Tutorial on Graphical Models and Causal Inference

Vanessa Didelez

Department of Statistical Science
UNIVERSITY COLLEGE LONDON

Perugia, April 2005

Overview

- Graphical Models
  - undirected graphs
  - directed graphs (DAGs)
- Causality
  - causal graphs
  - intervention graphs
- Identifiability
  - ‘back-door’ / ‘front-door’ criteria
  - lots of examples
- Graphs and counterfactuals
- Further topics
  - sequential decisions
  - instrumental variables
  - direct / indirect effects

Graphical Models

Statistical models

- for multivariate rv's $\mathbf{X} = (X_1, \ldots, X_K)$
- characterized through specific conditional independencies
- that have to be representable in a graph.

$\rightarrow$ Graph $\mathcal{G} = (V, E)$ with

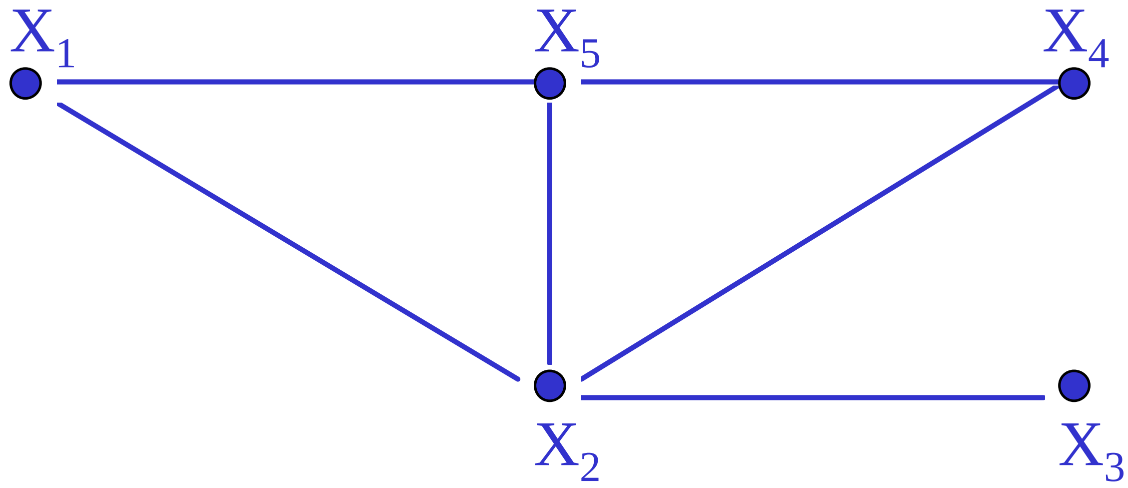
$V = \{1, \ldots, K\}$ set of vertices $\hat{=}$ variables

$E \subset V \times V$ set of edges $\hat{=}$ conditional dependence

Distinguish:

- Undirected graphs (edges: $\{i, j\}$, $i, j \in V$)
- Directed graphs (edges: (i, j))

Example:



$V = \{1, \ldots, 5\}$, $E = \{\{1,2\},\{1,5\},\{2,4\},\{2,5\},\{3,2\},\{4,5\}\}$

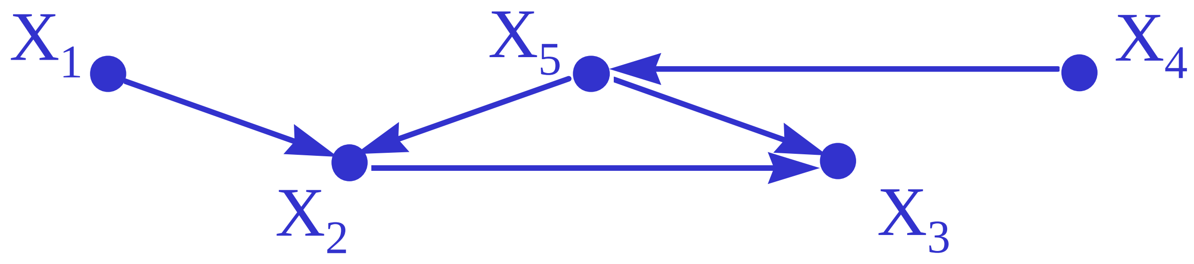
$\Rightarrow X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_5)$ etc.

Definition:

Distribution $P_{\mathbf{X}}$ is called $\mathcal{G}$–Markov, if for all disjoint $A, B, C \subset V$:

C separates A and B in $\mathcal{G} \Rightarrow \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$.

Example



$V = \{1, \ldots, 5\}$, $E = \{(1,2), (2,3), (4,5), (5,2), (5,3)\}$

Notations: $A \subset V$

$\mathrm{pa}(A) \hat{=}$ parents	$\mathrm{ch}(A) \hat{=}$ children	$\mathrm{de}(A) \hat{=}$ descendants
$\mathrm{an}(A) \hat{=}$ ancestors	$\mathrm{nd}(A) \hat{=}$ non–descendants	$\mathrm{An}(A) = \mathrm{an}(A) \cup \{A\}$

Definition: Distribution $P_{\mathbf{X}}$ is called $\mathcal{G}$–Markov, if

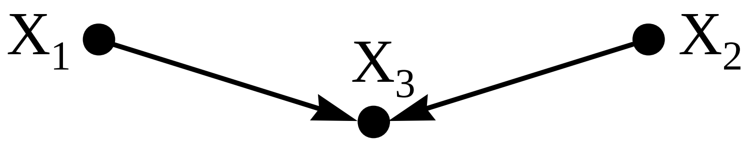
$$X_i \perp\!\!\!\perp \mathbf{X}_{\mathrm{nd}(i)\setminus\mathrm{pa}(i)} \mid \mathbf{X}_{\mathrm{pa}(i)} \quad \forall i \in V.$$

$\Leftrightarrow$ factorization: $P(X_1, \ldots, X_K) = \prod_{i \in V} P(X_i | \mathbf{X}_{\mathrm{pa}(i)})$

$\rightarrow$ system of univariate regressions

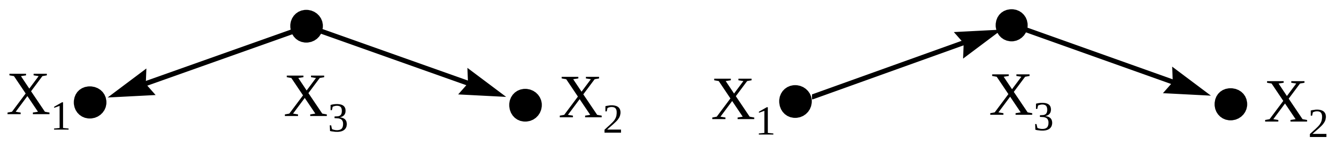
Important for the interpretation:

- conditioning on common child (selection) $\Rightarrow$ dependence



here: $X_1 \perp\!\!\!\perp X_2$, but $X_1 \not\perp\!\!\!\perp X_2 \mid X_3$

- marginalizing w.r.t. common parent (confounder) or intermediate variables $\Rightarrow$ dependence



here: $X_1 \perp\!\!\!\perp X_2 \mid X_3$, but $X_1 \not\perp\!\!\!\perp X_2$

'Markov equivalence' $\Rightarrow$ DAGs are not causal

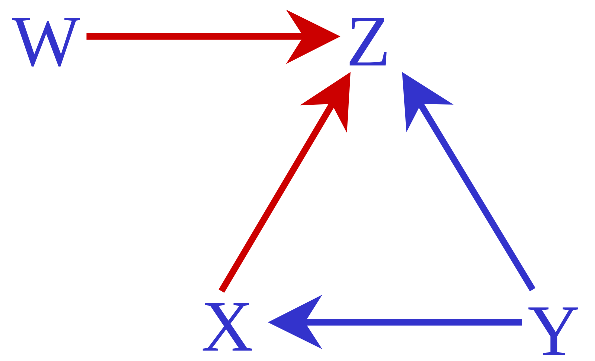
Markov equivalence

(Verma & Pearl, 1990)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (encode exactly the same conditional independencies) iff

