

TECHNISCHE
UNIVERSITÄT
DARMSTADT



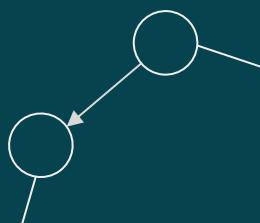
Causal Identification and Estimation

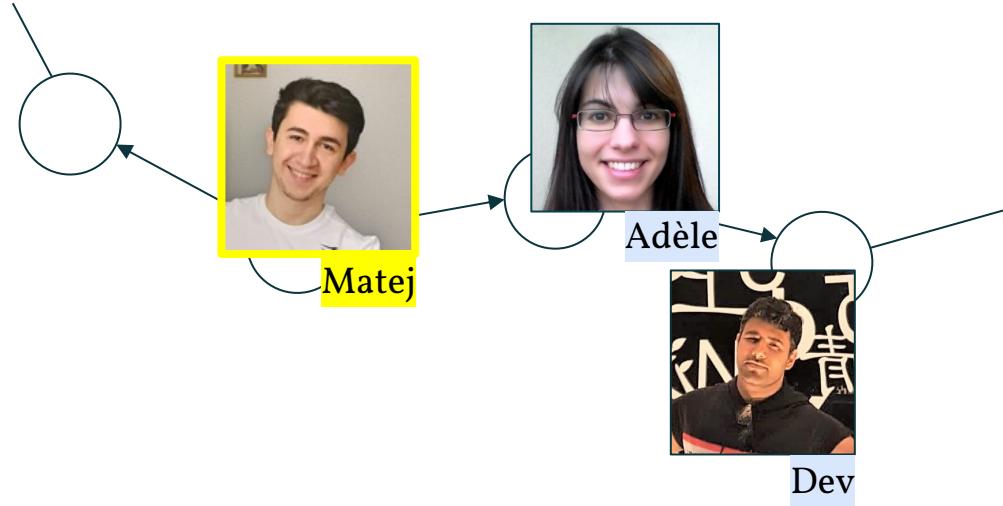
Matej Zečević

26th July 2023



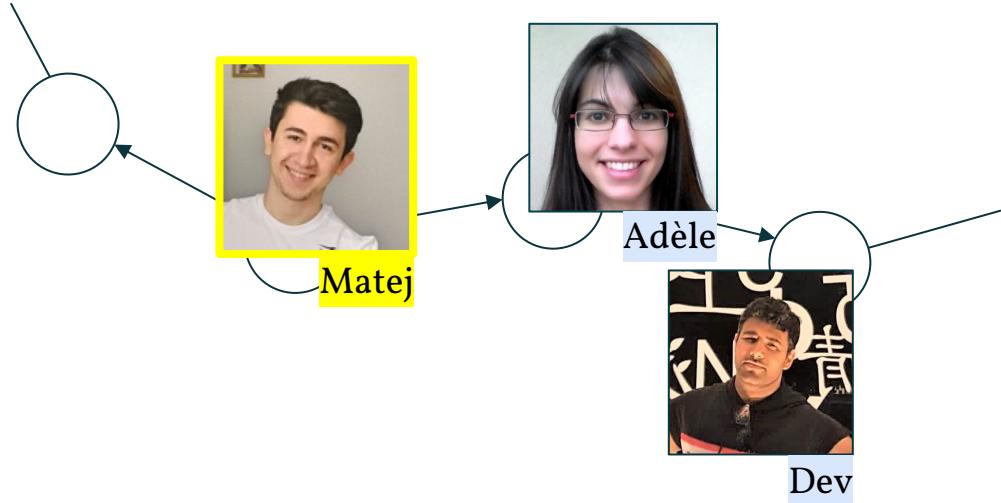
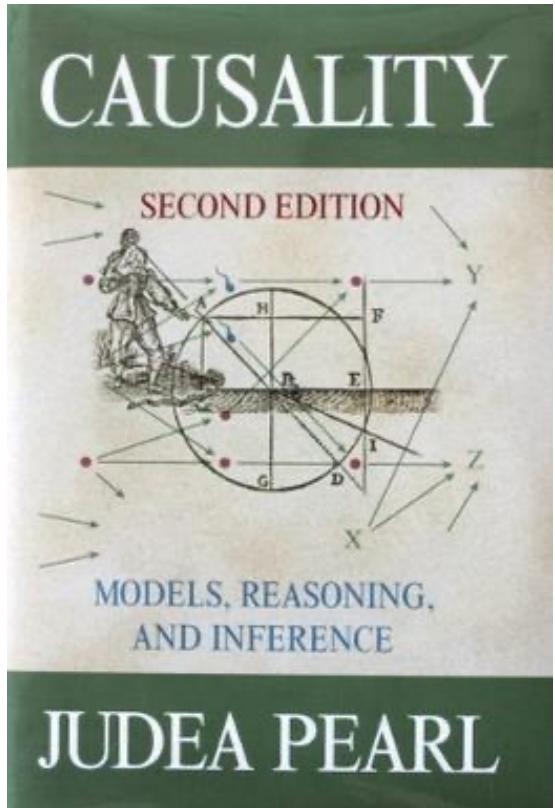
Machines Climbing Pearl's Ladder of Causation





