Causality

Jonas Peters
University of Copenhagen

Academic year 2021/2022 – Block 4

UNIVERSITY OF COPENHAGEN

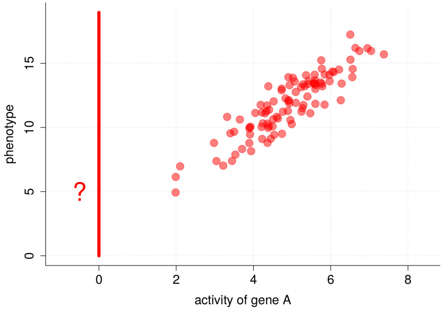
WELCOME!

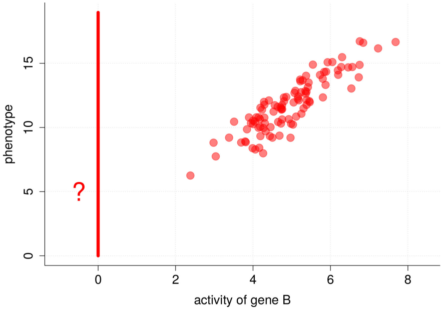
These slides are used only for visualization. They are not stand-alone material but should be considered as an addition to the reading material, in particular

- Book. Peters, Janzing, Schölkopf: Elements of Causal Inference, MIT Press (see also errata).
- Hand-written notes.
- Code examples.

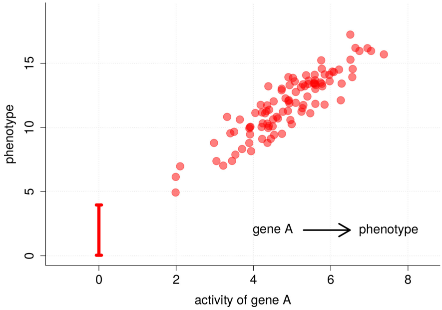
The slides contain many ideas and concepts that are developed by others and these are sometimes not cited properly. For references, please see the above mentioned book.

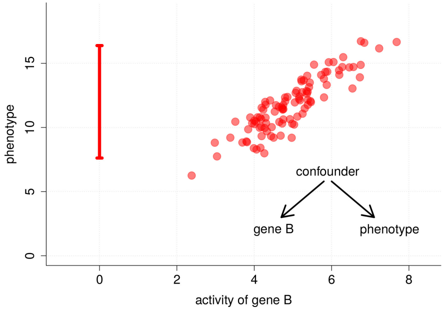
Consider the following problem.





Causality matters!





Example: smoking

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example, in the quarter of a century between 1922 and 1947 the annual number of deaths recorded increased from 612 to

whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

Possible Causes of the Increase

Two main causes have from time to time been put for-

Example: smoking

BRITISH MEDICAL JOURNAL

TABLE VII.—*Estimate of Total Amount of Tobacco Ever Consumed by Smokers; Lung-carcinoma Patients and Control Patients with Diseases Other Than Cancer*

Disease Group	No. Who have Smoked Altogether					Probability Test
	365 Cigs.–	50,000 Cigs.–	150,000 Cigs.–	250,000 Cigs.–	500,000 Cigs. +	
Males:						
Lung-carcinoma patients (647)	19 (2·9%)	145 (22·4%)	183 (28·3%)	225 (34·8%)	75 (11·6%)	$\chi^2 = 30{\cdot}60$; $n=4$; $P<0{\cdot}001$
Control patients with diseases other than cancer (622)..	36 (5·8%)	190 (30·5%)	182 (29·3%)	179 (28·9%)	35 (5·6%)	
Females:						
Lung-carcinoma patients (41)..	10 (24·4%)	19 (46·3%)	5 (12·2%)	7 (17·1%)	0 (0·0%)	$\chi^2 = 12{\cdot}97$; $n=2$; $0{\cdot}001 < P < 0{\cdot}01$ (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28) ..	19 (67·9%)	5 (17·9%)	3 (10·7%)	1 (3·6%)	0 (0·0%)	

UNG

uncil

Director of the Statistical

no one would deny that it
utory. As a corollary, it is
other causes.

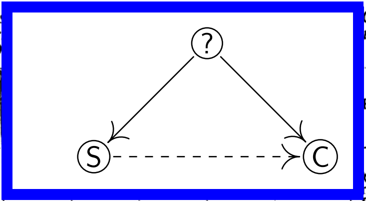
of the Increase

m time to time been put for-

Example: smoking

BRITISH MEDICAL JOURNAL

TABLE VII.—E… Consumed by Smoker… …tients with Diseases O…



Disease Group						Probability Test
Males:						
Lung-carcinoma patients (647)						$\chi^2=30{\cdot}60$; n=4; $P<0{\cdot}001$
Control patients with diseases other than cancer (622)..	36 (5·8%)	190 (30·5%)	182 (29·3%)	179 (28·9%)	35 (5·6%)	
Females:						
Lung-carcinoma patients (41)..	10 (24·4%)	19 (46·3%)	5 (12·2%)	7 (17·1%)	0 (0·0%)	$\chi^2=12{\cdot}97$; n=2; $0{\cdot}001< P < 0{\cdot}01$ (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28) ..	19 (67·9%)	5 (17·9%)	3 (10·7%)	1 (3·6%)	0 (0·0%)	

…UNG

…uncil

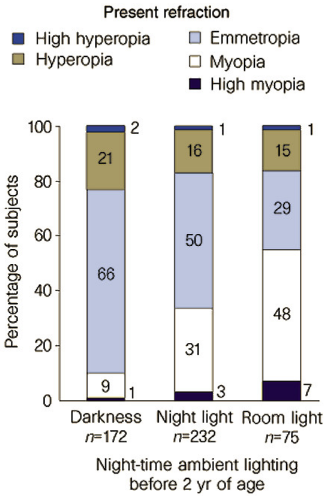
… Director of the Statistical

… no one would deny that it …utory. As a corollary, it is … other causes.

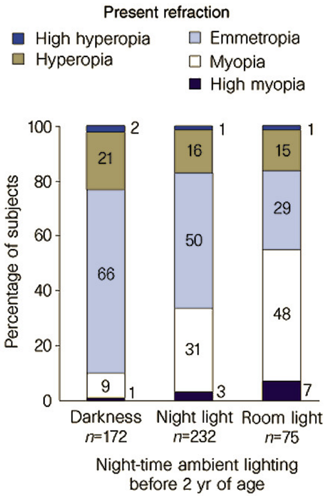
… of the Increase

…m time to time been put for-

Example: myopia



Example: myopia



"the strength of the association . . . does suggest that the absence of a daily period of darkness during childhood is a potential precipitating factor in the development of myopia"

Quinn, Shin, Maguire, Stone: *Myopia and ambient lighting at night*, Nature 1999

Example: myopia

Patente

Night light with sleep timer

US 20050007889 A1

ZUSAMMENFASSUNG

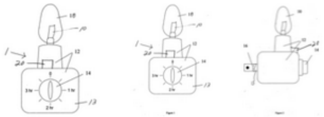
A timer a light and an optional music source is located on or in a housing of a nightlight assembly. When this assembly is plugged into a source of electric power, the timer is set to a selected time for the light and optional music to remain on. After this selected time has elapsed, the light and music automatically turns off, allowing for sleep in appropriate darkness and silence.

Veröffentlichungsnummer	US20050007889 A
Publikationstyp	Anmeldung
Anmeldenummer	US 10/614,245
Veröffentlichungsdatum	13. Jan. 2005
Eingetragen	8. Juli 2003
Prioritätsdatum ⓘ	8. Juli 2003
Erfinder	Karin Peterson
Ursprünglich Bevollmächtigter	Peterson Karin Lyn
Zitat exportieren	BiBTeX, EndNote, R

Klassifizierungen (4)

Externe Links: USPTO, USPTO-Zuordnung, Esp

BILDER (3)



BESCHREIBUNG

ANSPRÜCHE (18)

Example: myopia

Patente

Night light with sleep timer

US 20050007889 A1

ZUSAMMENFASSUNG

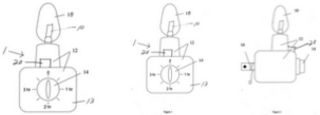
A timer a light and an optional music source is located on or in a housing of a nightlight assembly. When this assembly is plugged into a source of electric power, the timer is set to a selected time for the light and optional music to remain on. After this selected time has elapsed, the light and music automatically turns off, allowing for sleep in appropriate darkness and silence.

Veröffentlichungsnummer	US20050007889 A
Publikationstyp	Anmeldung
Anmeldenummer	US 10/614,245
Veröffentlichungsdatum	13. Jan. 2005
Eingetragen	8. Juli 2003
Prioritätsdatum	8. Juli 2003
Erfinder	Karin Peterson
Ursprünglich Bevollmächtigter	Peterson Karin Lyn
Zitat exportieren	BiBTeX, EndNote, R

Klassifizierungen (4)

Externe Links: USPTO, USPTO-Zuordnung, Esp

BILDER (3)



Question: Does the night light with sleep timer help?

BESCHREIBUNG

ANSPRÜCHE (18)

Example: kidney stones

	Treatment A	Treatment B
	$\frac{273}{350} = 0.78$	$\frac{289}{350} = 0.83$
	$\frac{562}{700} = 0.80$	

Assume: treatment is chosen only based on size of stones.