(i) they have the same ‘skeleton’

(ii) they have the same ‘V–structures’



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

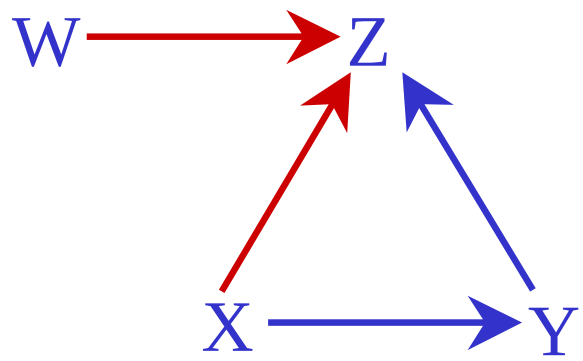
Markov equivalence

(Verma & Pearl, 1990)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (encode exactly the same conditional independencies) iff

(i) they have the same 'skeleton'

(ii) they have the same 'V–structures'



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

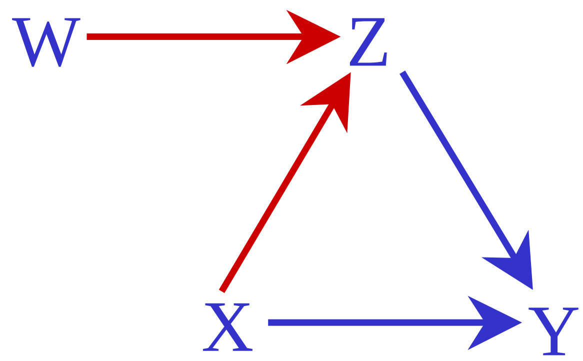
Markov equivalence

(Verma & Pearl, 1990)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (encode exactly the same conditional independencies) iff

(i) they have the same 'skeleton'

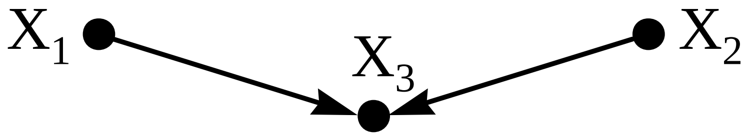
(ii) they have the same 'V–structures'



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

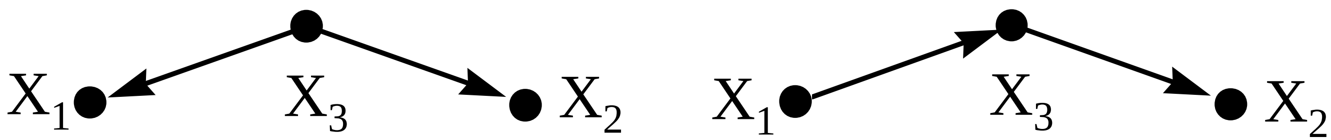
Important for the interpretation:

- conditioning on common child (selection) $\Rightarrow$ dependence



here: $X_1 \perp\!\!\!\perp X_2$, but $X_1 \not\perp\!\!\!\perp X_2 \mid X_3$

- marginalizing w.r.t. common parent (confounder) or intermediate variables $\Rightarrow$ dependence



here: $X_1 \perp\!\!\!\perp X_2 \mid X_3$, but $X_1 \not\perp\!\!\!\perp X_2$

$\Rightarrow$ how do we read off all conditional independencies?

Separation in DAGs

Construct moral graph $\mathcal{G}^m$

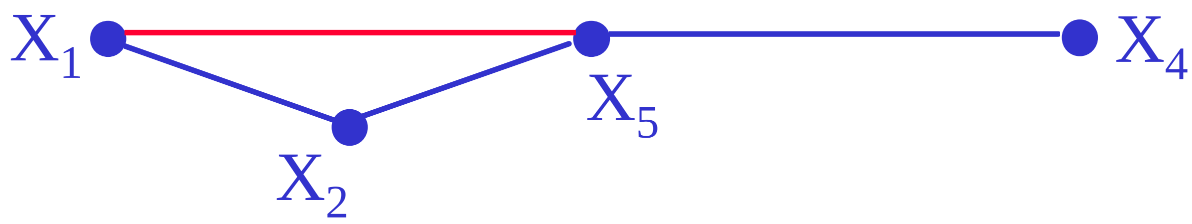
- connect parents of common children
- make all edges undirected

Theorem

$\mathcal{G}$ DAG, $P_{\mathbf{X}}$ $\mathcal{G}$–Markov, and $A, B, C \subset V$ disjoint:

C separates A and B in $\mathcal{G}^m_{\mathrm{An}(A \cup B \cup C)} \Rightarrow \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$

Example (continued): $\mathcal{G}^m_{\mathrm{An}(2,4,5)}$



$\Rightarrow X_2 \perp\!\!\!\perp X_4 \mid X_5$

Alternatively: d–Separation

(Pearl, 1988)

Given DAG $\mathcal{G} = (V, E)$. A path between X and $Y \in V$ is **blocked** by $S \subset V \backslash \{X, Y\}$ if

(i) it contains a constellation $\longleftarrow Z \longrightarrow$ or $\longleftarrow Z \longleftarrow$ and $Z \in S$ or

(ii) it contains a constellation $\longrightarrow Z \longleftarrow$ and neither Z nor any descendants of Z are elements of S.

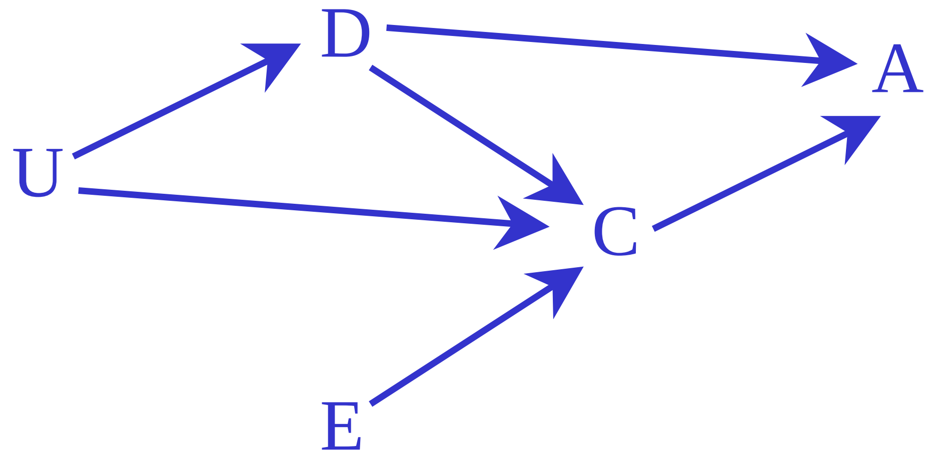
A and $B \subset V$ are d–separated by $S \subset V \backslash (A \cup B)$ if every path between A and B is blocked by S.

