow p(y \mid \check{x}, z, w) = p(y \mid \check{x}, w)$

② $Y \perp\!\!\!\perp F_Z \mid X, F_X \neq \phi, Z, W$

$\Rightarrow p(y \mid \check{x}, \check{z}, w) = p(y \mid \check{x}, z, w)$

③ $Y \perp\!\!\!\perp F_Z \mid X, F_X \neq \phi, W$

$\Rightarrow p(y \mid \check{x}, \check{z}, w) = p(y \mid \check{x}, w)$

[COMPLETE for DAGs]

If we have a DAG representation, (augmented)

can easily check CI

condition graphically (d-separation / moralisation)

[– conditioning on $(X, F_X \neq \emptyset)$

we can remove arrows into X]

Pearl has equivalent but more

complex conditions on "core" graph $\mathcal{C}$

e.g. 3

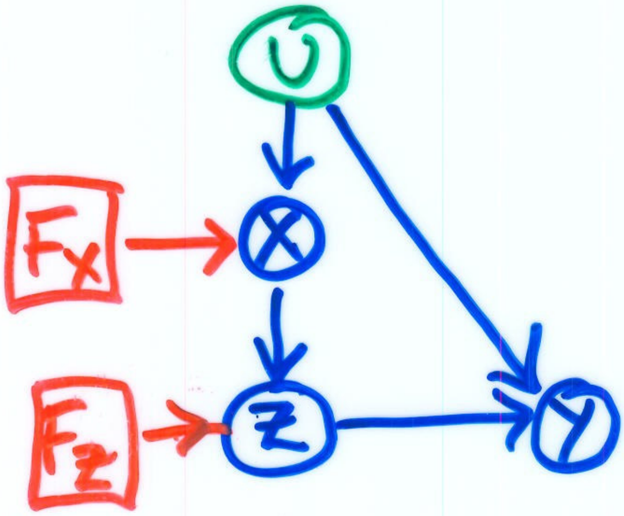
$$Y \perp_{\mathcal{D}_{\bar{X}}} F_Z \mid X, W$$

$$\Updownarrow$$

$$Y \perp_{\mathcal{C}_{\overline{X \cup (Z \setminus an(W))}}} Z \mid X, W$$

– unnecessary complication

Front Door Criterion



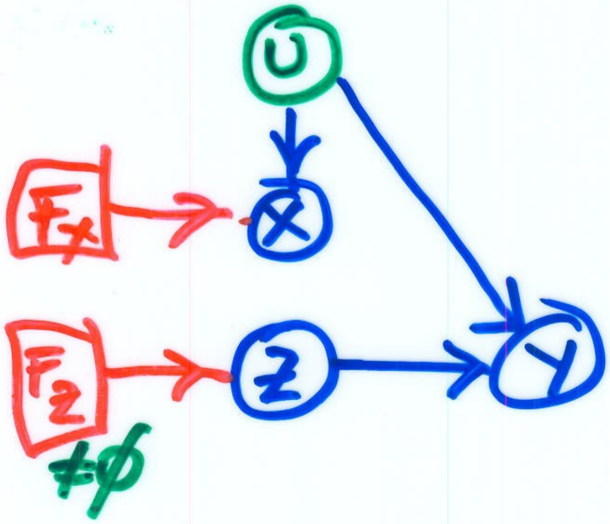
① $X \perp\!\!\!\perp F_Z \mid F_X$

② $Z \perp\!\!\!\perp F_X \mid X, F_Z$

③ $Y \perp\!\!\!\perp F_Z \mid X, Z, F_X$

X a sufft. covariate for eff. of Z on Y

→ can identify $p(y \mid \check{z})$



④ $Y \perp\!\!\!\perp F_X \mid Z, \underline{F_Z \neq \emptyset}$

$$p(y \mid \check{x}) = \sum_z \underbrace{p(y \mid \check{x}, z)}_{\overset{③}{=}\; p(y \mid \check{x}, \check{z}) \;\overset{④}{=}\; p(y \mid \check{z})\ ✓} \; \underbrace{p(z \mid \check{x})}_{\overset{②}{=}\; p(z \mid x)\ ✓}$$