Charig et al.: *Comparison of treatment of renal calculi by open surgery, (...)* , British Medical Journal, 1986

Example: kidney stones

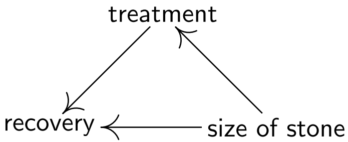
	Treatment A	Treatment B
Small Stones ($\frac{357}{700} = 0.51$)	$\frac{81}{87} = 0.93$	$\frac{234}{270} = 0.87$
Large Stones ($\frac{343}{700} = 0.49$)	$\frac{192}{263} = 0.73$	$\frac{55}{80} = 0.69$
	$\frac{273}{350} = 0.78$	$\frac{289}{350} = 0.83$
	$\frac{562}{700} = 0.80$	

Assume: treatment is chosen only based on size of stones.

Charig et al.: *Comparison of treatment of renal calculi by open surgery, (...)* , British Medical Journal, 1986

Example: kidney stones

underlying ground truth:



Example: kidney stones

underlying ground truth:

$$\begin{tikzcd} & \text{treatment} \arrow[dl] & \\ \text{recovery} & & \text{size of stone} \arrow[ul] \arrow[ll] \end{tikzcd}$$

Question: What is the expected recovery if all get treatment B?

(Make treatment independent of size.)

- Classical statistics:
  statistical model:

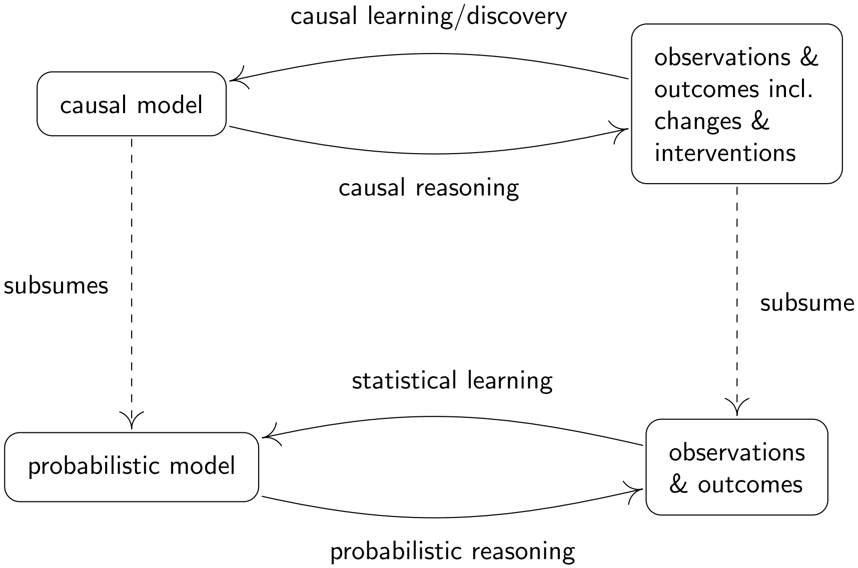
- Classical statistics:
  statistical model: $\{P_\theta, \theta \in \Theta\}$

- Classical statistics:
  statistical model: $\{P_\theta, \theta \in \Theta\}$
  observed data: from P_{θ_0}

- Classical statistics:
  statistical model: $\{P_\theta, \theta \in \Theta\}$
  observed data: from P_{θ_0}
  inference: investigate θ_0
  prediction: use parts of P_{θ_0}

- Classical statistics:
  statistical model: $\{P_\theta, \theta \in \Theta\}$
  observed data: from P_{θ_0}
  inference: investigate θ_0
  prediction: use parts of P_{θ_0}
- Causality is often about

- Classical statistics:
  statistical model: $\{P_\theta, \theta \in \Theta\}$
  observed data: from P_{θ_0}
  inference: investigate θ_0
  prediction: use parts of P_{θ_0}
- Causality is often about asking questions about distributions different from the one we have data from.
- We need models relating these distributions.
- We need tools to do causal inference.


causal learning/discovery
causal model
observations & outcomes incl. changes & interventions
causal reasoning
subsumes
subsume
statistical learning
probabilistic model
observations & outcomes
probabilistic reasoning

- Questions: lectures, TA sessions and padlet (better than emails/absalon messages)
- Format: lecture (1–4), inverted lecture (5–8), ?
- TA: Sorawit (James) Saengkyongam

- Questions: lectures, TA sessions and padlet (better than emails/absalon messages)
- Format: lecture (1–4), inverted lecture (5–8), ?
- TA: Sorawit (James) Saengkyongam

Exam Students:

- 5 mandatory assignments (up to two people)
- 4 assignments need to be passed (choose two exercises, hand-in in time)
- oral exam (21.6./22.6./also: 24.6.)

- Questions: lectures, TA sessions and padlet (better than emails/absalon messages)
- Format: lecture (1–4), inverted lecture (5–8), ?
- TA: Sorawit (James) Saengkyongam

Exam Students:

- 5 mandatory assignments (up to two people)
- 4 assignments need to be passed (choose two exercises, hand-in in time)
- oral exam (21.6./22.6./also: 24.6.)

Exam PhD Students:

- No hand-in of assignments.
- Report at the end (22.6.) about own research problem (alternatively: paper or data study).
- Cannot contain recycled material.

All information on absalon.

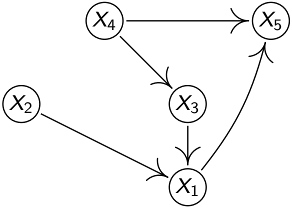
Maybe, there will be a bit of flexibility at the end. What are you interested in?

Hand-written notes 1.

Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

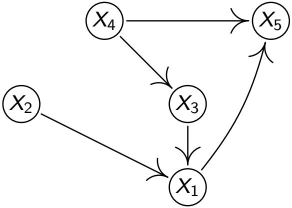
Check, whether all paths blocked!!



Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

Check, whether all paths blocked!!



$\circ \cdots \rightarrow \circ \rightarrow \cdots \circ$ blocks a path.

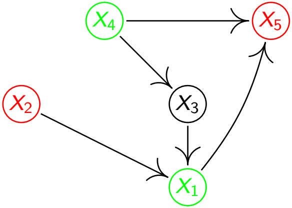
$\circ \cdots \leftarrow \circ \rightarrow \cdots \circ$ blocks a path.

$\circ \cdots \rightarrow \circ \leftarrow \cdots \circ$ blocks a path.

Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

Check, whether all paths blocked!!



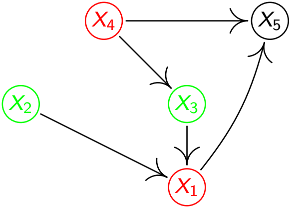
$\circ \cdots \rightarrow \circ \rightarrow \cdots \circ$ blocks a path.
$\circ \cdots \leftarrow \circ \rightarrow \cdots \circ$ blocks a path.
$\circ \cdots \rightarrow \circ \leftarrow \cdots \circ$ blocks a path.

X_2 and X_5 are d-sep. by $\{X_1, X_4\}$

Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

Check, whether all paths blocked!!



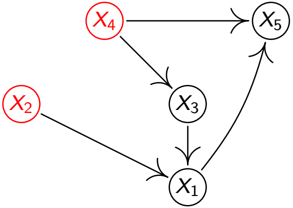
$\circ \cdots \rightarrow \circ \rightarrow \cdots \circ$ blocks a path.
$\circ \cdots \leftarrow \circ \rightarrow \cdots \circ$ blocks a path.
$\circ \cdots \rightarrow \circ \leftarrow \cdots \circ$ blocks a path.

X_2 and X_5 are d-sep. by $\{X_1, X_4\}$
X_4 and X_1 are d-sep. by $\{X_2, X_3\}$

Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

Check, whether all paths blocked!!



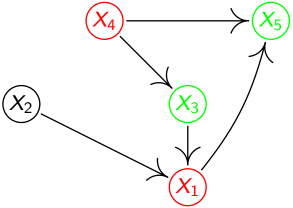
$\circ \cdots \rightarrow \circ \rightarrow \cdots \circ$	blocks a path.
$\circ \cdots \leftarrow \circ \rightarrow \cdots \circ$	blocks a path.
$\circ \cdots \rightarrow \circ \leftarrow \cdots \circ$	blocks a path.

X_2 and X_5 are d-sep. by $\{X_1, X_4\}$
X_4 and X_1 are d-sep. by $\{X_2, X_3\}$
X_2 and X_4 are d-sep. by $\{\}$

Definition: d-separation

X_i and X_j are d-separated by $\mathcal{S}$ if all paths between X_i and X_j are blocked by $\mathcal{S}$.

Check, whether all paths blocked!!



$\circ \cdots \rightarrow \circ \rightarrow \cdots \circ$ blocks a path.

$\circ \cdots \leftarrow \circ \rightarrow \cdots \circ$ blocks a path.

$\circ \cdots \rightarrow \circ \leftarrow \cdots \circ$ blocks a path.