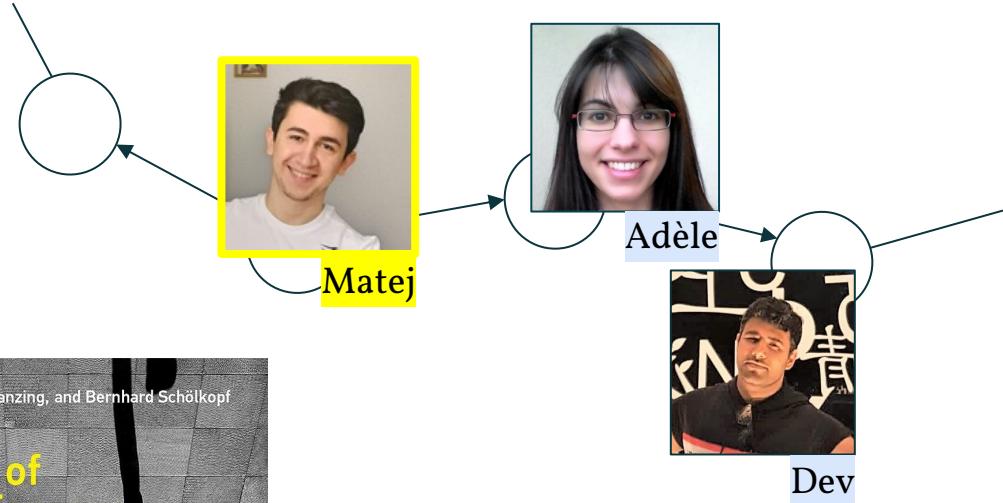
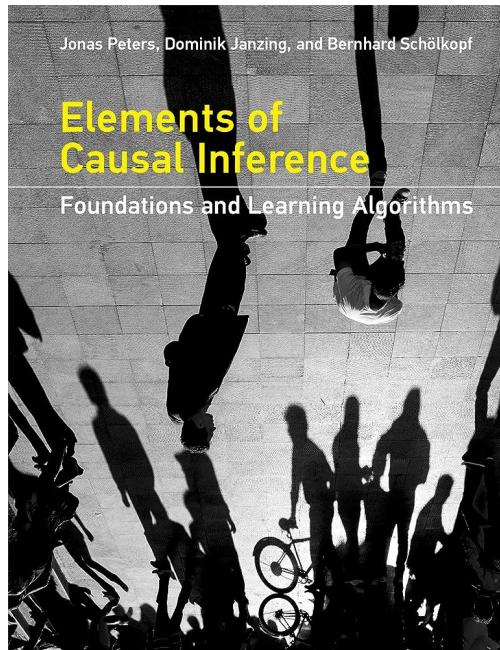
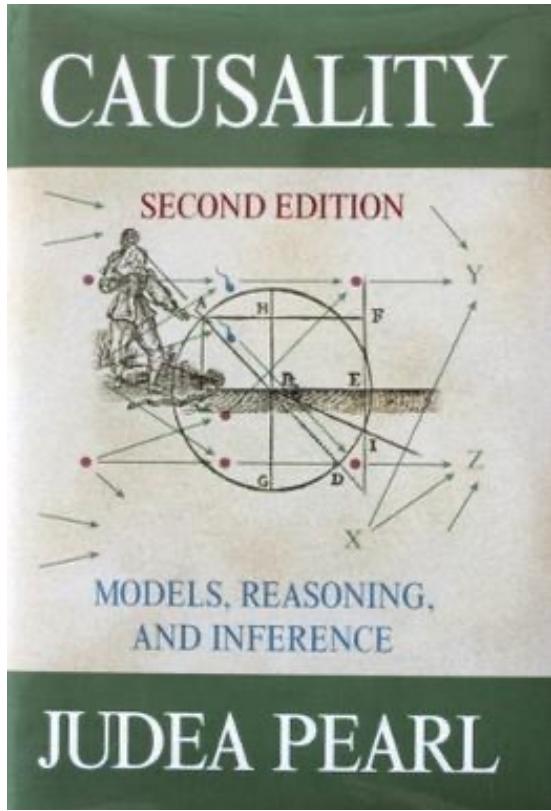
ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Overview

- Defining Causal Effect Identifiability

Overview

- Defining Causal Effect Identifiability
- Identifying using: (a) Adjustment Sets,

Overview

- Defining Causal Effect Identifiability
- Identifying using: (a) Adjustment Sets,
(b) Truncated Factorization,

Overview

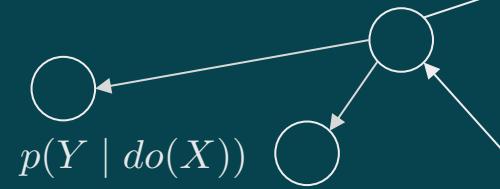
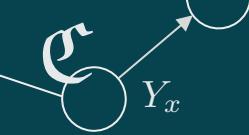
- Defining Causal Effect Identifiability
- Identifying using: (a) Adjustment Sets,
(b) Truncated Factorization,
(c) Pearl's do-Calculus

Overview

- Defining Causal Effect Identifiability
- Identifying using: (a) Adjustment Sets,
(b) Truncated Factorization,
(c) Pearl's do-Calculus and
(d) optimization (Partial Identification)

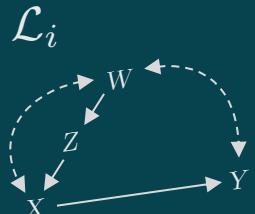
Overview

- Defining Causal Effect Identifiability
- Identifying using: (a) Adjustment Sets,
(b) Truncated Factorization,
(c) Pearl's do-Calculus and
(d) optimization (Partial Identification)
- Estimation with ML, Cases: iSPN & NCM



O Brief Recap

All Crucial (Formal) Concepts
for this Lecture



Machines Climbing Pearl's Ladder of Causation

Structural Causal Model

Definition

A structural causal model \mathcal{M} (or data generating model) is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P_{\mathbf{U}} \rangle$, where

Structural Causal Model

Definition

A structural causal model \mathcal{M} (or data generating model) is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P_{\mathbf{U}} \rangle$, where

\mathbf{V} are endogenous variables

\mathbf{U} are exogenous variables

\mathcal{F} are functions determining \mathbf{V} i.e., $v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)$

Structural Causal Model

Definition

A structural causal model \mathcal{M} (or data generating model) is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P_{\mathbf{U}} \rangle$, where

\mathbf{V} are endogenous variables

\mathbf{U} are exogenous variables

\mathcal{F} are functions determining \mathbf{V} i.e., $v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)$

$P_{\mathbf{U}}$ is the probability distribution over \mathbf{U} .