Reduction of sufficient covariate

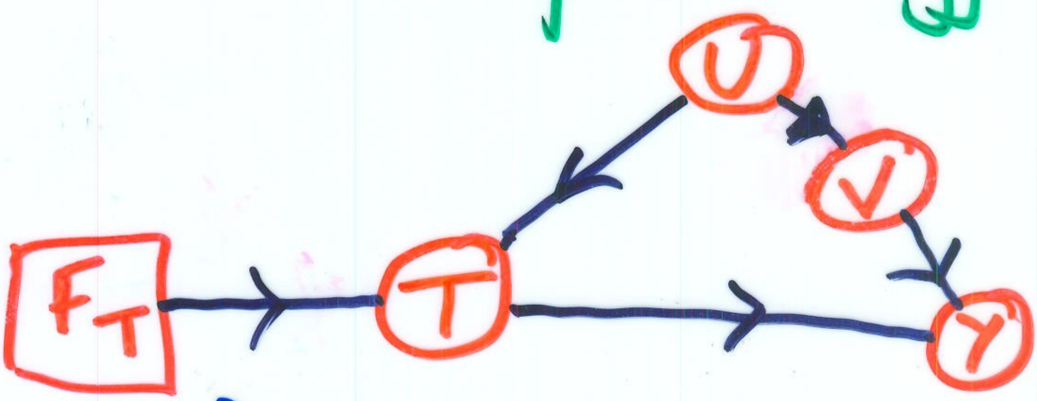
$U \perp\!\!\!\perp F_T$

$Y \perp\!\!\!\perp F_T \mid (T, U)$

V a function of U s.t.

$Y \perp\!\!\!\perp U \mid (T, V)$

– "response sufficient"



$\stackrel{?}{\Rightarrow}$ V is a sufft. cov.

PROPENSITY ANALYSIS

U a sufficient covariate

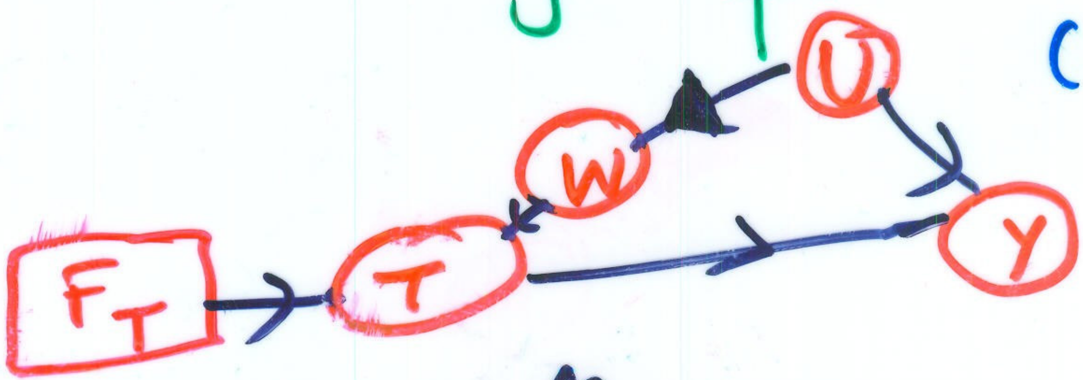
W a function of U s.t.

$$T \perp\!\!\!\perp U \mid W, F_T \quad (=\emptyset)$$

- e.g. (T binary)

$$W = \mathrm{pr}(T=1 \mid U, F_T=\emptyset)$$

- or, W is sufficient for observational distributions of "data" U given "parameter" T. (LR)



W is a suff. covariate

DYNAMIC DECISION MAKING

Data Situation

$A_1, \ldots, A_N$ "action" variables $\longrightarrow$ can be 'manipulated'

$L_1, \ldots, L_N$ covariates $\longrightarrow$ (available) background information

$Y = L_{N+1}$ response variable

all measured over time, L_i before A_i

$\mathbf{A}^{<i} = (A_1, \ldots, A_{i-1})$ past up to before i; $\mathbf{A}^{\leq i}$, $\mathbf{A}^{>i}$ etc. similarly

$$L_1 \rightarrow A_1 \rightarrow L_2 \rightarrow A_2 \cdots \rightarrow A_N \rightarrow Y$$

time $\rightarrow$

Example

Consider patients receiving anticoagulant treatment $\Rightarrow$ has to be monitored and adjusted.

$$L_i = \text{blood test results, other health indicators.}$$

$$A_i = \text{dose of anticoagulant drug.}$$

Plausibly, 'optimal' dose A_i will be a function of $\mathbf{L}^{\leq i}$ (and poss. $\mathbf{A}^{<i}$).