X_2 and X_5 are d-sep. by $\{X_1, X_4\}$
X_4 and X_1 are d-sep. by $\{X_2, X_3\}$
X_2 and X_4 are d-sep. by $\{\}$
X_4 and X_1 are NOT d-sep. by $\{X_3, X_5\}$

Hand-written notes 2.

p	number of DAGs with p nodes
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	1213442454842881
10	4175098976430598143
11	31603459396418917607425
12	521939651343829405020504063
13	18676600744432035186664816926721
14	1439428141044398334941790719839535103
15	237725265553410354992180218286376719253505
16	83756670773733320287699303047996412235223138303
17	62707921196923889899446452602494921906963551482675201
18	99421195322159515895228914592354524516555026878588305014783
19	332771901227107591736177573311261125883583076258421902583546773505
20	2344880451051088988152559855229099188899081192234291298795803236068491263
21	34698768283588750028759328430181088222313944540438601719027559113446586077675521
22	1075822921725761493652956179327624326573727662809185218104090000500559527511693495107583
23	69743329837281492647141549700245804876504274990515985894109106401549811985510951501377122074625

`https://oeis.org/A003024/b003024.txt`

Definition

Given a directed graph (V, E), a permutation

$$\pi : V \to V \text{ (bijective)}$$

is called a causal/topological order if

$$j \in DE_i \Rightarrow \pi(i) < \pi(j).$$

(Remember: $\pi^{-1}(1)$ does not have any parents.)

Definition

Given a directed graph (V, E), a permutation

$$\pi : V \to V \text{ (bijective)}$$

is called a causal/topological order if

$$j \in DE_i \Rightarrow \pi(i) < \pi(j).$$

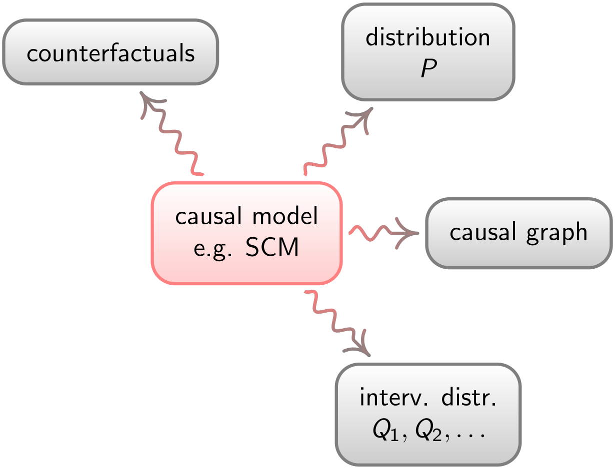
(Remember: $\pi^{-1}(1)$ does not have any parents.)

Proposition (Prop. B.2)

For any DAG, there exists a causal order.

Hand-written notes 3.

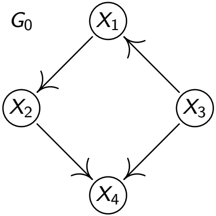
Recap:


counterfactuals
distribution
P
causal model
e.g. SCM
causal graph
interv. distr.
$Q_1, Q_2, \ldots$

SCMs $(\mathbf{S}, P^{\mathbf{N}})$: structural equations with noise distribution.

$$
\begin{aligned}
X_1 &:= f_1(X_3, N_1) \\
X_2 &:= f_2(X_1, N_2) \\
X_3 &:= f_3(N_3) \\
X_4 &:= f_4(X_2, X_3, N_4)
\end{aligned}
$$

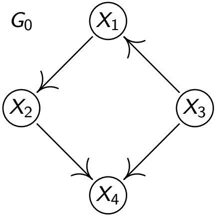
- N_i jointly independent
- G_0 has no cycles



SCMs $(\mathbf{S}, P^{\mathbf{N}})$ model observational distributions over $X_1, \ldots, X_d$. Call it P.

$$
\begin{aligned}
X_1 &:= f_1(X_3, N_1)\\
X_2 &:= f_2(X_1, N_2)\\
X_3 &:= f_3(N_3)\\
X_4 &:= f_4(X_2, X_3, N_4)
\end{aligned}
$$

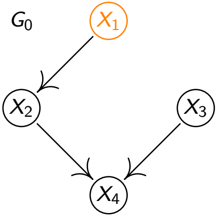
- N_i jointly independent
- G_0 has no cycles



SCMs $(\mathbf{S}, P^{\mathbf{N}})$ model interventions, too. Call it: $P_{do(X_1:=0)}$.

$$
\begin{aligned}
X_1 &:= 0 \\
X_2 &:= f_2(X_1, N_2) \\
X_3 &:= f_3(N_3) \\
X_4 &:= f_4(X_2, X_3, N_4)
\end{aligned}
$$

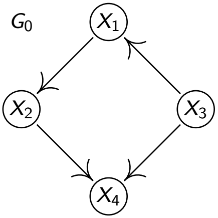
- N_i jointly independent
- G_0 has no cycles



SCMs model observational distributions over $X_1, \ldots, X_d$. Call it: P.

$$
\begin{aligned}
X_1 &:= f_1(X_3, N_1) \\
X_2 &:= f_2(X_1, N_2) \\
X_3 &:= f_3(N_3) \\
X_4 &:= f_4(X_2, X_3, N_4)
\end{aligned}
$$

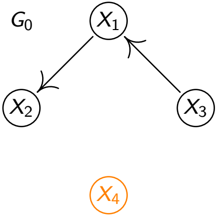
- N_i jointly independent
- G_0 has no cycles



SCMs model interventions, too. Call it $P_{do(X_4:=13)} \neq P(.\,|\,X_4 = 13)$

$$
\begin{aligned}
X_1 &:= f_1(X_3, N_1)\\
X_2 &:= f_2(X_1, N_2)\\
X_3 &:= f_3(N_3)\\
X_4 &:= 13
\end{aligned}
$$

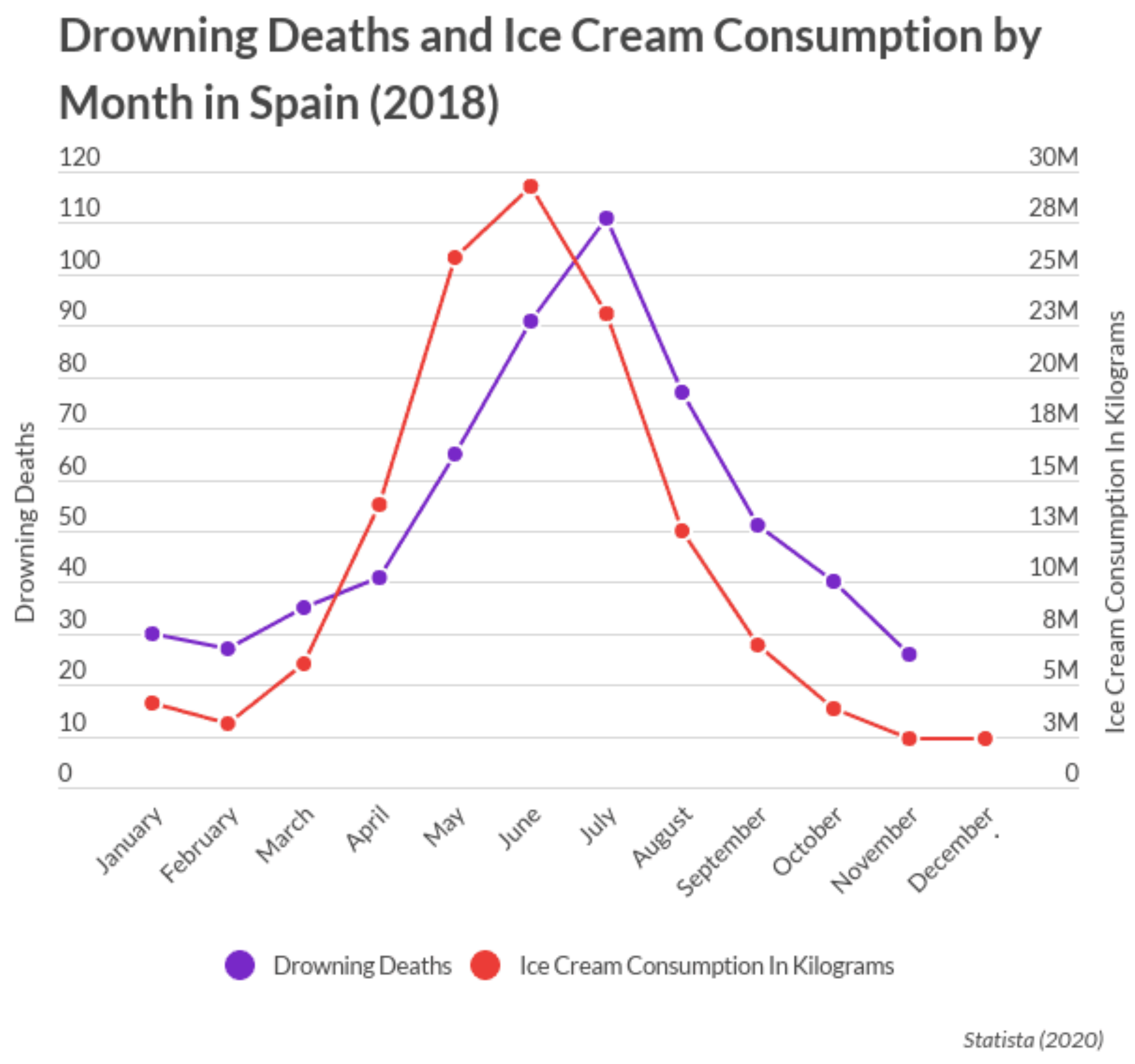
- N_i jointly independent
- G_0 has no cycles