Structural Causal Model

Definition

A structural causal model \mathcal{M} (or data generating model) is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P_{\mathbf{U}} \rangle$, where

\mathbf{V} are endogenous variables

\mathbf{U} are exogenous variables

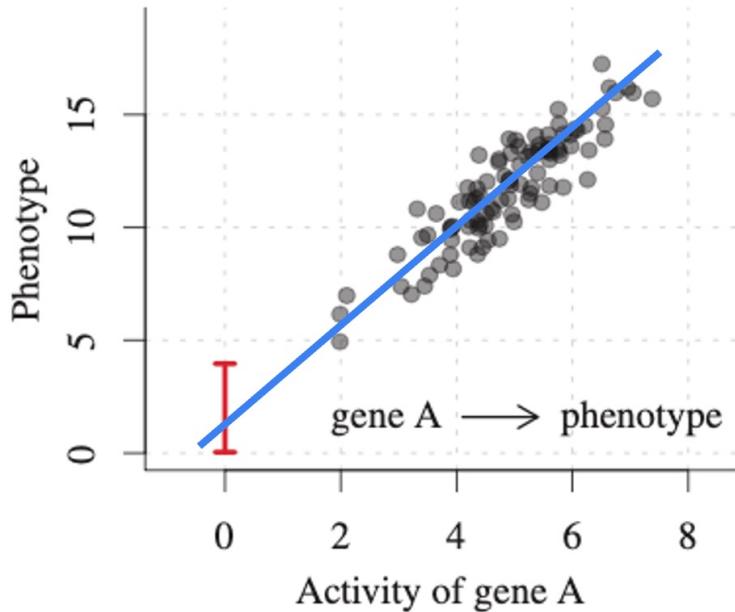
\mathcal{F} are functions determining \mathbf{V} i.e., $v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)$

$P_{\mathbf{U}}$ is the probability distribution over \mathbf{U} .

Assumption: \mathcal{M} is recursive i.e., there are no feedback (cyclic) mechanisms

SCM as Data-Generating Processes

The Observational Distribution (Non-causal)



$$P(Y = y \mid X = x)$$

Interventions

Changing the SCM's structural equations

$$\mathcal{M} \left\{ \begin{array}{l} \mathbf{V} = \{X, Y\} \quad \mathbf{U} = \{U_{XY}, U_X, U_Y\} \quad P_{\mathbf{U}} \\ \mathcal{F} = \left\{ \begin{array}{l} X = f_X(U_X, U_{XY}) \\ Y = f_Y(X, U_Y, U_{XY}) \end{array} \right. \end{array} \right.$$

Interventions

Changing the SCM's structural equations

$$\mathcal{M} \left\{ \begin{array}{l} \mathbf{V} = \{X, Y\} \quad \mathbf{U} = \{U_{XY}, U_X, U_Y\} \quad P_{\mathbf{U}} \\ \mathcal{F} = \left\{ \begin{array}{ll} X = f_X(U_X, U_{XY}) & \\ Y = f_Y(X, U_Y, U_{XY}) & P(Y = y \mid X = x) \neq P(Y = y) \end{array} \right. \end{array} \right.$$

Interventions

Changing the SCM's structural equations

$$\mathcal{M} \left\{ \begin{array}{l} \mathbf{V} = \{X, Y\} \quad \mathbf{U} = \{U_{XY}, U_X, U_Y\} \quad P_{\mathbf{U}} \\ \mathcal{F} = \left\{ \begin{array}{ll} X = f_X(U_X, U_{XY}) & \\ Y = f_Y(X, U_Y, U_{XY}) & P(Y = y \mid X = x) \neq P(Y = y) \end{array} \right. \end{array} \right.$$

↓ $do(X = x)$

$$\mathcal{M}_x \quad \mathcal{F} = \left\{ \begin{array}{l} X = x \\ Y = f_Y(x, U_Y, U_{XY}) \end{array} \right.$$

Interventions

Changing the SCM's structural equations

$$\mathcal{M} \left\{ \begin{array}{l} \mathbf{V} = \{X, Y\} \quad \mathbf{U} = \{U_{XY}, U_X, U_Y\} \quad P_{\mathbf{U}} \\ \mathcal{F} = \left\{ \begin{array}{ll} X = f_X(U_X, U_{XY}) & \\ Y = f_Y(X, U_Y, U_{XY}) & P(Y = y \mid X = x) \neq P(Y = y) \end{array} \right. \end{array} \right.$$

↓
 $do(X = x)$

$$\mathcal{M}_x \quad \mathcal{F} = \left\{ \begin{array}{ll} X = x & \\ Y = f_Y(x, U_Y, U_{XY}) & P(Y = y \mid do(X = x)) \neq P(Y = y) \end{array} \right.$$

The Causal Graph

An induced property of the SCM

latent

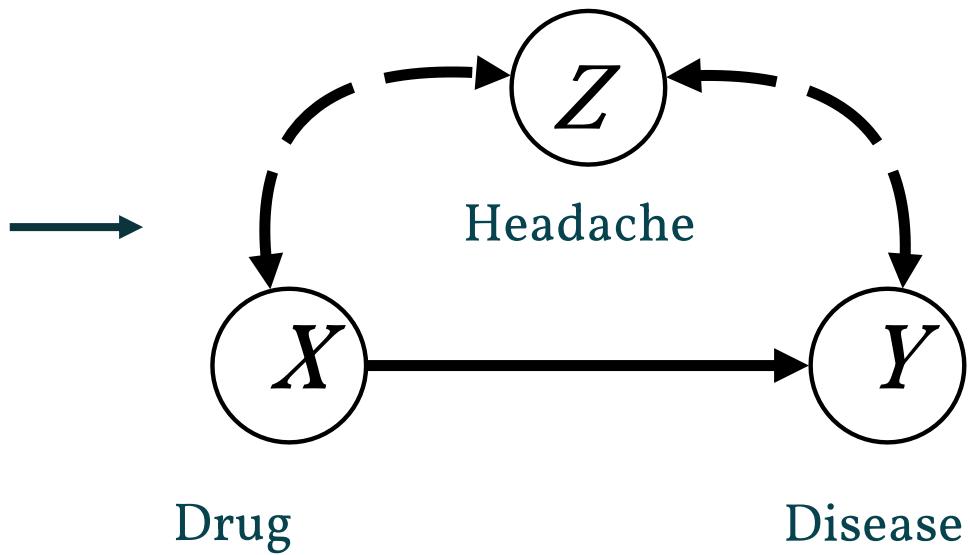
$$\mathcal{F} = \begin{cases} X = f_X(U_X, U_{XZ}) \\ Y = f_Y(X, U_Y, U_{YZ}) \\ Z = f_Z(U_Z) \end{cases}$$