$\Leftrightarrow$ **separation using moralisation criterion**, i.e. $A \perp\!\!\!\perp B \mid S$

Example

(Robins, 2001)

D = endometrial cancer
E = oestrogens
C = vaginal bleeding
A = ascertained cancer
U = unknown uterine abnormality



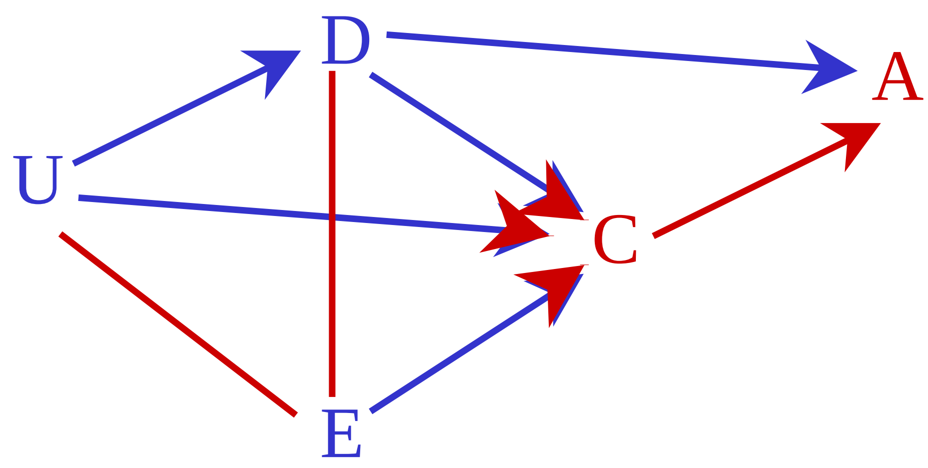
Graph implies: $E \perp\!\!\!\perp (D, U)$ and $A \perp\!\!\!\perp (U, E) \mid (D, C)$

- H_0: no 'causal effect' of E on D implies no association
- Case control study

Example

(Robins, 2001)

D = endometrial cancer
E = oestrogens
C = vaginal bleeding
A = ascertained cancer
U = unknown uterine abnormality



Graph implies: $E \perp\!\!\!\perp (D, U)$ and $A \perp\!\!\!\perp (U, E) \mid (D, C)$

- H_0: no 'causal effect' of E on D implies no association
- Case control study $\Rightarrow$ always conditional on A

$\Rightarrow$ H_0 cannot be tested in this design as $E \not\perp\!\!\!\perp D \mid A$ and $E \not\perp\!\!\!\perp D \mid (A, C)$

Causality: A Basic Distinction

seeing $\leftrightarrow$ doing

Association: seeing X helps to predict Y

$\Rightarrow$ usual conditional probability $P(Y \mid X = x)$

Causation: setting X helps to predict Y

$\Rightarrow$ new notation, e.g.

$P(Y \mid \text{do}(X = x))$ or $P(Y \mid\mid X = x)$ or

$P(Y \mid X, F_X = x)$ where F_X intervention indicator

Causal Graphs

Definition: (Lauritzen, 2000)

DAG $\mathcal{G}$, $P_{\mathbf{X}}$ $\mathcal{G}$–Markov. Then, $\mathcal{G}$ causal wrt $B \subset V$ if for any $A \subset B$

$$P(\mathbf{X} \mid \mathrm{do}(\mathbf{X}_A = a)) = \prod_{i \in V \setminus A} P(X_i \mid \mathbf{X}_{\mathrm{pa}(i)}) \Big|_{\mathbf{X}_A = a}$$

in words:

- all conditional specifications on $V \setminus A$ remain the same
- $\mathbf{X}_A$ are simply fixed to a when appearing in $\mathbf{X}_{\mathrm{pa}(i)}$

Example 1

- as DAG: $P(A, B) = P(B)P(A \mid B)$ — no restrictions
- if causal wrt B: $P(A \mid \text{do}(B = b)) = P(A \mid B = b)$
- if not causal (association due to other factor): $P(A \mid \text{do}(B = b)) = P(A)$

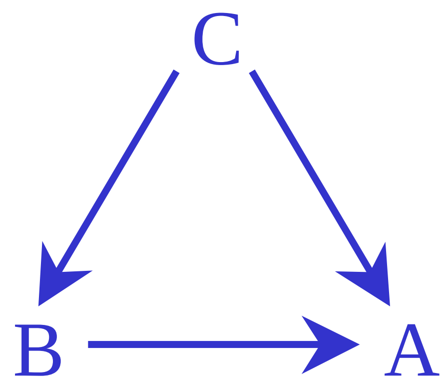
NOTE: in causal graph $\text{do}(B = b)$ is like usual conditioning if $\text{pa}(B) = \emptyset$.

Example 2

- as DAG: $P(A, B, C) = P(C)P(B \mid C)P(A \mid B, C)$
- if causal wrt B:

  $P(A, C \mid \text{do}(B = b)) = P(C)P(A \mid B = b, C)$

  $\rightarrow$ remove arrow into B and condition on B



Causal effect of B on A is

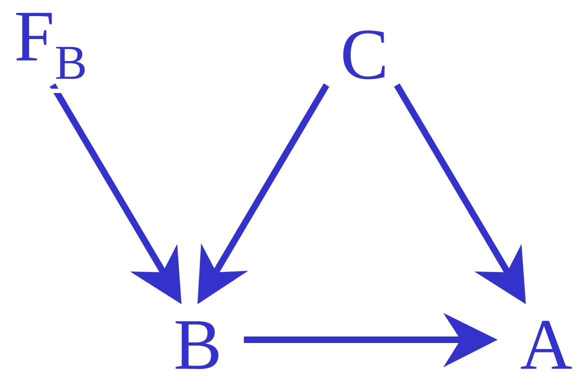
$$P(A \mid \text{do}(B = b)) = \sum_C P(C)P(A \mid B = b, C)$$

which is not the usual conditioning

$$P(A \mid B = b) = \frac{\sum_C P(C)P(B \mid C)P(A \mid B = b, C)}{\sum_C P(C)P(B \mid C)}$$

Alternative: Influence Diagrams

(Dawid, 2002)



Define **intervention indicator**

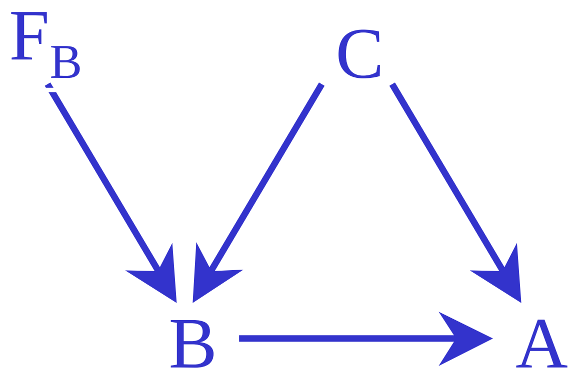
$$F_B = \begin{cases} \emptyset & = \text{idle} = \text{no intervention} \\ b & = \text{active} = \text{do}(B = b) \end{cases}$$

so that $P(B \mid F_B = \emptyset, C)$ = observational distribution

and $P(B \mid F_B = b, C) = I_{\{B=b\}}$ when intervening

- intervention is made explicit in the graph
- graph is again 'just' conditional independence graph
- can be generalised to random / conditional interventions
- NOTE: statements always conditional on F_B

Example:



- DAG implies

  $C \perp\!\!\!\perp F_B$ and $A \perp\!\!\!\perp F_B | (B, C)$ or

  $P(A, B, C \mid F_B) = P(C) P(B \mid F_B, C) P(A \mid B, C)$

- under intervention

  $P(A, C \mid F_B = b) = P(C) I_{\{B=b\}} P(A \mid B = b, C)$ $\rightarrow$ same as before

Causal effect

$$P(A \mid F_B = b) = \sum_C P(C) P(A \mid B = b, C)$$

Use of Causal Graphs

$\rightarrow$ simple and intuitive (?) 'language' for complex multivariate dependence structures

$\rightarrow$ important characteristics of the statistical model can be read off the graph, e.g. independencies under selection (case control studies) or latent structures (marginalisation)

Identifiability

- can we identify causal effect from observational data?
- what do we need to condition on?
- what must we not condition on?