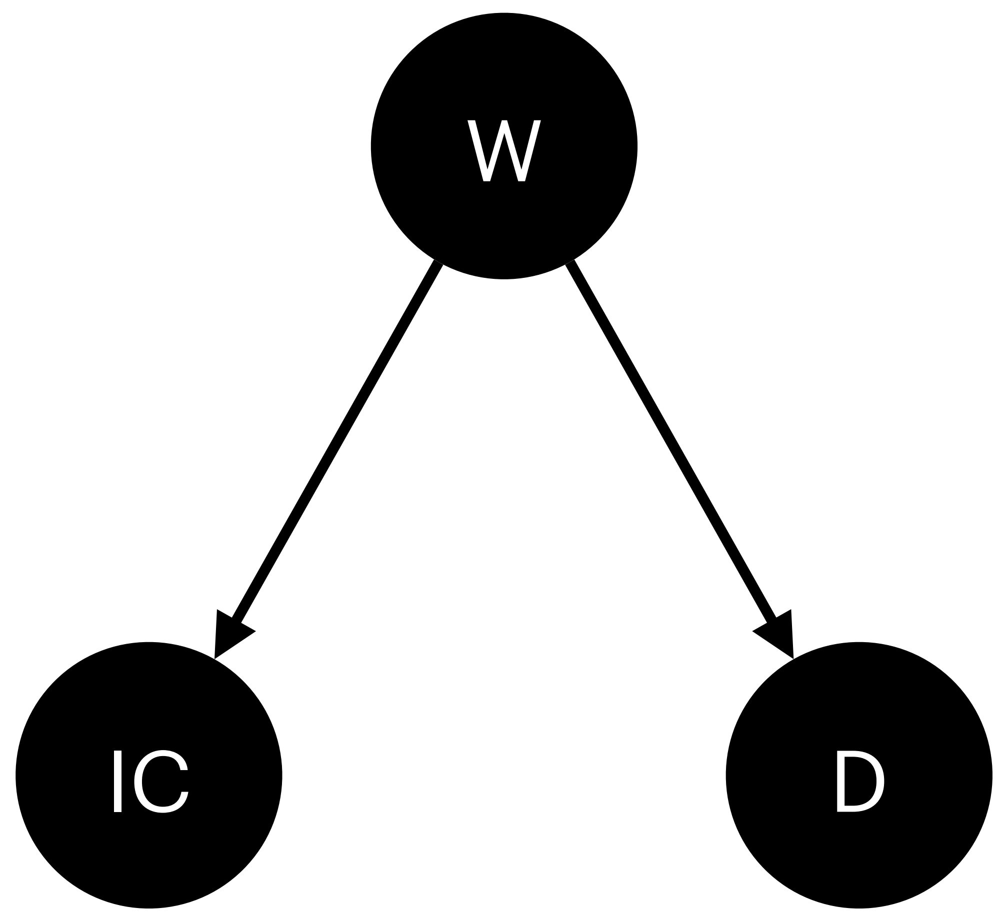
(Exercise-SCM.R)

Total Causal Effect

Total Causal Effect



Is there a *total causal effect* from IC to D ?



$$\mathbb{P}_D^{do(IC:=99999)} = \mathbb{P}_D^{do(IC:=0)} = \mathbb{P}_D$$

There is "no effect" from IC to D.
(Note, however, that we usually have that IC is dependence with D).

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

(i) $\exists \tilde{N}_X$ such that X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

(i) $\exists \tilde{N}_X$ such that X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

(ii) $\exists x_1, x_2$ such that $\mathbb{P}_{Y}^{\mathbb{C};do(X:=x_1)} \neq \mathbb{P}_{Y}^{\mathbb{C};do(X:=x_2)}$

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

(i) $\exists \tilde{N}_X$ such that X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

(ii) $\exists x_1, x_2$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x_1)} \neq \mathbb{P}_Y^{\mathbb{C};do(X:=x_2)}$

(iii) $\exists x$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x)} \neq \mathbb{P}_Y^{\mathbb{C}}$

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

(i) $\exists \tilde{N}_X$ such that X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

(ii) $\exists x_1, x_2$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x_1)} \neq \mathbb{P}_Y^{\mathbb{C};do(X:=x_2)}$

(iii) $\exists x$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x)} \neq \mathbb{P}_Y^{\mathbb{C}}$

(iv) $\forall \tilde{N}_X$ with full support we have X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

Total Causal Effect

Definition & Proposition (total causal effect)

Consider an SCM $\mathbb{C}$, we say that there is a (total) causal effect from X to Y if one of the following **equivalent** statements (i), (ii), (iv) hold

(i) $\exists \tilde{N}_X$ such that X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

(ii) $\exists x_1, x_2$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x_1)} \neq \mathbb{P}_Y^{\mathbb{C};do(X:=x_2)}$

(iii) $\exists x$ such that $\mathbb{P}_Y^{\mathbb{C};do(X:=x)} \neq \mathbb{P}_Y^{\mathbb{C}}$

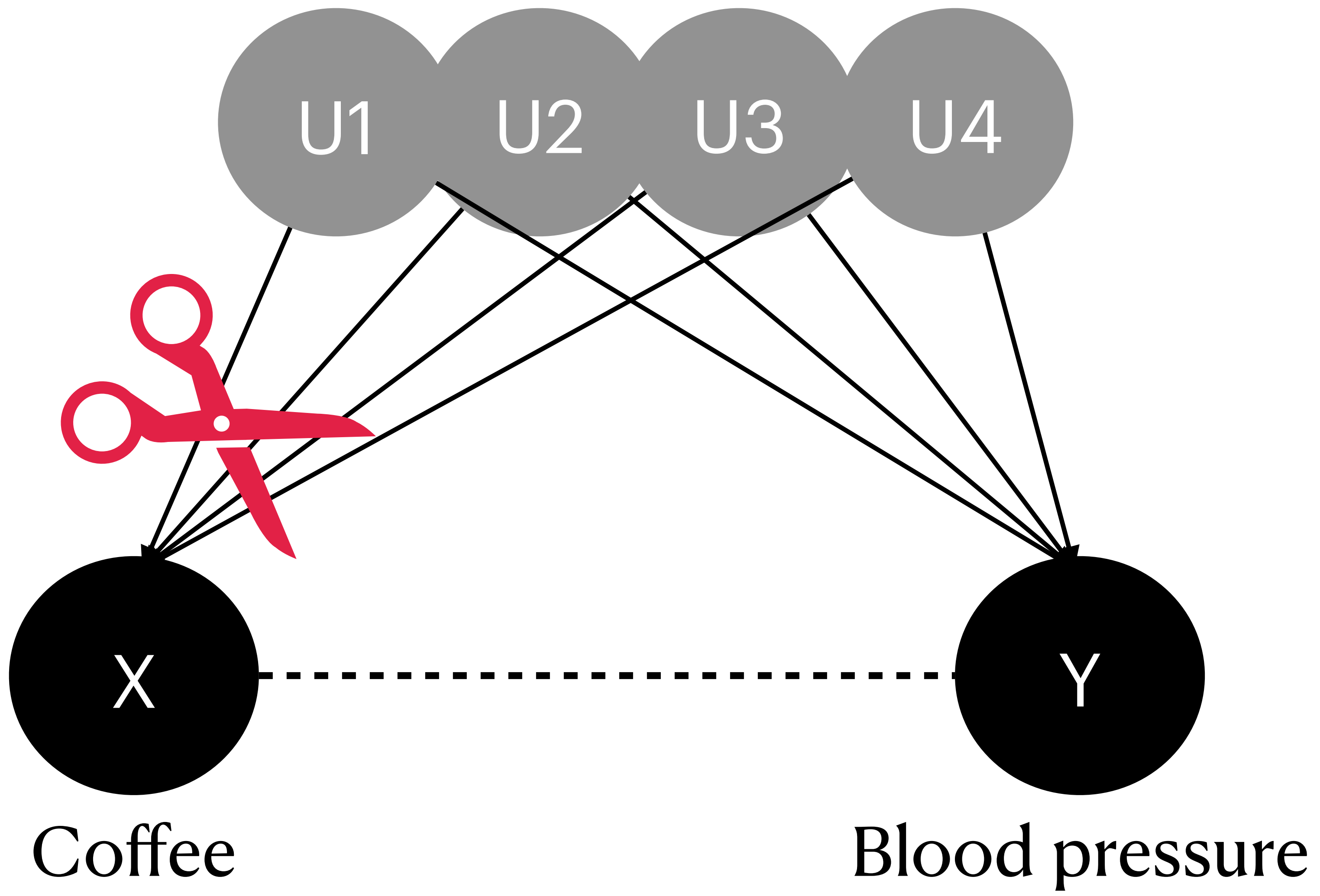
(iv) $\forall \tilde{N}_X$ with full support we have X not ind. Y in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

Each of (i), (ii), (iv) implies (iii).

See Proposition 6.13

Total Causal Effect

Is there a *total causal effect* from Coffee to Blood pressure?



Randomised trails!

Raomdize X with $\tilde{N}_X$. This yields $\mathbb{P}_{X,Y}^{do(X:=\tilde{N}_X)}$. Then test whether $X \perp\!\!\!\perp Y$.