Strategies

Strategy $\mathbf{s} = (s_1, \ldots, s_N)$ set of functions assigning an action

$a_i = s_i(\mathbf{a}^{<i}, \mathbf{l}^{\leq i})$ to each history $(\mathbf{a}^{<i}, \mathbf{l}^{\leq i})$

(Could be stochastic, then dependence on $\mathbf{a}^{<i}$ relevant.)

Also called: **conditional** / **dynamic** / **adaptive** strategies.

Evaluation

Let $p(\cdot;\mathbf{s})$ be distribution under strategy $\mathbf{s}$.

Let $k(\cdot)$ be a loss function. Want to evaluate $E(k(Y);\mathbf{s})$.

Define

$$f(\mathbf{a}^{\leq j},\mathbf{l}^{\leq i}) := E\{k(Y)|\mathbf{a}^{\leq j},\mathbf{l}^{\leq i};\mathbf{s}\} \quad i=1,\ldots,N; j=i-1,i.$$

Then obtain $f(\emptyset)=E(k(Y);\mathbf{s})$ from $f(\mathbf{a}^{\leq N},\mathbf{l}^{\leq N})$ iteratively by:

$$f(\mathbf{a}^{<i},\mathbf{l}^{\leq i}) = \sum_{a_i} \underbrace{p(a_i|\mathbf{a}^{<i},\mathbf{l}^{\leq i};\mathbf{s})}_{\text{known by s}} \times f(\mathbf{a}^{\leq i},\mathbf{l}^{\leq i})$$

$$f(\mathbf{a}^{<i},\mathbf{l}^{<i}) = \sum_{l_i} \underbrace{p(l_i|\mathbf{a}^{<i},\mathbf{l}^{<i};\mathbf{s})}_{?} \times f(\mathbf{a}^{<i},\mathbf{l}^{\leq i}).$$

(cf. extensive form analysis)

Identifiability

Problem: doctors are following their 'gut feeling' (and not a specific strategy) in modifying the dose of anticoagulant drug.

Identifiability: *can* we find the optimal strategy from such data?

In particular: $p(l_i|\mathbf{a}^{<i}, \mathbf{l}^{<i}; \mathbf{s})$ then not known.

But could *estimate* $p(l_i|\mathbf{a}^{<i}, \mathbf{l}^{<i}; o)$ under observational regime.

Introduce indicator

$$\sigma = \begin{cases} o, & \text{observational regime} \\ s, & \mathbf{s} \in \mathcal{S} = \text{ set of strategies} \end{cases}$$

Simple Stability

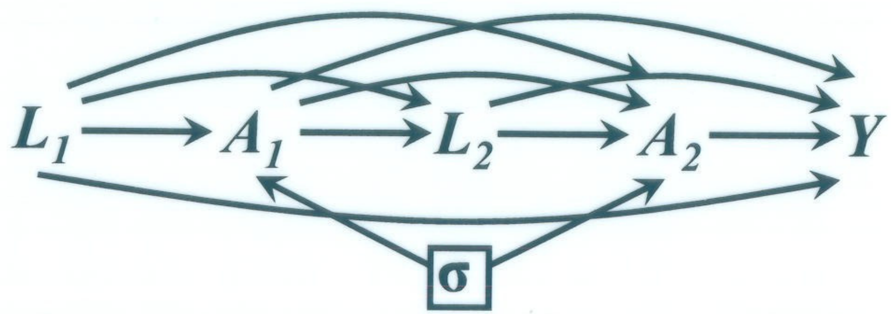
Sufficient for identifiability is

$$p(l_i|\mathbf{a}^{<i},\mathbf{l}^{<i};\mathrm{s}) = p(l_i|\mathbf{a}^{<i},\mathbf{l}^{<i};o) \quad \text{for all } i = 1,\ldots,N+1$$

or (via intervention indicator)

$$L_i \perp\!\!\!\perp \sigma | (\mathbf{A}^{<i},\mathbf{L}^{<i}) \quad \text{for all } i = 1,\ldots,N+1$$