$\rightarrow$ $P(Y \mid \text{do}(X = x))$ or $P(Y \mid F_X = \text{'active'})$ has to be expressed with observable quantities only.

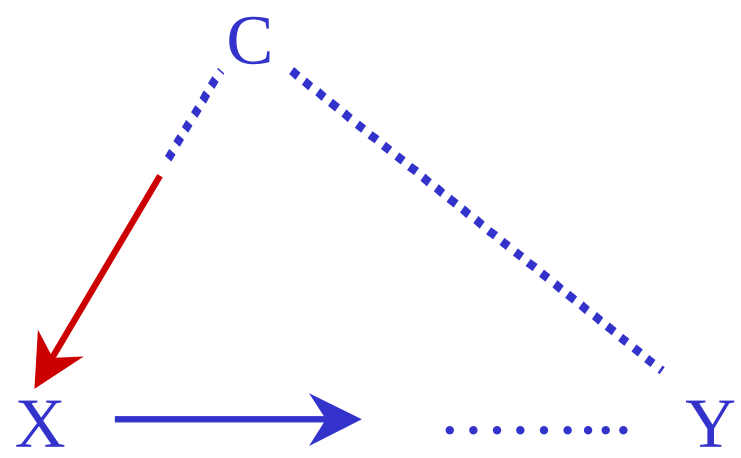
‘Back-Door’ Criterion with Causal Graphs

Theorem (Pearl, 1995)

Given causal graph $\mathcal{G}$ on V. Then $C \subset V \backslash \{X, Y\}$ identifies causal effect of X on Y if

(i) C is non-descendant of X and

(ii) all ‘back-door’ paths from X to Y are blocked by C



‘Back-Door’ Criterion with Intervention Graphs

Theorem (Lauritzen, 2000)

Given intervention DAG $\mathcal{G}$ on V with X, Y, $F_X \in V$. Then $C \subset V \setminus \{X, Y, F_X\}$ identifies causal effect of X on Y if

(i) $C \perp\!\!\!\perp F_X$ ($\Leftrightarrow$ C non-descendant of X)

(ii) $Y \perp\!\!\!\perp F_X \mid (X, C)$ ($\Leftrightarrow$ C blocks back-door paths)

Causal effect is then calculated as

$$P(Y \mid F_X = x) = \sum_C P(C) P(Y \mid X = x, C)$$

How can we use the Back-door Criterion?

- Construct the DAG based on knowledge of
  - subject matter (basic biology etc.)
  - temporal ordering
  - study design
  - statistical evidence

$\Rightarrow$ will typically include unobservable variables

- check for which choice of C (if any) (i) and (ii) hold
  $\rightarrow$ check for separations

Example: Drug Use During Pregnancy

(Robins, 2001)

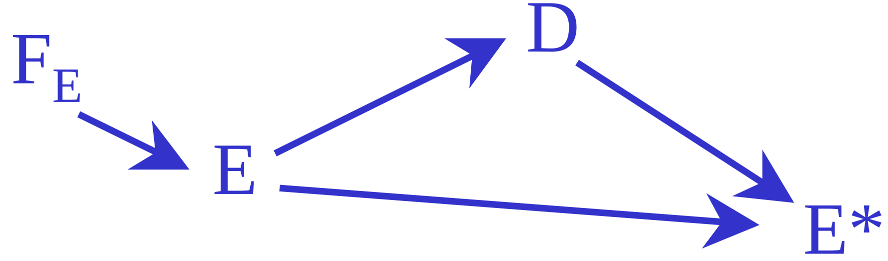
Exposure $=E=$ drug intake during pregnancy (from medical record)
Disease $=D=$ congenital defect of child
Surrogate $=E^*=$ self reported drug use (after diagnosis of D)
$\rightarrow$ case-control study

$D \perp\!\!\!\perp F_E \mid E$

$\Rightarrow$ causal effect of E on D
can be identified through OR_{ED}

$F_E \not\!\perp\!\!\!\perp E^*$

NOTE: should and must ignore E^*!



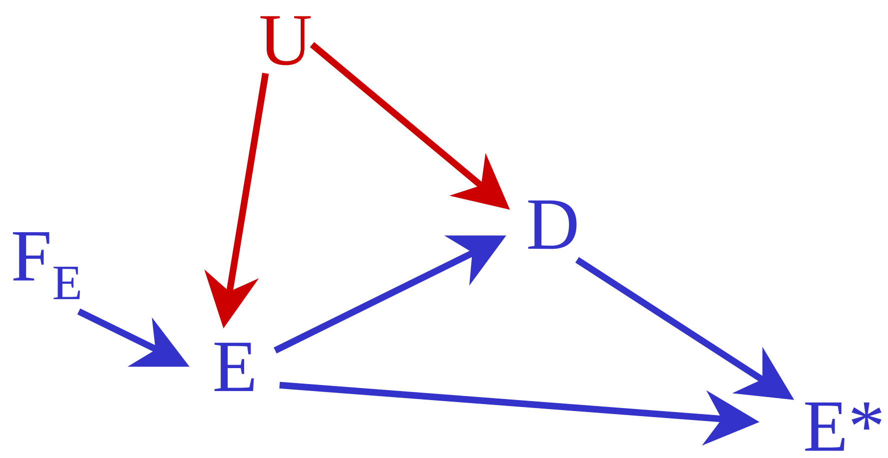
Example: Drug Use During Pregnancy

(Robins, 2001)

Exposure $= E =$ drug intake during pregnancy (from medical record)
Disease $= D =$ congenital defect of child
Surrogate $= E^* =$ self reported drug use (after diagnosis of D)
$\rightarrow$ case-control study

$D \not\perp\!\!\!\perp F_E \mid E$

$\Rightarrow$ causal effect of E on D
can not be identified



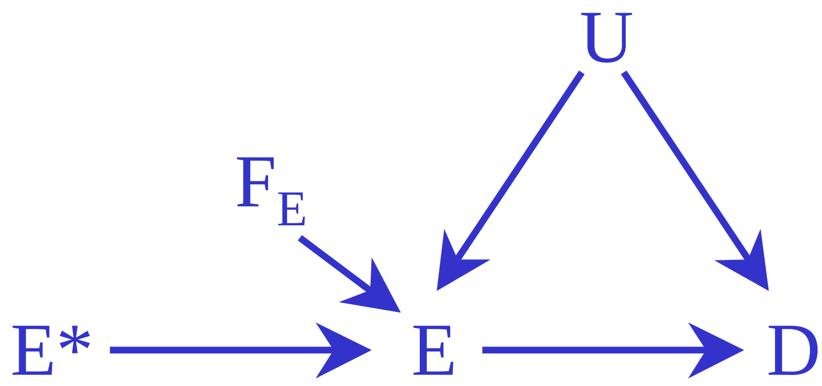
Example: Partial Compliance

E = actual drug intake

D = recovery

E^* = assigned drug intake

$\rightarrow$ randomized trial



randomisation $\Rightarrow$ $E^* \perp\!\!\!\perp U$

and further $D \perp\!\!\!\perp E^* \mid (E, U)$

but $D \not\!\perp\!\!\!\perp F_E \mid E$ and $D \not\!\perp\!\!\!\perp F_E \mid (E, E^*)$