Total Causal Effect

Proposition: Assume an SCM $\mathbb{C}$ with corresponding graph G.

(i) $\nexists$ directed path from X to $Y \implies \nexists$ causal effect from X to Y.

(ii) Sometimes there is a directed path but no total causal effect.

Proof (i): Let $\tilde{N}_X$ be ind. of all noise variables.

$\nexists$ directed path from X to $Y \implies X \notin AN_Y \cup \{Y\}$

$\nexists$ directed path from X to $Y \implies X \perp\!\!\!\perp Y$ in $\mathbb{P}_{X,Y}^{\mathbb{C};do(X:=\tilde{N}_X)}$

Because $X = \tilde{N}_X$ and $\exists \tilde{f}: Y = \tilde{f}(N_{AN_Y}, N_Y)$

Total Causal Effect

Proposition: Assume an SCM $\mathbb{C}$ with corresponding graph G.

(i) $\nexists$ directed path from X to Y $\implies$ $\nexists$ causal effect from X to Y.

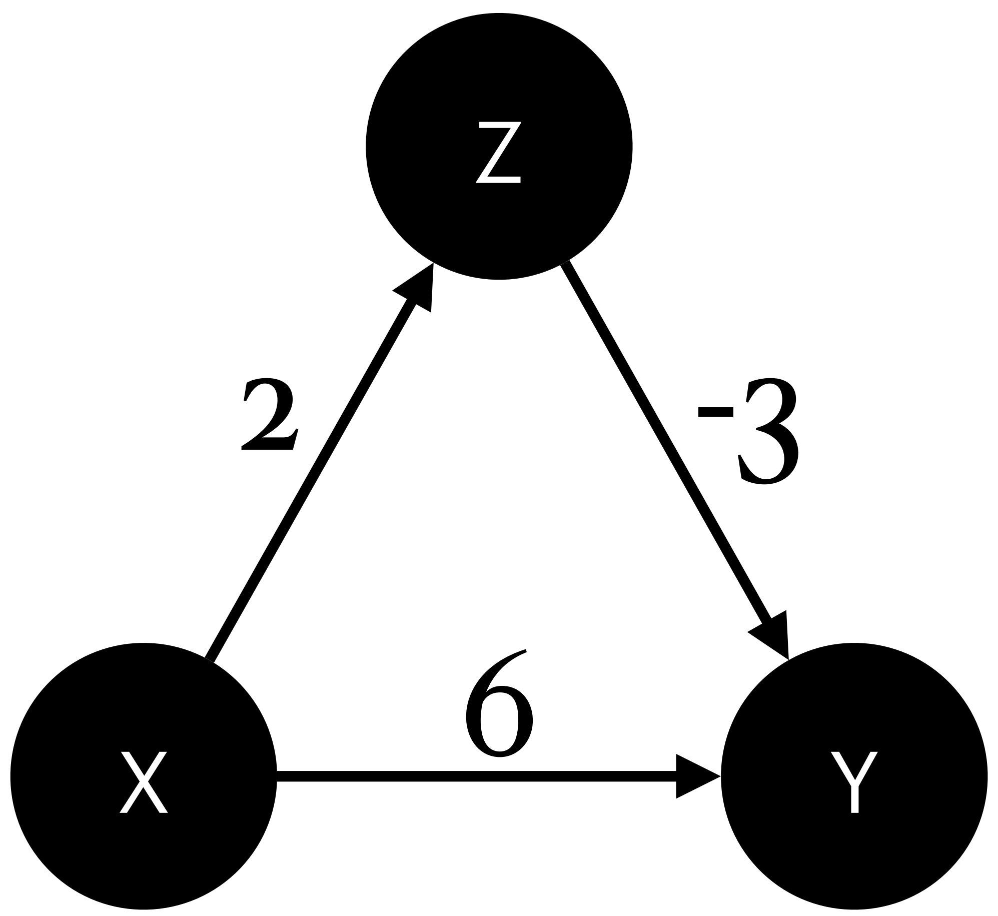
(ii) Sometimes there is a directed path but no total causal effect.

Proof (ii):

$X := N_X$

$Z := 2X + N_Z$

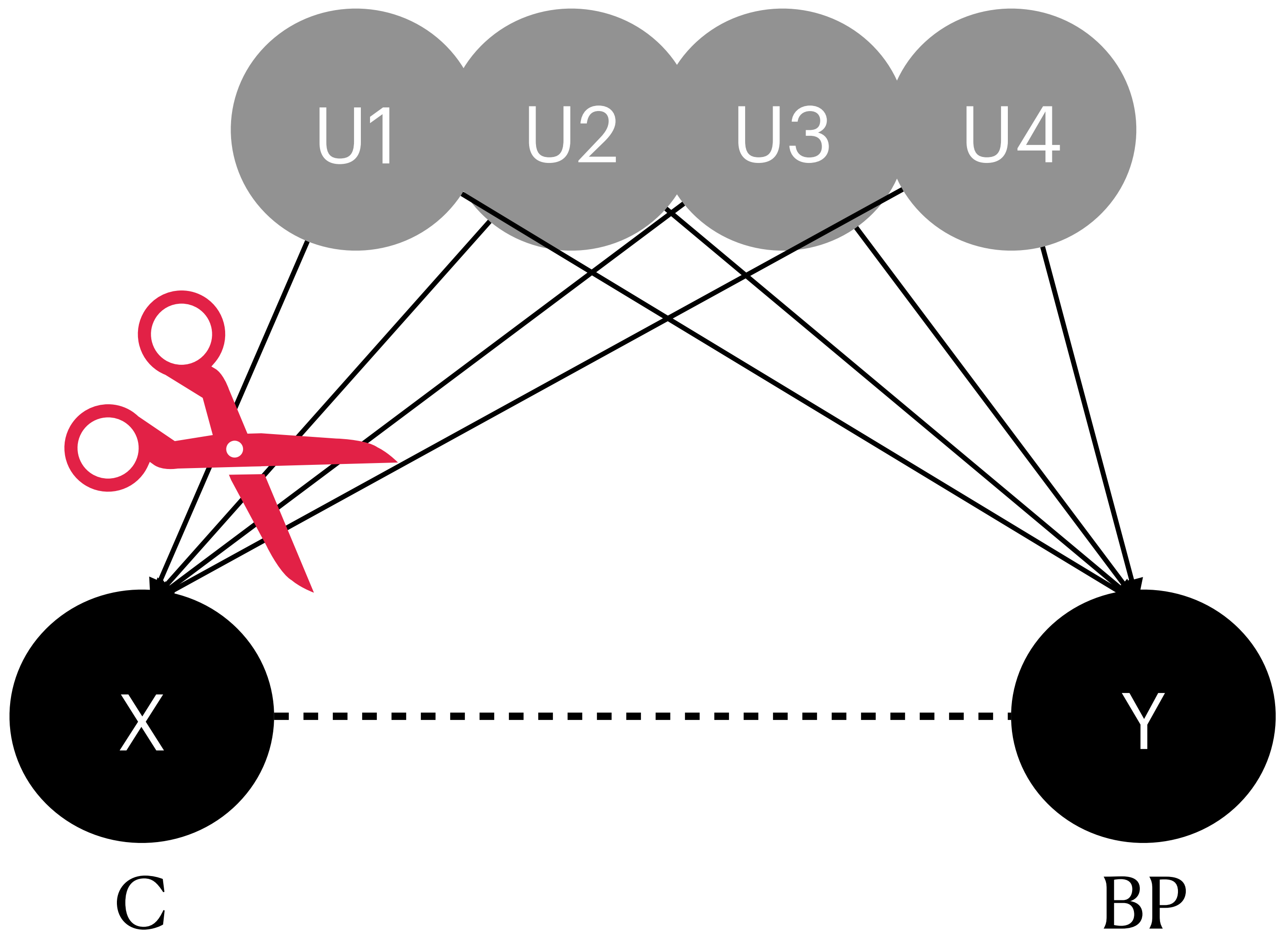
$Y := -3Z + 6X + N_Y$



Then: $Y = -6X - 3N_Z + 6X + N_Y = -3N_Z + N_Y$

Total Causal Effect

Is there a directed path from Coffee to Blood pressure?