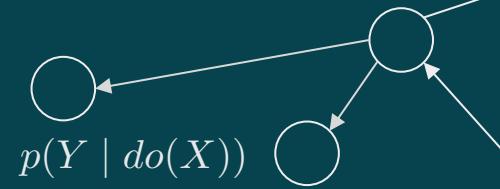
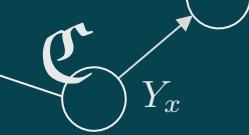
The Causal Graph

An induced property of the SCM

latent

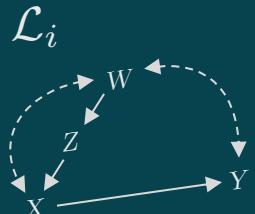
$$\mathcal{F} = \left\{ \begin{array}{l} X = f_X(U_X, U_{XZ}) \\ Y = f_Y(X, U_Y, U_{YZ}) \\ Z = f_Z(U_Z) \end{array} \right.$$





I | Causal Effect Identifiability

Definition and Intuition



Machines Climbing Pearl's Ladder of Causation

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X

on a (set of) outcome variable(s) Y is said to be identifiable

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X
on a (set of) outcome variable(s) Y is said to be identifiable
from a causal diagram G and observational distribution P(V)

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X

on a (set of) outcome variable(s) Y is said to be identifiable

from a causal diagram G and observational distribution $P(V)$

if the interventional distribution $P(Y | \text{do}(X))$ is uniquely computable.

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X

on a (set of) outcome variable(s) Y is said to be identifiable

That is, if for every pair of SCMs M_1 and M_2

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X

on a (set of) outcome variable(s) Y is said to be identifiable

That is, if for every pair of SCMs M_1 and M_2

that induce G and have $P_{M_1}(V) = P_{M_2}(V) = P(V)$ with $P(V) > 0$,

‘Formal’ Definition

Definition 1, Causal Effect Identifiability:

The causal effect of a (set of) treatment variable(s) X

on a (set of) outcome variable(s) Y is said to be identifiable

That is, if for every pair of SCMs M_1 and M_2

that induce G and have $P_{M_1}(V) = P_{M_2}(V) = P(V)$ with $P(V) > 0$,

it must hold: $P_{M_1}(Y | \text{do}(X)) = P_{M_2}(Y | \text{do}(X)) = P(Y | \text{do}(X))$.

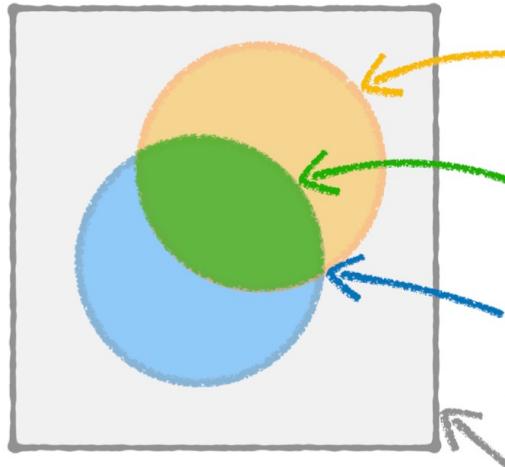


ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

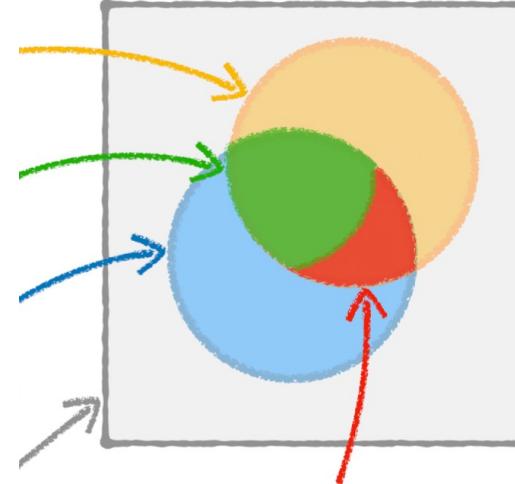
Machines Climbing Pearl’s Ladder of Causation