$\Rightarrow$ causal effect can not be identified

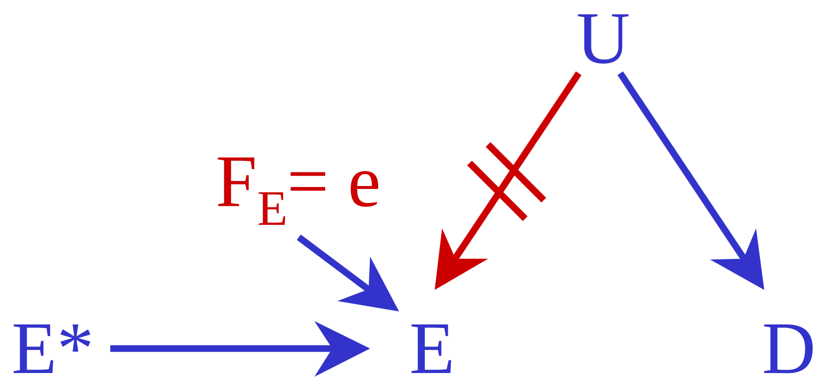
Example: Partial Compliance

E = actual drug intake

D = recovery

E^* = assigned drug intake

$\rightarrow$ randomized trial



H_0 : no causal effect of E on $D \Rightarrow E^* \perp\!\!\!\perp D$

$\Rightarrow H_0$ can be tested!

NOTE: can compute bounds for causal effect of E on D.

Example: Mineworkers Exposed to Radon

(Robins, 2001)

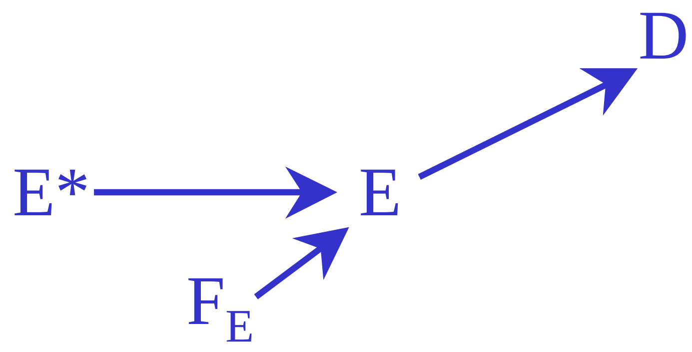
E = exposure to radon (lung dosimetry)

D = mortality (no censoring)

E^* = radon air level in mine

$\rightarrow$ prospective cohort study

data shows: E^* and D associated given E



Example: Mineworkers Exposed to Radon

(Robins, 2001)

E = exposure to radon (lung dosimetry)

D = mortality (no censoring)

E^* = radon air level in mine

$\rightarrow$ prospective cohort study

data shows: E^* and D associated given E

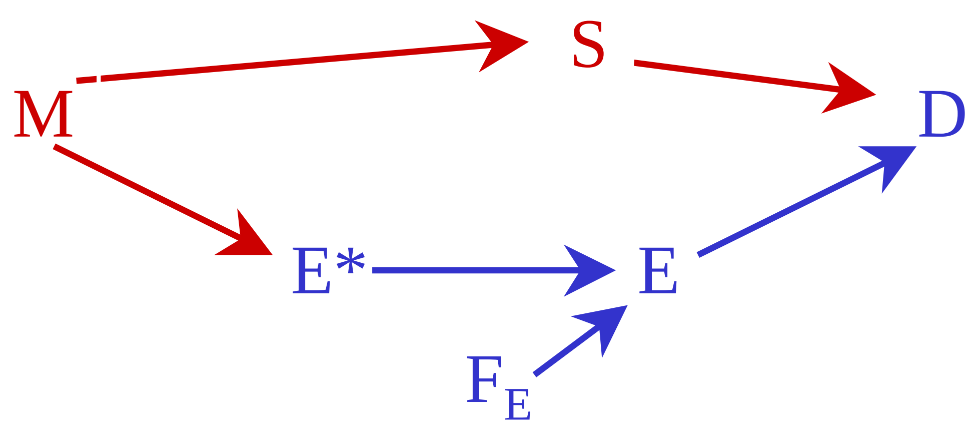
M = mine

S = silica

$D \not\perp\!\!\!\perp F_E \mid E$

but $D \perp\!\!\!\perp F_E \mid (E, E^*)$ and $E^* \perp\!\!\!\perp F_E$

$\Rightarrow$ E^* identifies the causal effect



Example: Mineworkers Exposed to Radon

(Robins, 2001)

E = exposure to radon (lung dosimetry)

D = mortality (no censoring)

E^* = radon air level in mine

$\rightarrow$ prospective cohort study

data shows: E^* and D associated given E

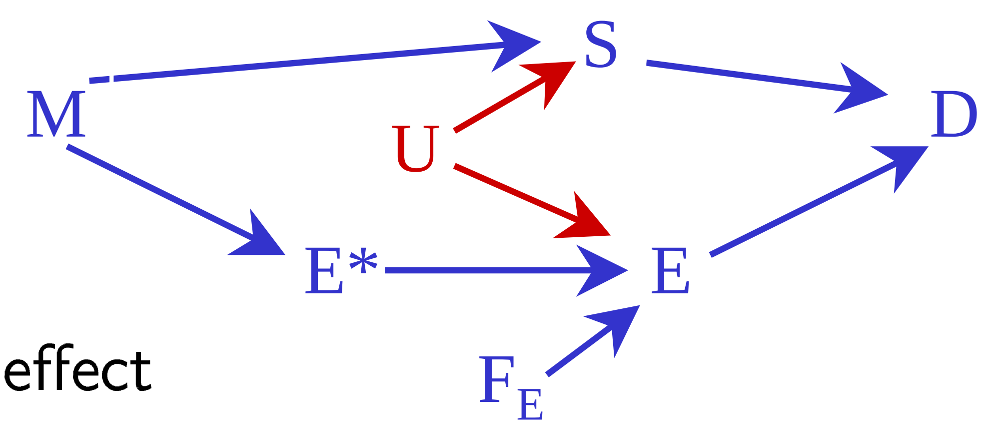
M = mine

S = silica

e.g. U = physical exertion

now $D \not\perp\!\!\!\perp F_E \mid (E, E^*)$

$\Rightarrow$ E^* does not identify the causal effect

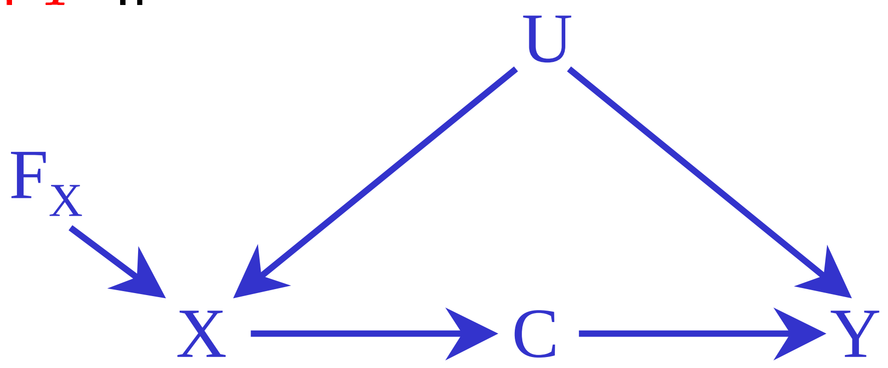


Front-door criterion

Theorem (Pearl, 1995)

Given intervention DAG $\mathcal{G}$ on V with X, Y, $F_X \in V$. Let $C \subset V \backslash \{X, Y, F_X\}$ and $U = V \backslash \{X, Y, F_X, C\}$.