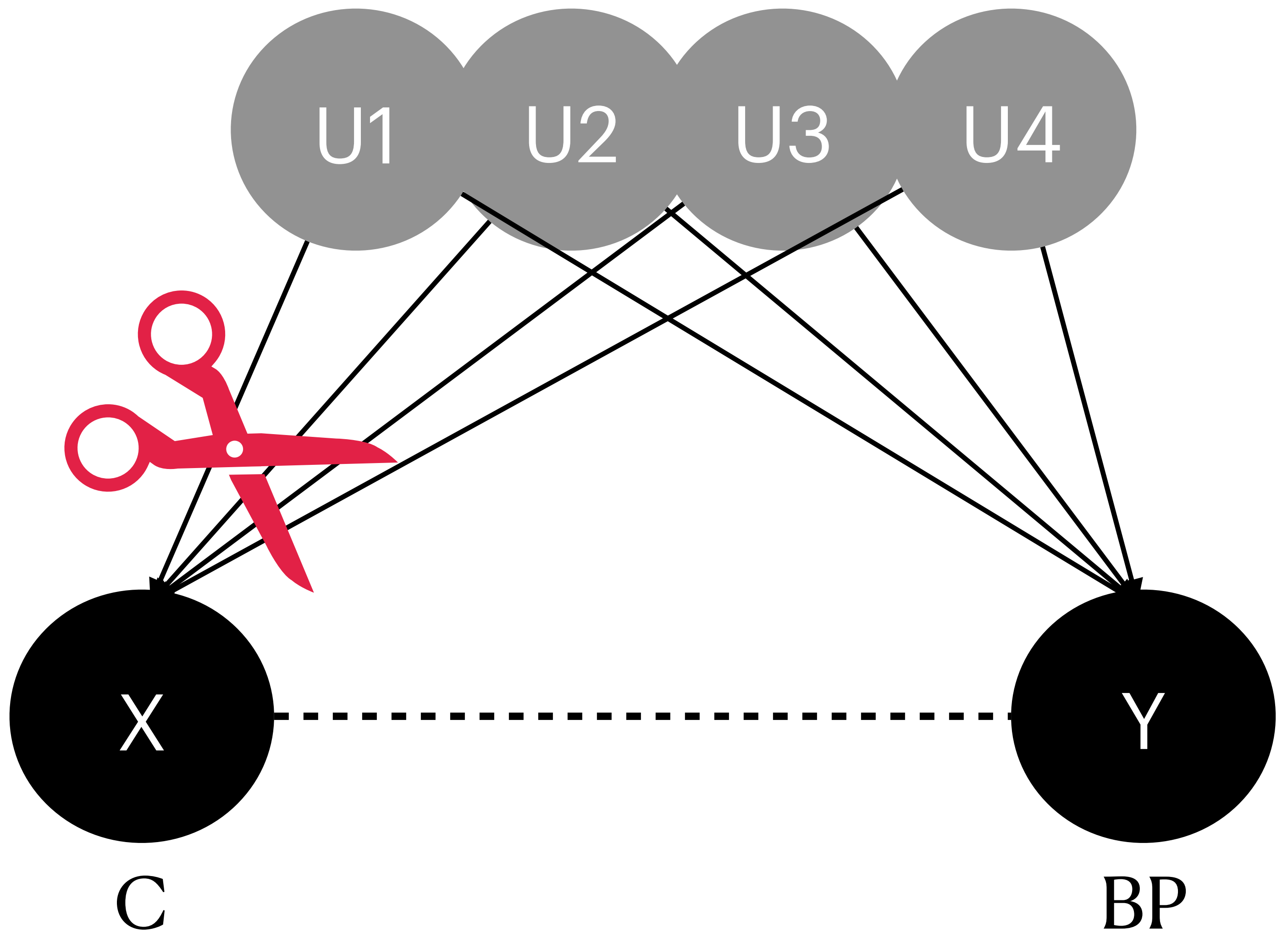


Randomised trails!

In $\mathbb{C}$; $do(C := \tilde{N}_C)$ the incoming edges into C are all removed.

Total Causal Effect

Is there a directed path from Coffee to Blood pressure?



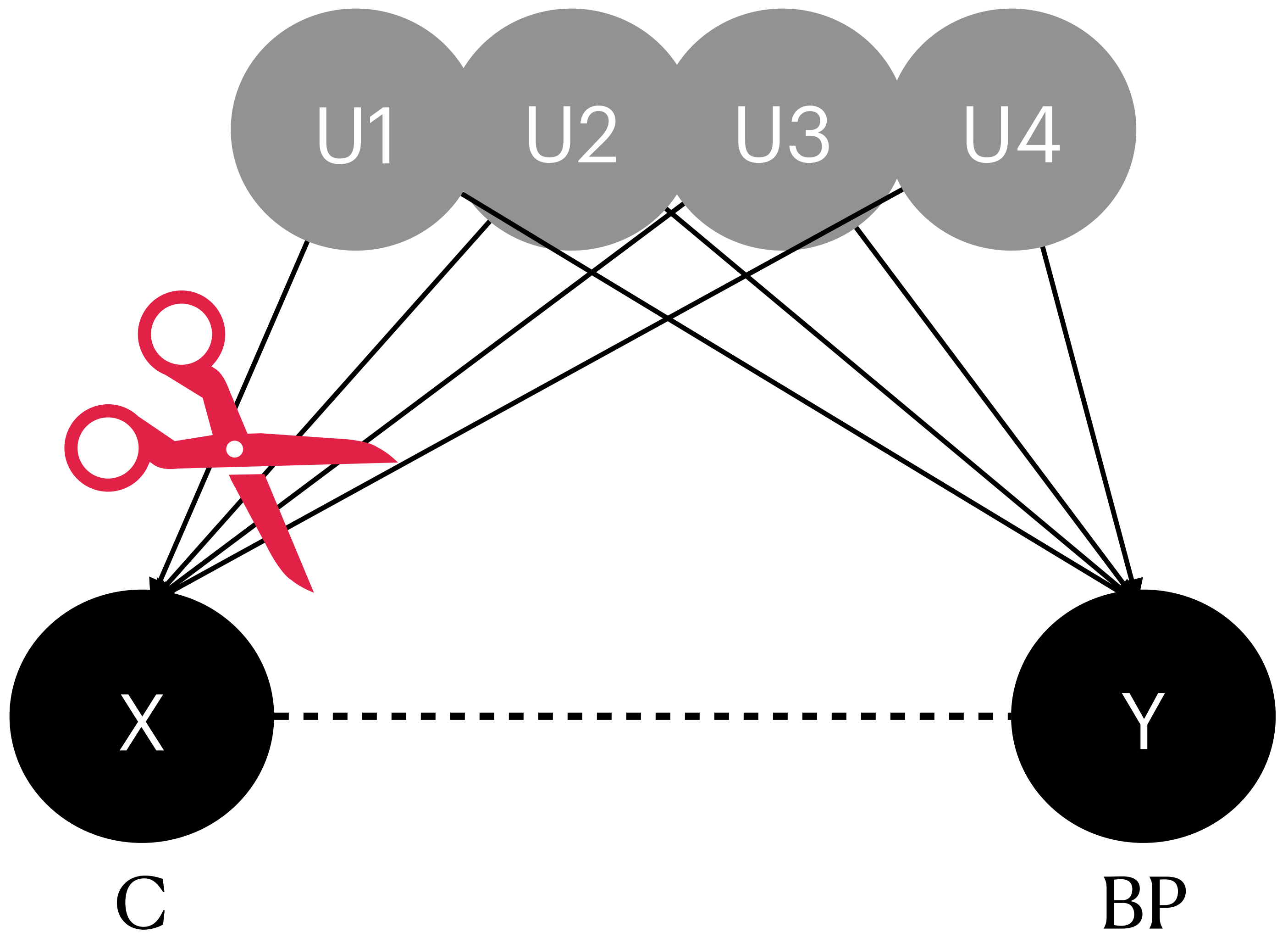
Randomised trails!

In $\mathbb{C}; do(C := \tilde{N}_C)$ the incoming edges into C are all removed.

Thus, if C not ind. BP in $\mathbb{P}^{\mathbb{C};do(C:=\tilde{N}_C)}$, there is a total causal effect from C to BP.

Total Causal Effect

Is there a directed path from Coffee to Blood pressure?



Randomised trails!

In $\mathbb{C}; do(C := \tilde{N}_C)$ the incoming edges into C are all removed.