Intuition for Def.I: Set / Venn Diagram View

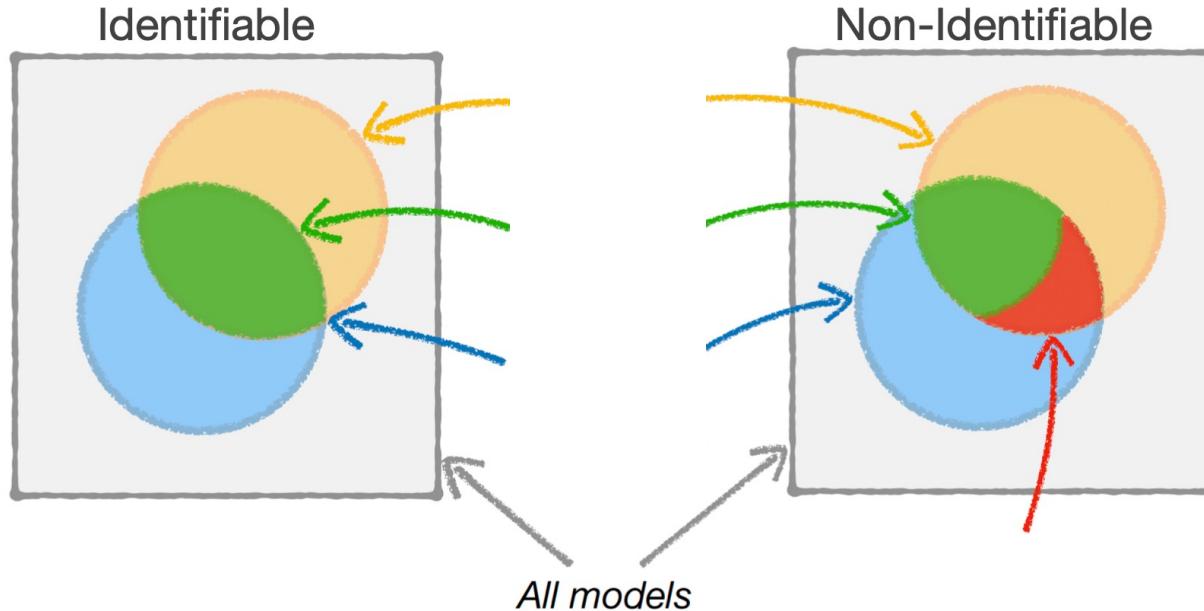
Identifiable



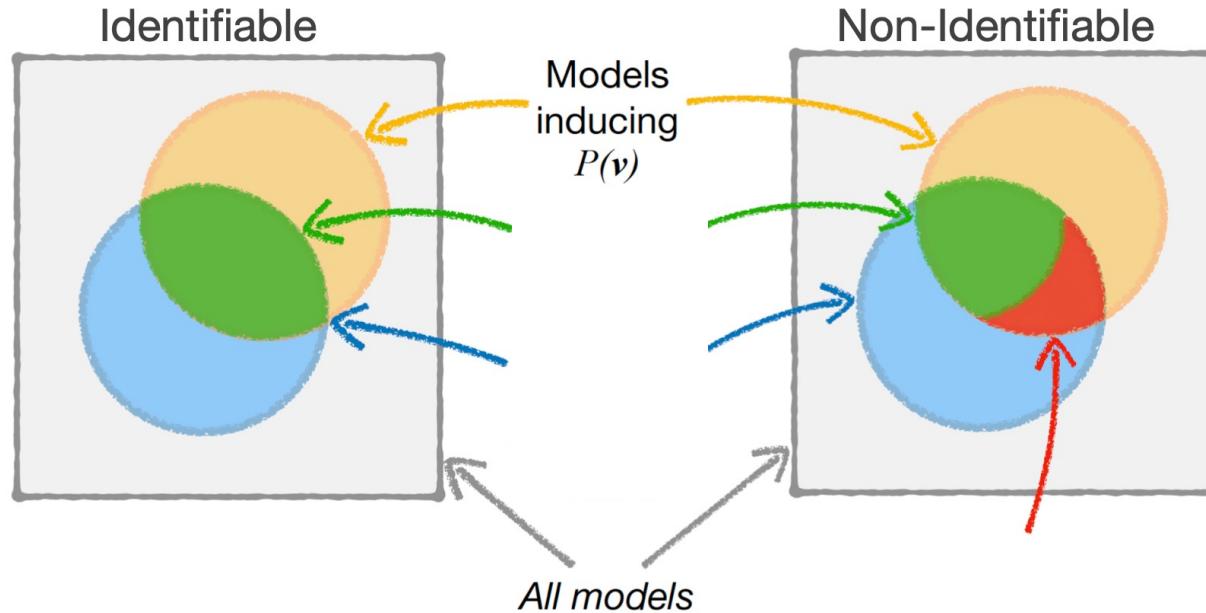
Non-Identifiable



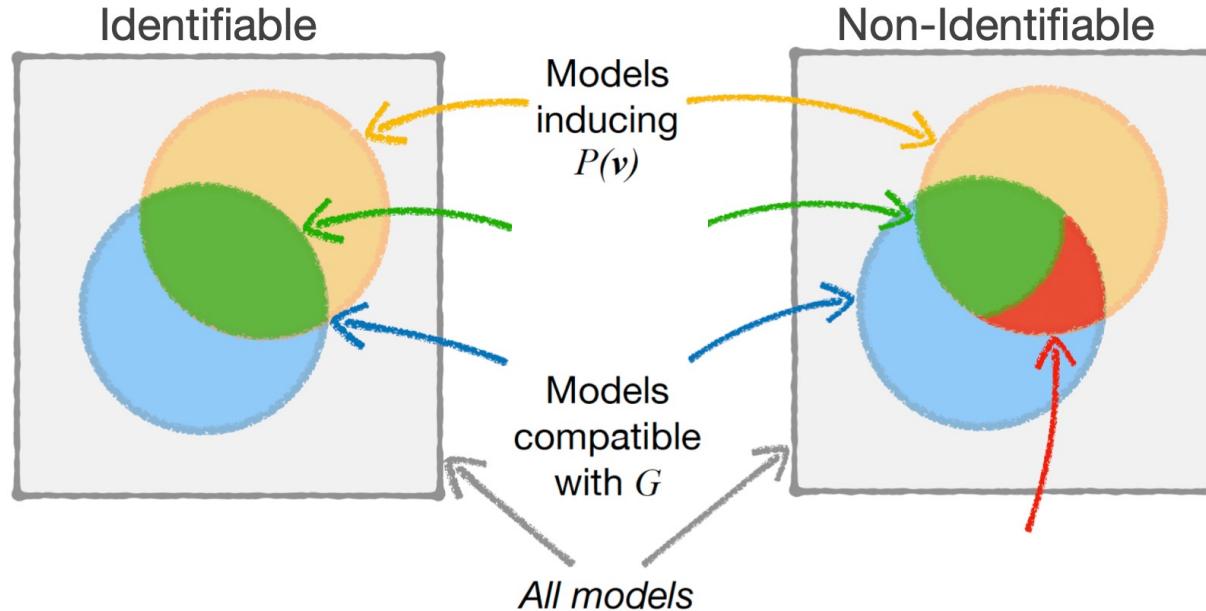
Intuition for Def.I: Set / Venn Diagram View