Then C identifies causal effect of X on Y if



(i) $U \perp\!\!\!\perp F_X$

(ii) $C \perp\!\!\!\perp (U, F_X) \mid X$

(iii) $Y \perp\!\!\!\perp F_X \mid (C, U)$

Causal effect is then calculated as

$$P(Y \mid \mathrm{do}(X = x^*)) = \sum_C P(C \mid X = x^*) \sum_X P(Y \mid C, X) P(X)$$

Graphs and Counterfactuals

consider binary treatment $X \in \{0, 1\}$

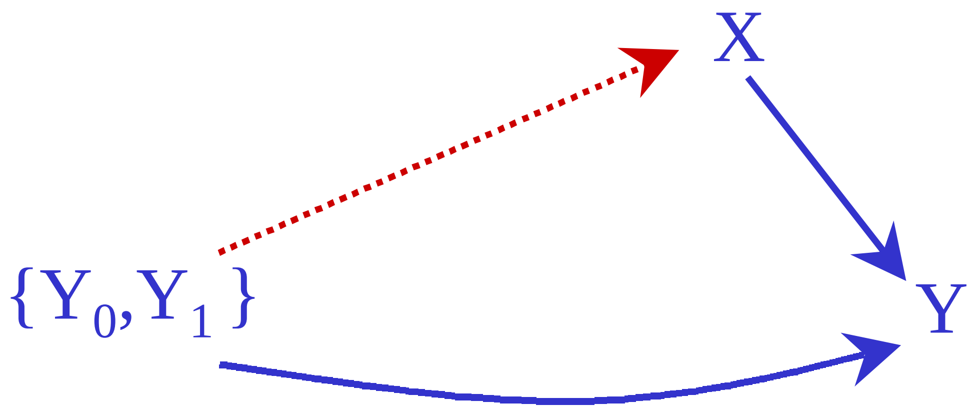
Y_0 = response if $X = 0$

Y_1 = response if $X = 1$ for same subject (at the same time)

$\Rightarrow$ can never be observed together

$\{Y_0, Y_1\}$ unobserved variable always affecting Y

$\Rightarrow$ X must not depend on $\{Y_0, Y_1\}$



Graphs and Counterfactuals

consider binary treatment $X = 0, 1$

Y_0 = response if $X = 0$

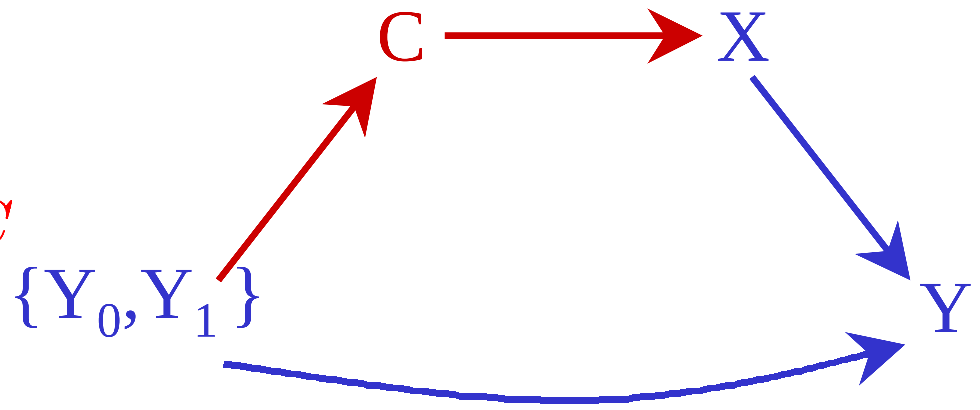
Y_1 = response if $X = 1$ for same subject (at the same time)

$\Rightarrow$ can never be observed together

$\{Y_0, Y_1\}$ unobserved variable always affecting Y

$\Rightarrow$ X must not depend on $\{Y_0, Y_1\}$

possibly conditional on covariates C



Counterfactuals, Graphs and Structural Equations

(Pearl, 2000)

X = treatment / exposure, Y = response, C = covariate

Structural equation model (SEM) — ingredients:

- DAG $\mathcal{G}$
- equations: $$X := f_X(\text{pa}(X), U_X)$$ $$Y := f_Y(\text{pa}(Y), U_Y)$$ $$C := f_C(\text{pa}(C), U_C)$$

  where f_X, f_Y, f_C describe 'stable' functional relations
- probability distribution on (U_X, U_Y, U_C)

$\Rightarrow$ induce probability distribution on (X, Y, C)
which is $\mathcal{G}$–Markov if U_X, U_Y, U_C independent

in this framework we have

$$Y_0 = f_Y(\mathrm{pa}(Y)\backslash X, X = 0, U_Y)$$
$$Y_1 = f_Y(\mathrm{pa}(Y)\backslash X, X = 1, U_Y)$$

with the same U_Y

$\Rightarrow$ distribution on (U_X, U_Y, U_C) also induces a probability distribution on (Y_0, Y_1, X, Y, C)

... in particular a joint distribution for (Y_0, Y_1)!

$\Rightarrow$ allows to address questions like

"Would Mr. Smith (who has cancer) have developed cancer had he not (counter to the fact) been exposed to radiation?"

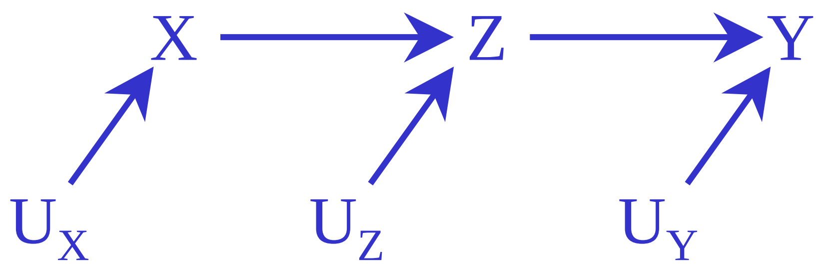
e.g. $P(Y_0 = 1 \mid X = 1, Y = 1)$

Twin Networks

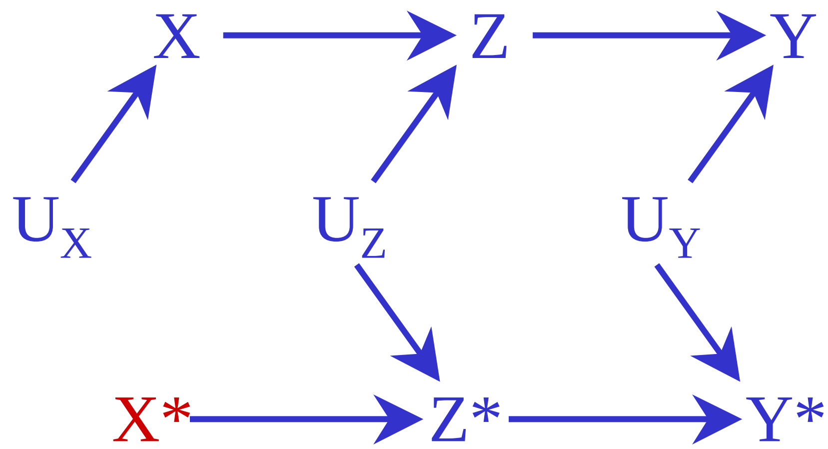
(Balke & Pearl, 1994)

Idea: include counterfactual variables in graph such that conditional independencies of joint distribution are represented.