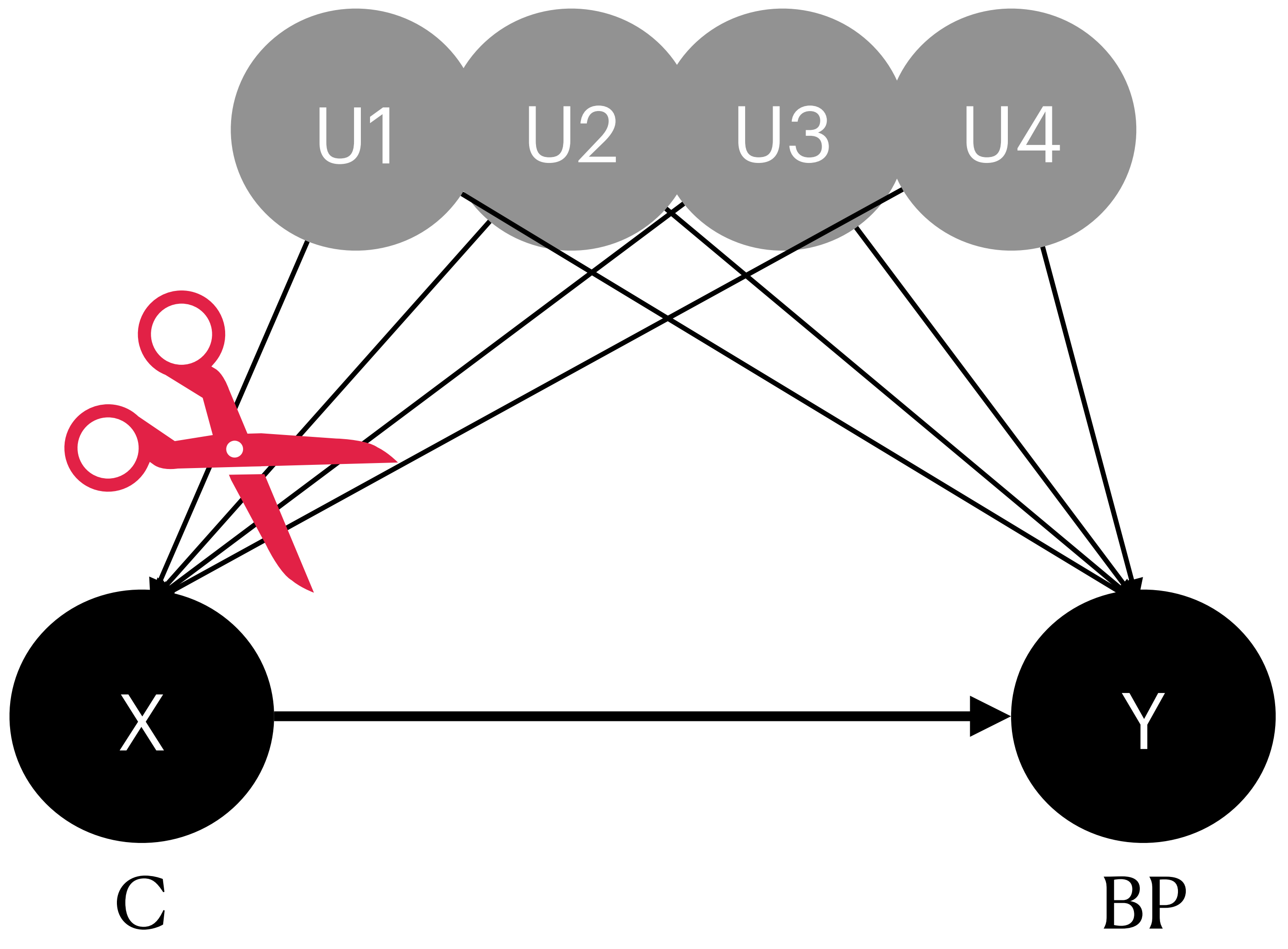
Thus, if C not ind. BP in $\mathbb{P}^{\mathbb{C}; do(C := \tilde{N}_C)}$, there is a total causal effect from C to BP.

Then by Prop (i), there must be a directed path from C to BP.

$$\nexists \text{ directed path from } X \text{ to } Y \implies \nexists \text{ causal effect from } X \text{ to } Y.$$

Total Causal Effect

Is there a directed path from Coffee to Blood pressure?



Randomised trails!

In $\mathbb{C}; do(C := \tilde{N}_C)$ the incoming edges into C are all removed.

Thus, if C not ind. BP in $\mathbb{P}^{\mathbb{C};do(C:=\tilde{N}_C)}$, there is a total causal effect from C to BP.

Then by Prop (i), there must be a directed path from C to BP.

$\nexists$ directed path from X to Y $\implies$ $\nexists$ causal effect from X to Y.

Equivalent of Causal Models

Definition (Equiv. of causal models)

Two models are {**probabilistically**/ **interventionally**/ **counterfactually**} equivalent

if they entail the same {obs./ obs. and int./ obs., int., and counterf.} distributions.

This gives us a way of falsifying models:

(i) probabilistic models -> standard statistical inference.

Assume someone claims drinking coffee (C) is associated with high blood pressure (BP).

Collect data (coffee consumption and blood pressure) then conduct a statistical test of independence between C and BP.

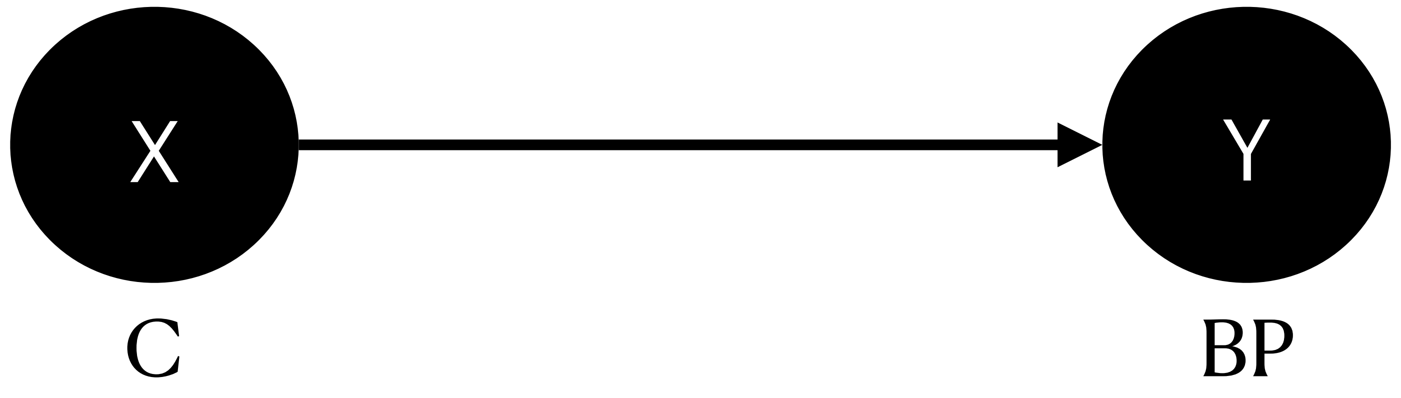
Equivalent of Causal Models

Definition (Equiv. of causal models)

Two models are {**probabilistically**/ **interventionally**/ **counterfactually**} equivalent if they entail the same {obs./ obs. and int./ obs., int., and counterf.} distributions.

This gives us a way of falsifying models:

(ii) interventional models -> randomised trails



Assume someone claims drinking coffee is a cause of high blood pressure.

Randomize coffee intake and measure blood pressure.

Then, if C not ind. BP in $\mathbb{P}^{\mathbb{C};do(C:=\tilde{N}_C)}$, there must be a directed path from C to BP.

Remark: one needs to agree on the notion of "interventions" to falsify the interventional models.

Predictability and Interventions

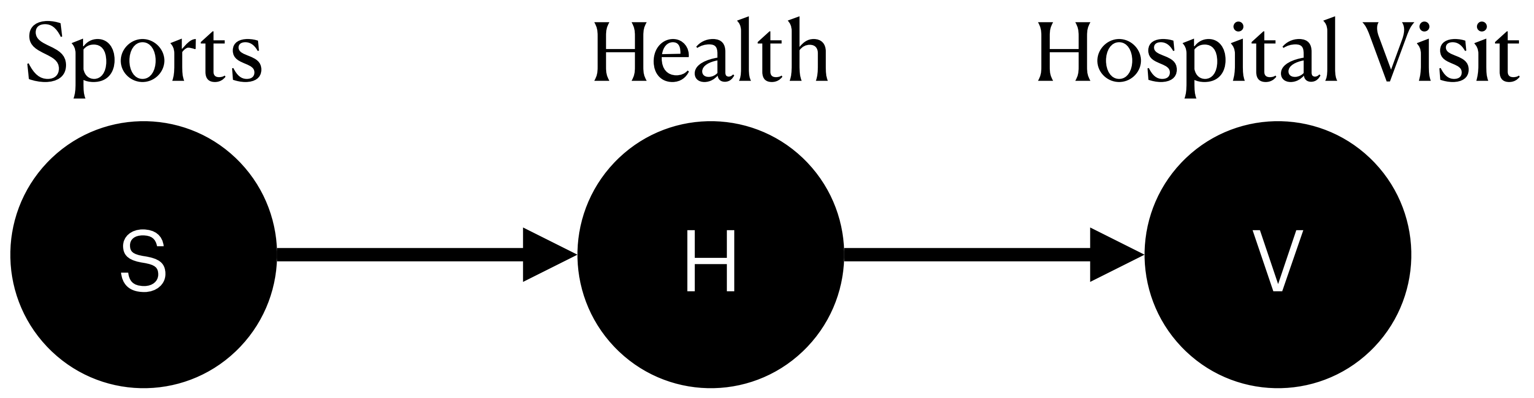
Predictability and Interventions

Example: Consider the following SCM:

$$S := N_A$$

$$H := A \oplus N_H$$

$$V := H \oplus N_B$$



Where $N_A \sim Ber(1/2), N_H \sim Ber(1/3),$ and $N_B \sim Ber(1/20)$, and $\oplus$ is addition modulo 2.