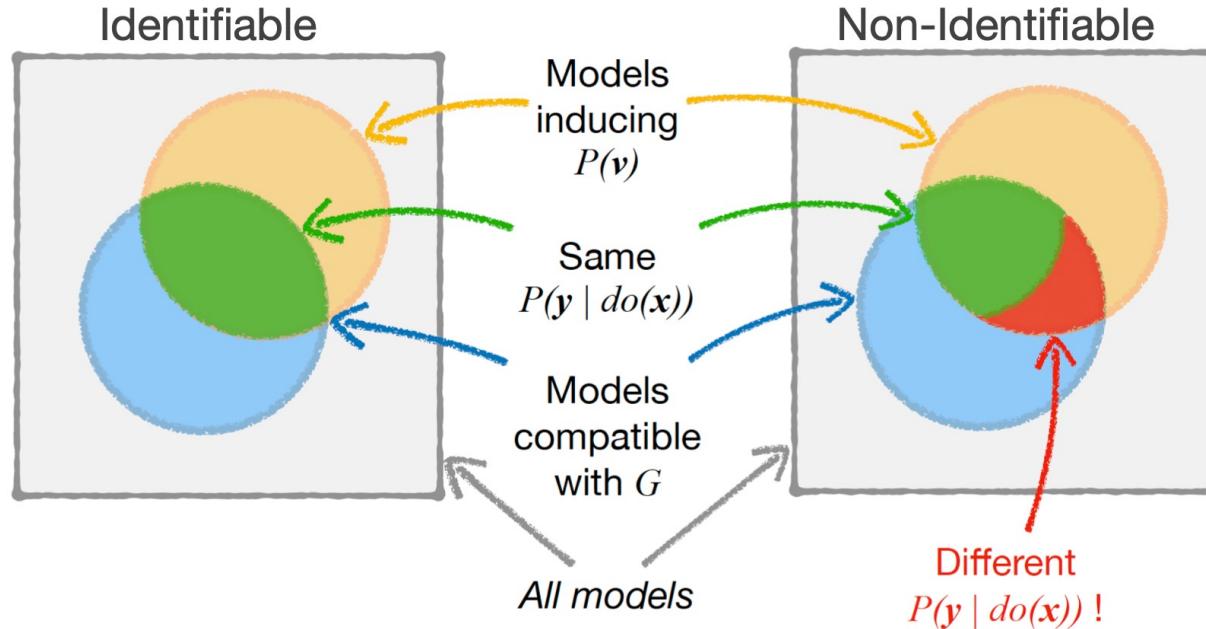
Intuition for Def.I: Set / Venn Diagram View



Intuition for Def.I: Set / Venn Diagram View



Intuition for Def.I: Set / Venn Diagram View



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

A Special Sub-class of SCM

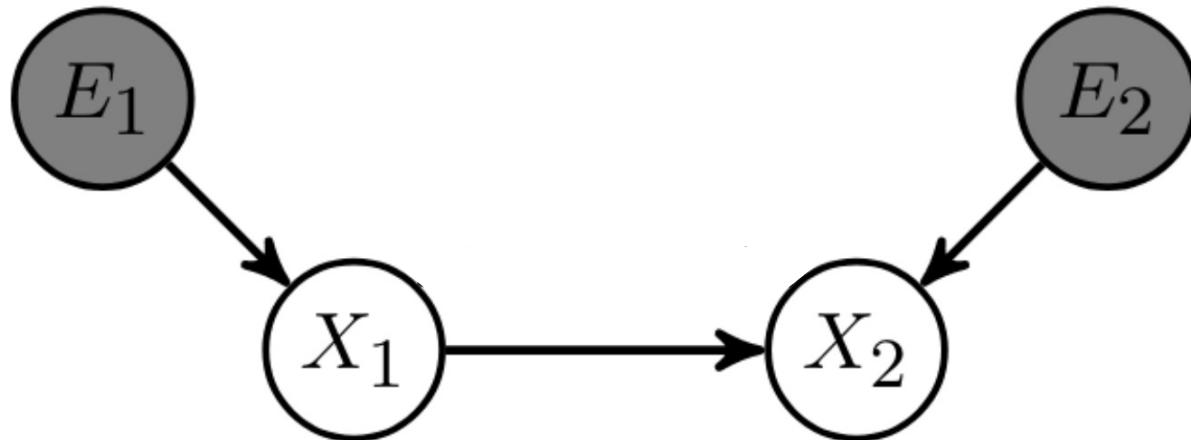
Definition 2, Markovian SCM:

Let $M = (V, U, F, P(U))$ be an SCM.

If $P(U)$ is a product distribution over U ,

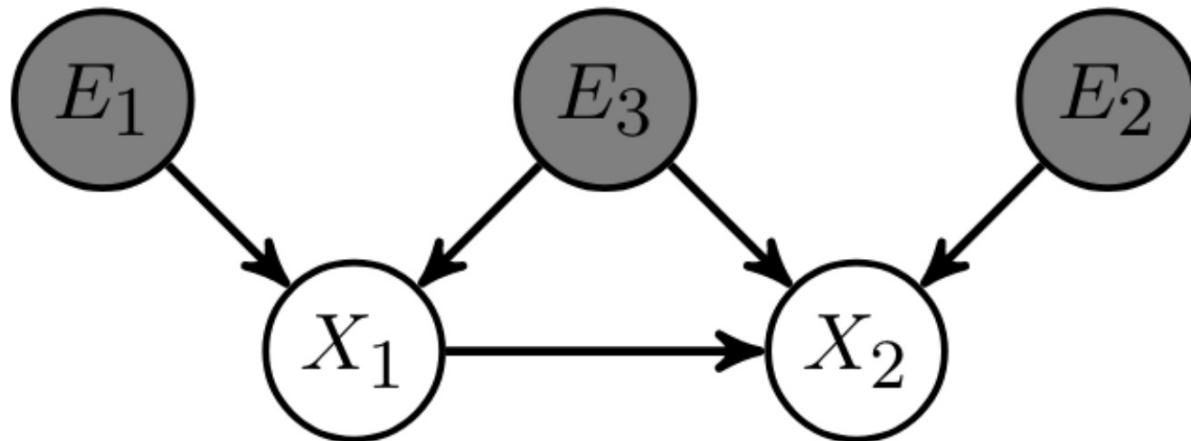
then M is called Markovian.

Intuition for Def.2: Graphical View



depicted $G_a(M)$ and Def.2 satisfied, M is Markovian

Intuition for Def.2: Graphical View



Def.2 not satisfied, M is non-/semi-Markovian

Theorem for Markovian SCM:

Theorem for Markovian SCM:

Any causal effect is identifiable. \square

What are causal effects?

Definition 3, Causal Effect:

Any quantity Q derived from $P(Y | \text{do}(X))$

that tells us how much Y changes due to an intervention $\text{do}(X)$.

What are causal effects?

Definition 3, Causal Effect:

Any quantity Q derived from $P(Y | \text{do}(X))$

that tells us how much Y changes due to an intervention $\text{do}(X)$.