Or graphically:

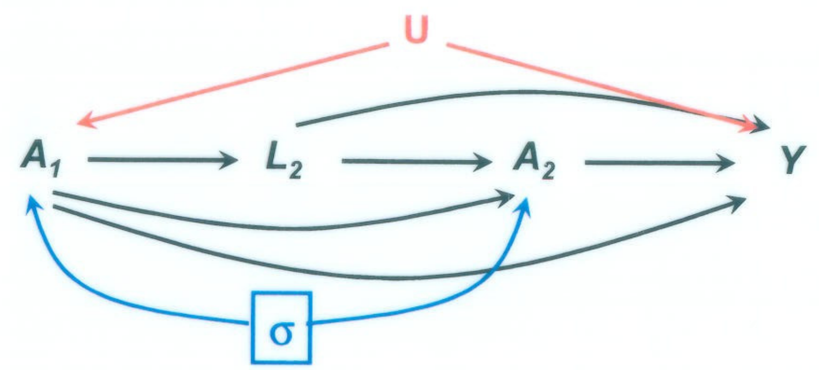


Extended Stability

Might not be able to assess simple stability without taking unobserved variables into account.

$\Rightarrow$ extend covariates $\mathbf{L}$ to include unobserved / hidden variables $\mathbf{U} = (U_1, \ldots, U_N)$ and check if simple stability can be deduced.

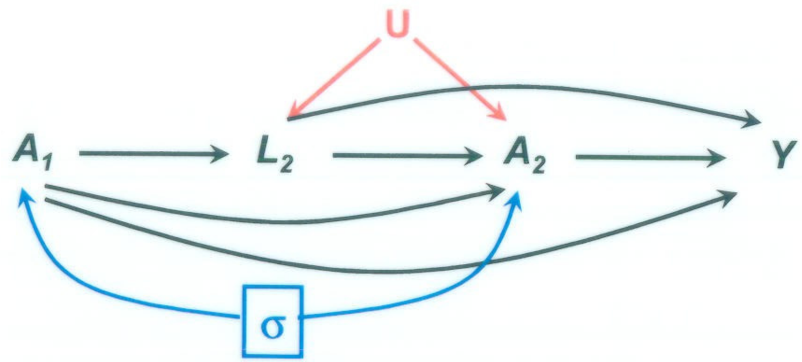
Example 1: particular underlying structure (note: $L_1 = \emptyset$)



Simple stability violated as $Y \not\perp\!\!\!\perp \sigma \mid (A_1, A_2, L_2)$.

Examples

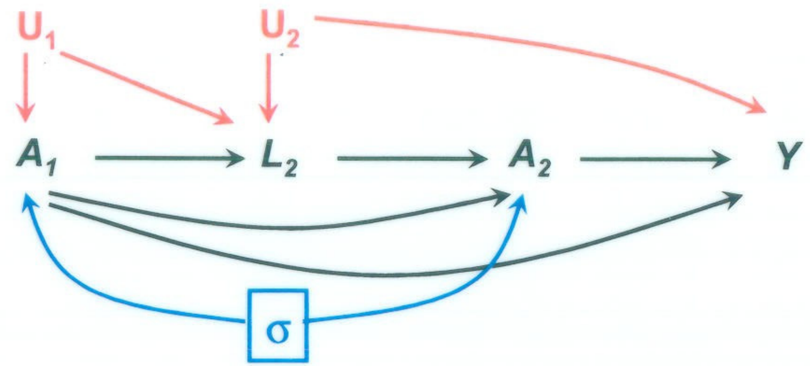
Example 2: different underlying structure



Simple stability satisfied.

Examples

Example 3: another different underlying structure



Simple stability violated: $L_2 \not\perp\!\!\!\perp \sigma \mid A_1$ and $Y \not\perp\!\!\!\perp \sigma \mid (A_1, A_2, L_2)$

SUMMARY

- Causal relationships can be clearly expressed in terms of conditional independence (involving decision variables as well as stochastic variables)

- When DAG-representable, we can apply standard semantics and manipulations

How about other (graphical/algebraic/...) representations of CI structures?