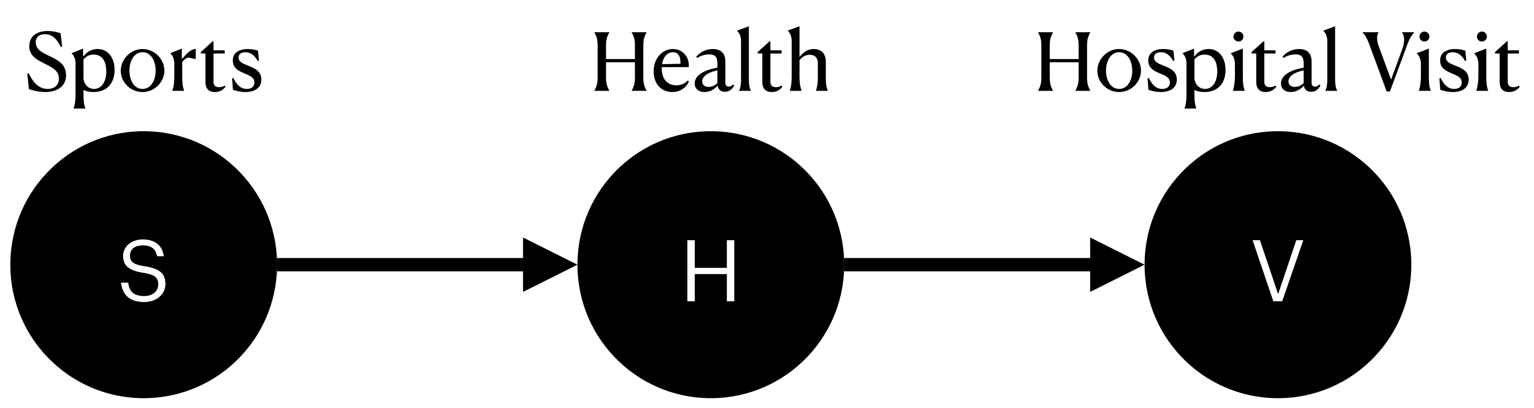
Predictability and Interventions

Example: Consider the following SCM:

$$S := N_A$$

$$H := A \oplus N_H$$

$$V := H \oplus N_B$$



Where $N_A \sim Ber(1/2), N_H \sim Ber(1/3),$ and $N_B \sim Ber(1/20)$, and $\oplus$ is addition modulo 2.

We have that both S and V contain information about H. That is why they both help predicting H. But if you want to improve your health "do" S, not V.

Forcing B to 1 -> $P_H^{do(B:=1)} = P_H = Ber(1/2)$. Forcing A to 1 -> $P_H^{do(A:=1)} Ber(2/3) \neq P_H$.

Counterfactuals

Counterfactuals

Example (taken from Causal Inference in Statistics: A Primer Chapter 4)

Imagine that you were driving home and had to make a choice: to take the highway ($X = 1$) or take a regular street ($X = 0$). You took the regular street and it turned out that the traffic was really bad. When you arrive home one hour later, you said to yourself: “I should have taken the highway.” In other words you think that “My driving time (T) would have been less than one hour had I taken the highway ($X = 1$).”

Counterfactuals

Example (taken from Causal Inference in Statistics: A Primer Chapter 4)

Imagine that you were driving home and had to make a choice: to take the highway ($X = 1$) or take a regular street ($X = 0$). You took the regular street and it turned out that the traffic was really bad. When you arrive home one hour later, you said to yourself: "I should have taken the highway." In other words you think that "My driving time (T) would have been less than one hour had I taken the highway ($X = 1$)."

Observation: T = $1, X = 0$ -> this help infer the highway traffic condition on that day

Intervention: $X = 1$

Aim to estimate : Hypothetical driving time under $X = 1$ under the same highway traffic condition

Counterfactuals

Definition: Consider an SCM $\mathbb{C}$ over nodes $\mathbf{X}$ with noise distribution $\mathbb{P}_{\mathbf{N}}$. Given some observation $\mathbf{X}_S = x_s$ for $S \subseteq \{1, \ldots, d\}$, we define a counterfactual SCM by keeping all the assignments and replacing the distribution of noise variables $\mathbb{P}_{\mathbf{N}}$ by $\mathbb{P}_{\mathbf{N}|\mathbf{X}_S = x_s}$.

(For simplicity, we restrict ourselves to the discrete cases, and the cases where $\forall_j$ either $\mathbf{X}_S = x_s \implies \mathbf{N}_j = n_j$ or $\mathbf{N}_j \perp\!\!\!\perp \mathbf{X}_S$.) (The new noise variables do not need to be jointly independence anymore.)

Counterfactual statements are do-statements in the new counterfactual SCM. We write

$$\mathbb{P}_Z^{\mathbb{C}|X=3, W=2; do(Y:=7)}$$

The original SCM

Observed data

Intervention

Variable of interest

Counterfactuals

Example: Consider the SCM $\mathbb{C}$

$$X := N_X, \qquad Y := X^2 + N_Y, \qquad Z := 2Y + X + N_Z,$$

with $N_X, N_Y, N_Z \sim N(0,1)$. Assume we observe $(X, Y, Z) = (1,2,4)$. Can we compute what would Z have been had X been set to 2, i.e., $\mathbb{P}_Z^{\mathbb{C}|X=1,Y=2,Z=4;do(X:=2)}$?

Counterfactuals

Example: Consider the SCM $\mathbb{C}$

$$X := N_X, \qquad Y := X^2 + N_Y, \qquad Z := 2Y + X + N_Z,$$

with $N_X, N_Y, N_Z \sim N(0,1)$. Assume we observe $(X, Y, Z) = (1,2,4)$. Can we compute what would Z have been had X been set to 2, i.e., $\mathbb{P}_Z^{\mathbb{C}|X=1,Y=2,Z=4;do(X:=2)}$?

Three steps in computing counterfactuals

1. Abduction: Use observations (1,2,4) to determine the value of N_X, N_Y, N_Z.
2. Action: Modify $\mathbb{C}$ by setting $X := 2$ yielding $\mathbb{C}; do(X := 2)$
3. Prediction: Use the modified SCM $\mathbb{C}; do(X := 2)$ and the value of N_X, N_Y, N_Z to compute the value of Z.

Counterfactuals

Example: Consider the SCM $\mathbb{C}$

$$X := N_X, \qquad Y := X^2 + N_Y, \qquad Z := 2Y + X + N_Z,$$

with $N_X, N_Y, N_Z \sim N(0,1)$. Assume we observe $(X, Y, Z) = (1,2,4)$.

Abduction: we have $N_X, N_Y, N_Z = (1,1,-1)$.

Action: setting $X := 2$.

Prediction: $X := 2, \quad Y := 2^2 + 1, \quad Z := 2(5) + 2 + -1 = 11$

Z would have been 11 had X been 2

Y would have been 5 had X been 2 No transitivity in general!

Z would have been 10 had Y been 5

Counterfactuals

Remarks:
(i) We can think of interventional statements as a mathematical construct for (randomized) experiments. For counterfactual statements, there is no comparable correspondence in the real world. This means that, usually, counterfactual statements cannot be falsified. See Example 6.19 in the book (there exist two SCMs that induce the same obs. & int. distributions but no the same counterfactuals).

Counterfactuals

Remarks:
(i) We can think of interventional statements as a mathematical construct for (randomized) experiments. For counterfactual statements, there is no comparable correspondence in the real world. This means that, usually, counterfactual statements cannot be falsified. See Example 6.19 in the book (there exist two SCMs that induce the same obs. & int. distributions but no the same counterfactuals).

(ii) Popular in history. What would have happened had some events in the world been other than as they actually are.

Counterfactuals

Remarks:

(i) We can think of interventional statements as a mathematical construct for (randomized) experiments. For counterfactual statements, there is no comparable correspondence in the real world. This means that, usually, counterfactual statements cannot be falsified. See Example 6.19 in the book (there exist two SCMs that induce the same obs. & int. distributions but no the same counterfactuals).

(ii) Popular in history. What would have happened had some events in the world been other than as they actually are.

(iii) Humans often think in counterfactuals: “I should have taken the highway”, “I should have been more attentive in the class”, or “I should have bought Bitcoin in 2010”.

Counterfactuals

Remarks:

(i) We can think of interventional statements as a mathematical construct for (randomized) experiments. For counterfactual statements, there is no comparable correspondence in the real world. This means that, usually, counterfactual statements cannot be falsified. See Example 6.19 in the book (there exist two SCMs that induce the same obs. & int. distributions but no the same counterfactuals).

(ii) Popular in history. What would have happened had some events in the world been other than as they actually are.

(iii) Humans often think in counterfactuals: “I should have taken the highway”, “I should have been more attentive in the class”, or “I should have bought Bitcoin in 2010”.

(iv) Sometimes the noise variables cannot be uniquely determined from the observations. We instead then compute $P(U \mid \mathbf{X}_S = x_s)$ in the abduction step. (Bayesian flavour)

Counterfactuals

Remarks:
(v) Currently discussed a lot in the fields of Explainability and Fairness in AI.

(Example from Verma, Sahil et al. "Counterfactual explanations for machine learning: A review." *arXiv preprint arXiv:2010.10596* (2020).)

Suppose Alice walks into a bank and seeks a home mortgage loan. The decision is based on a machine learning classifier which considers Alice's features {Income, CreditScore, Education, Age}. Unfortunately, Alice is denied for the loan she seeks and is left wondering (1) why was the loan denied? and (2) what can she do differently so that the loan will be approved in the future? The former question might be answered with traditional explanations like: "CreditScore was too low". The latter question forms the basis of a counterfactual explanation: what small changes could be made to Alice's feature vector in order to receive the loan.

Counterfactuals

(vi) Legal systems - should people be punished based on counterfactual- or do-statements?

Example: Consider the SCM $\mathbb{C}$

$$T := N_T, \quad R := TN_R + (1-T)(1-N_R), \text{ with } N_R \sim Ber(0.99)$$

Patient Alice: $\quad T = 1, \quad R = 0 \quad (\implies N_R = 0)$

For Alice, $T = 0$ would have been better.

We have $\mathbb{P}^{\mathbb{C}|T=1,R=0;do(T:=0)}(R=1) = 1.$

But the doctor doesn't know N_R in advance, shall Alice receive $ from the doctor?