Example



observational graph
according to SEM

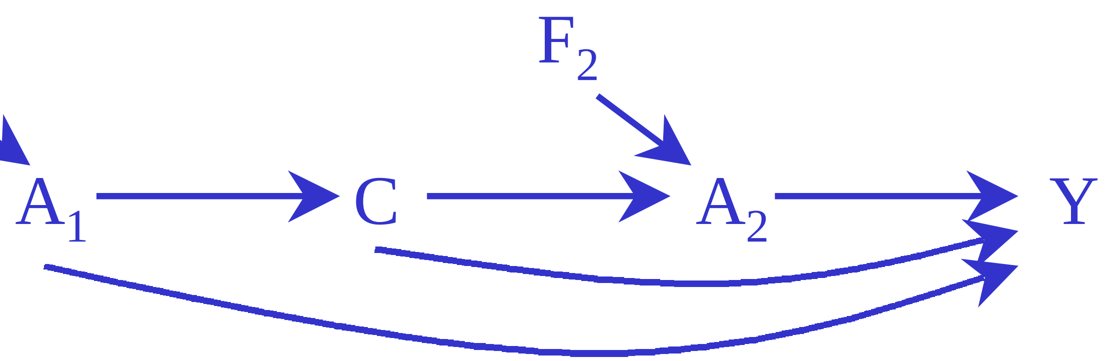


graph with counterfactuals
when setting X^*
read off, e.g. $Y^* \not\perp\!\!\!\perp X \mid Z$

Further Topics

- Sequential decisions / G-formula
- Instrumental variables
- Direct / indirect effects

Sequential Decisions / G-Formula



DAG implies

$Y \perp\!\!\!\perp F_1 \mid A_1$ and
$Y \perp\!\!\!\perp F_2 \mid (A_1, A_2, C)$ and $C \perp\!\!\!\perp F_1 \mid A_1$

or $P(Y, A_1, A_2, C|F_1, F_2) = P(A_1|F_1)P(C|A_1)P(A_2|F_2, A_1, C)P(Y|A_1, A_2, C)$

$\Rightarrow$ Causal effect **(G-Formula)** (Robins, 1986)

$$P(Y \mid F_1 = a, F_2 = b) = \sum_C P(C \mid A_1 = a)P(Y \mid C, A_1 = a, A_2 = b)$$

NOTE: $P(Y \mid F_1 = a, F_2 = b) \neq P(Y \mid A_1 = a, A_2 = b)$

Instrumental Variable

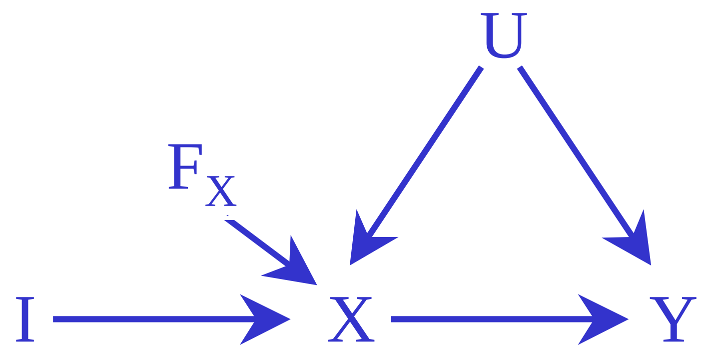
(Pearl, 1995)

Problem: causal effect of X on Y confounded by unobserved U.

⇒ an **instrumental variable** I satisfies:

(i) $I \perp\!\!\!\perp Y \mid (X, F_X = x)$ and

(ii) $I \not\!\perp\!\!\!\perp X$



⇒ allows to calculate bounds for the causal effect

Instrumental Variable

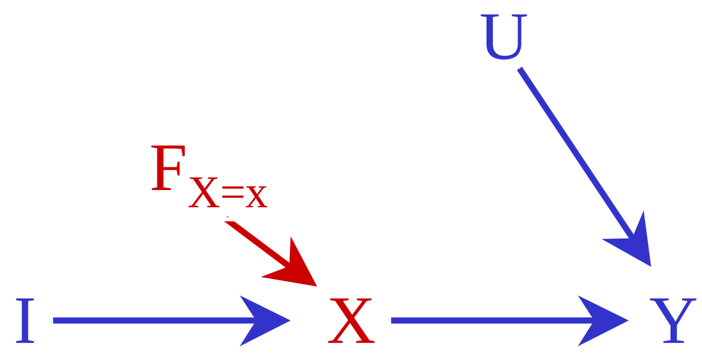
(Pearl, 1995)

Problem: causal effect of X on Y confounded by unobserved U.

$\Rightarrow$ an **instrumental variable** I satisfies:

(i) $I \perp\!\!\!\perp Y \mid (X, F_X = x)$ and

(ii) $I \not\perp\!\!\!\perp X$



$\Rightarrow$ allows to calculate bounds for the causal effect

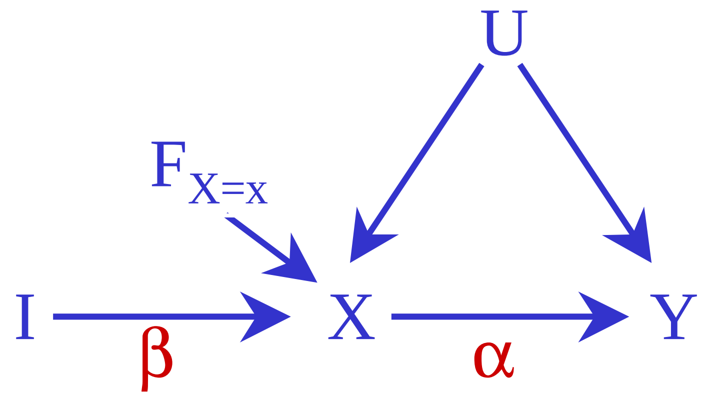
NOTE: (i) can not easily be tested as typically $I \not\perp\!\!\!\perp Y$ and $I \not\perp\!\!\!\perp Y \mid X$

Instrumental Variable

(Pearl, 1995)

Problem: causal effect of X on Y confounded by unobserved U.

$\Rightarrow$ an **instrumental variable** I satisfies:



(i) $I \perp\!\!\!\perp Y \mid (X, F_X = x)$ and

(ii) $I \not\perp\!\!\!\perp X$

$\Rightarrow$ allows to calculate the causal effect exactly in linear (multivariate normal) case:

$$\beta = r_{IX} \quad \text{and} \quad \alpha\beta = r_{IY} \quad \Rightarrow \quad \alpha = \frac{r_{IY}}{r_{IX}}$$

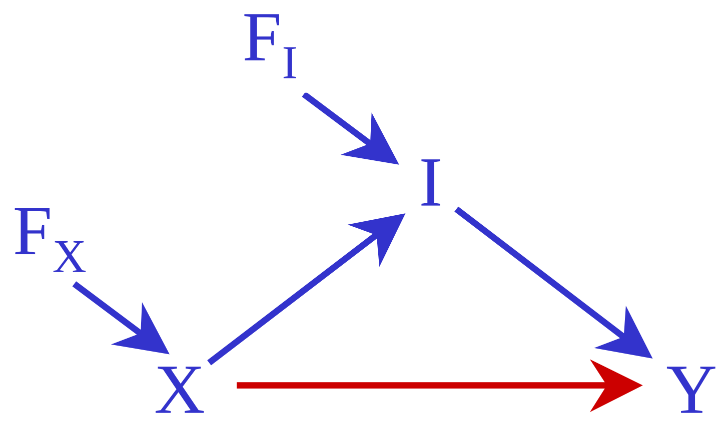
where r_{ab} LS regression coefficients

Direct Effect

X = treatment e.g. contraceptive pill

I = intermediate variable e.g. pregnancy

Y = response e.g. thrombosis

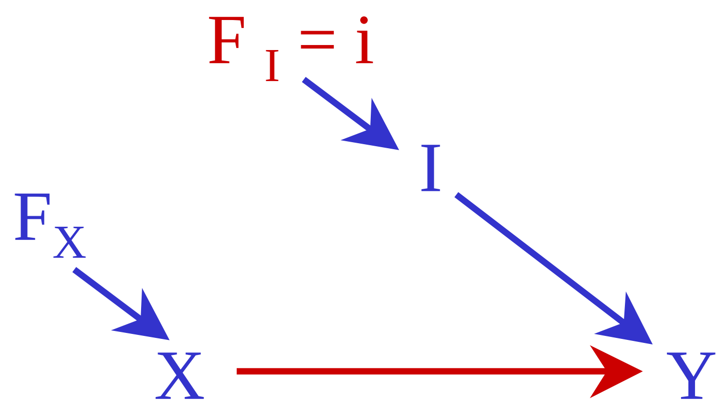


Direct Effect

X = treatment e.g. contraceptive pill

I = intermediate variable e.g. pregnancy

Y = response e.g. thrombosis



Direct effect

= effect of setting X while holding $I = i$ constant

$= P(Y | F_X = x, F_I = i)$ compared to $P(Y | F_X = x', F_I = i)$

(cf. 'controlled direct effect' in Pearl, 2001)

$\{F_X = x\}$ = taking the contraceptive pill

$\{F_X = x'\}$ = not taking the contraceptive pill

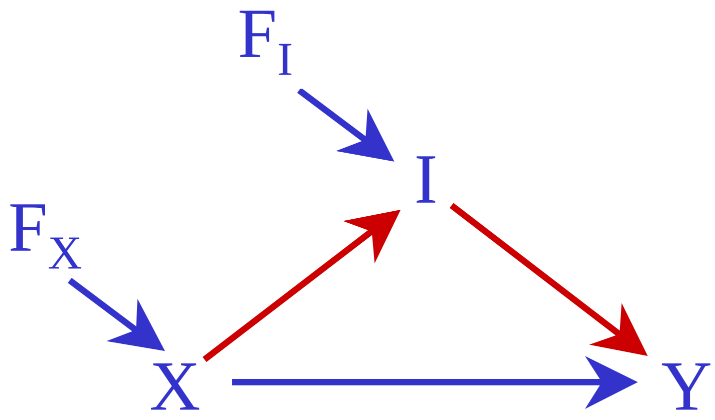
$\{F_I = i\}$ = preventing pregnancy by other method

Indirect Effect: The Placebo Effect

X = treatment, active agent or not

I = psychol. effect of receiving a pill or not

Y = response, e.g. recovery

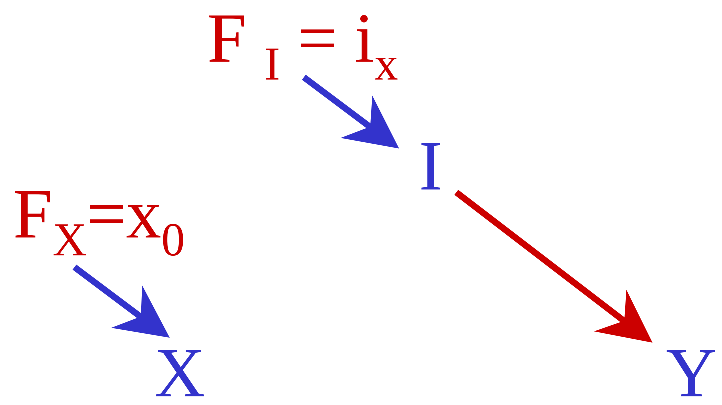


Indirect Effect: The Placebo Effect

X = treatment, active agent or not

I = psychol. effect of receiving a pill or not

Y = response, e.g. recovery



Indirect effect

= effect of setting I to values that arise for different X while $X = x_0$ is actually held constant such that x_0 has no direct effect

$= P(Y|F_X = x_0, F_I = i_x)$ compared to $P(Y|F_X = x_0, F_I = i'_x)$

$\{F_X = x_0\}$ = no active agent given

$\{F_I = i_x\}$ = no pill given

$\{F_I = i_{x'}\}$ = placebo given

Summary

Graphs help

- to make explicit the available / prior knowledge
- to make explicit the assumptions required for causal inference
- and to compare the two

Caution

- eliciting prior knowledge and formulating a graph is not easy
- all results provided are ‘non-parametric’
- back-door / front-door criteria are sufficient, not necessary
- DAGs difficult to combine with events in continuous time / feedback systems

References

Introductory / overview texts

Edwards D (2000): Causal inference. Chapter 8 in *Introduction to graphical modelling.* Springer, New York.

Lauritzen SL (2000): Causal inference from graphical models. In *Complex Stochastic Systems,* Eds. OE Bardorff-Nielsen, DR Cox, C Klüppelberg, pp. 63-107. CRC Press, London.

Pearl J (2000): *Causality – models, reasoning and inference.* Cambridge University Press.

Pearl J (2003): Statistics and causal inference: a review. *Test,* 12, 2, pp. 281-345.

Further references

Balke AA & Pearl J (1994): Counterfactual probabilities: computational methods, bounds and applications. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence,* pp. 46-54.

Dawid AP (2002): Influence diagrams for causal modelling and inference. *Int. Stat. Review,* 70, 2, pp. 161-189

Pearl J (1988): Probabilistic reasoning in intelligent systems. Morgan Kaufmann, CA.

Pearl J (1995): Causal diagrams for empirical research (with discussion). *Biometrika,* 82, 669-710.

Pearl J (1995): Causal inference from indirect experiments. *Artificial intelligence in Medicine,* 7, pp. 561-582.

Pearl J (2001): Direct and Indirect Effects. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence,* pp. 411-420.

Robins JM (1986): A new approach to causal inference in mortality studies with sustained exposure periods — application to control of the healthy worker effect. *Mathematical Modelling,* 7,pp. 1393-1512.

Robins JM (2001): Data, design and background knowledge in etiologic inference. *Epidemiology,* 12, 3, pp. 313-320.

Verma T & Pearl J (1990): Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence,* pp. 220-227.