Example Quantity:

$$Q := E(Y | \text{do}(X=x')) - E(Y | \text{do}(X=x))$$

What are causal effects?

Definition 3, Causal Effect:

Any quantity Q derived from $P(Y | \text{do}(X))$

that tells us how much Y changes due to an intervention $\text{do}(X)$.

Example Quantity:

$$Q := E(Y | \text{do}(X=x')) - E(Y | \text{do}(X=x))$$

$$E(Y | x) := \sum_y y \cdot P(y | x)$$

What are causal effects?

Definition 3, Causal Effect:

Any quantity Q derived from $P(Y | \text{do}(X))$

that tells us how much Y changes due to an intervention $\text{do}(X)$.

Example Quantity:

$$Q := E(Y | \text{do}(X=x')) - E(Y | \text{do}(X=x)) \quad E(Y | x) := \sum_y y \cdot P(y | x)$$

known as Average Treatment Effect (ATE)

The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain

1

Query

$$P(y | do(x))$$

The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain

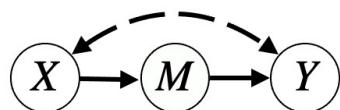
1

Query

$$P(y | do(x))$$

2

Causal Constraints



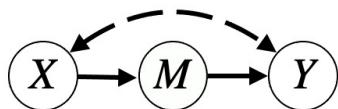
The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain

1 **Query**

$$P(y | do(x))$$

2 **Causal Constraints**

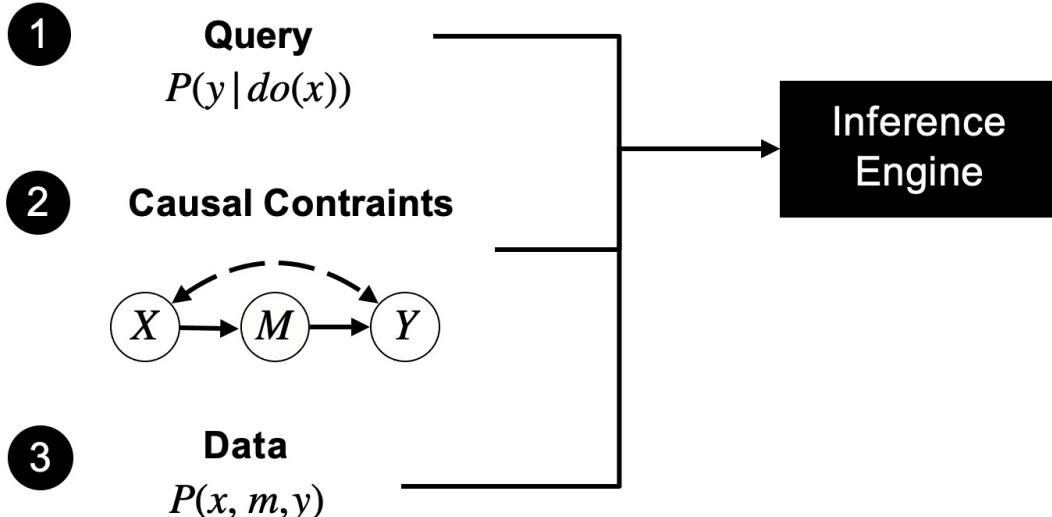


3 **Data**

$$P(x, m, y)$$

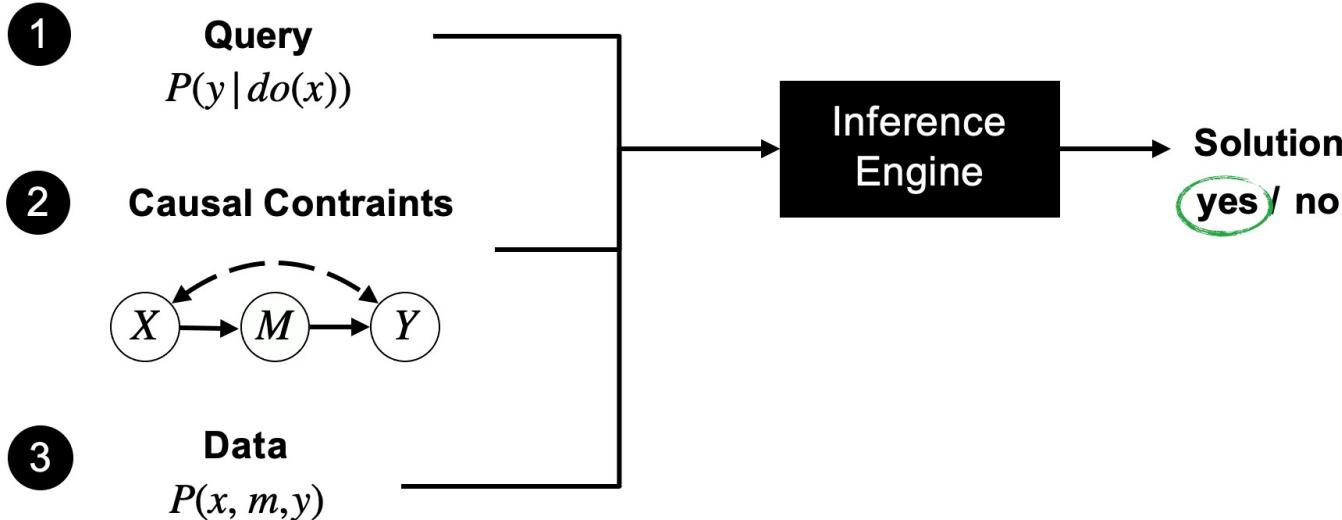
The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain



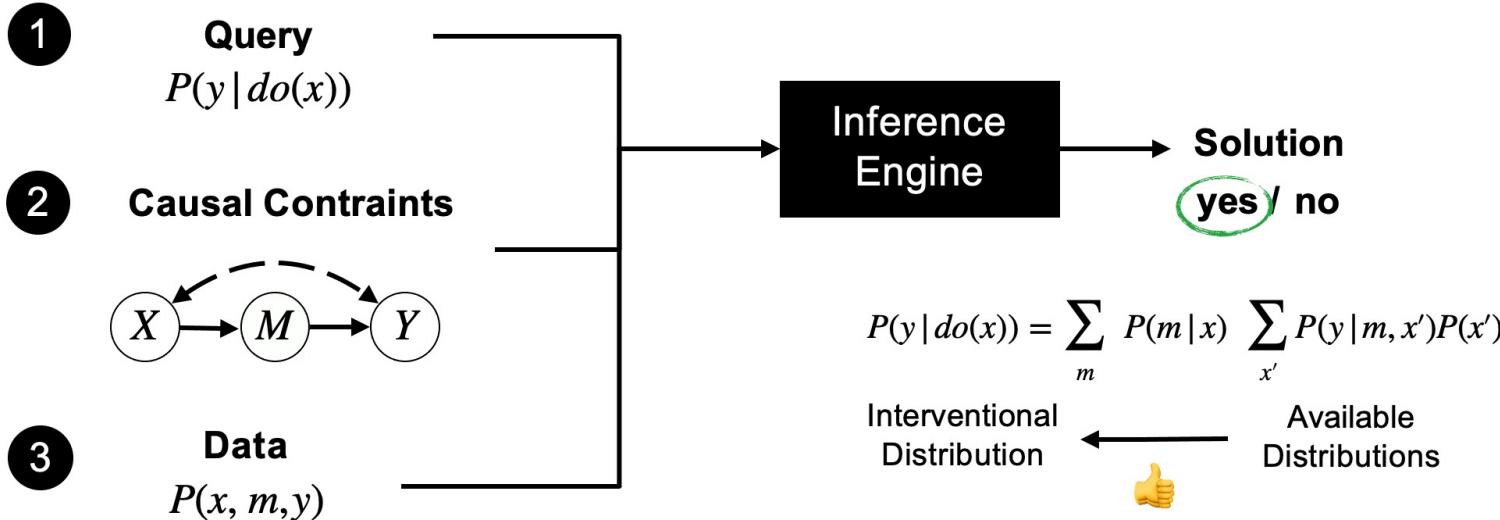
The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain



The Identification Problem ‘Big Picture’

Using Observational Data From One Population/Domain



Solving the Ident. Problem

Underlying Model

Corresponding Ident. Approach

- For Markovian Models use Truncated Factorization

Solving the Ident. Problem

Underlying Model

Corresponding Ident. Approach

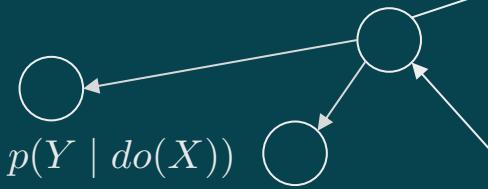
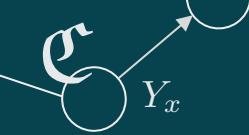
- For Markovian Models use Truncated Factorization
- For certain special graphs use Adjustment Sets

Solving the Ident. Problem

Underlying Model

Corresponding Ident. Approach

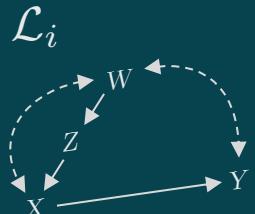
- For Markovian Models use Truncated Factorization
- For certain special graphs use Adjustment Sets
- For General SCM use Pearl's do-Calculus



2

Identifying using

Adjustment Sets, Truncated Factorization,
Pearl's do-Calculus and Optimization



Simpson's Paradox

$$P(R | A) < P(R | B)$$

Overall

Treatment *a*:
Open surgery 78% (273/350)

Treatment *b*:
Percutaneous
nephrolithotomy 83% (289/350)

Simpson's Paradox

$$P(R | A) < P(R | B)$$

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

Simpson's Paradox

$$P(R | A) < P(R | B)$$

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

however

$$P(R | \text{do}(A)) > P(R | \text{do}(B))$$

When Correlation ≠ Causation

Definition 4, Confounding:

If $P(Y | \text{do}(X)) \neq P(Y | X)$, then the causal mechanism from X to Y is called confounded.

When Correlation \neq Causation*

Definition 4, Confounding:

If $P(Y | \text{do}(X)) \neq P(Y | X)$, then the causal mechanism from X to Y is called confounded.

* confounding is not ‘evil’ per se! Synonymous to “common cause.”

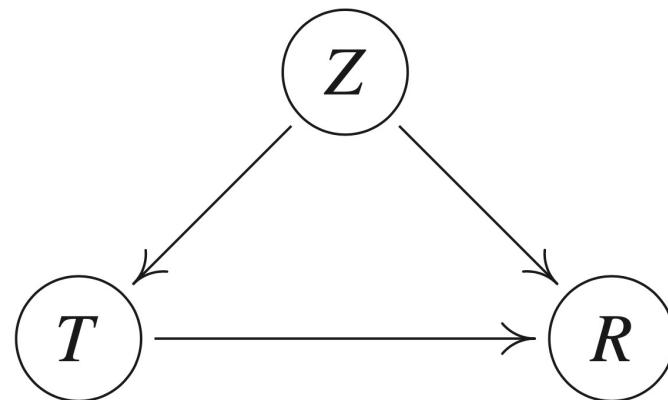
The Kidney Stone Example

For any patient we are given (binary) records on

Z , the kidney stone size

T , the treatment and

R , the recovery.



Conditioning ≠ Intervening

Example Calculation:

$$P(R=I \mid do(T=A)) = \dots$$

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



ESSAI & ACN 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Conditioning ≠ Intervening

Example Calculation:

$$P(R=I \mid do(T=A)) = 0.93 \cdot \quad + 0.73 \cdot \quad =$$

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

Conditioning ≠ Intervening

Example Calculation:

$$P(R=I \mid do(T=A)) = 0.93 \cdot (357/700) + 0.73 \cdot (343/700) = 0.832$$

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



ESSAI & ACN 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Conditioning \neq Intervening

Example Calculation:

$$P(R=I \mid \text{do}(T=A)) = 0.93 \cdot (357/700) + 0.73 \cdot (343/700) = 0.832$$

$$P(R=I \mid T=A) = 0.78$$

Conditioning \neq Intervening

Example Calculation:

$$P(R=I \mid \text{do}(T=A)) = 0.93 \cdot (357/700) + 0.73 \cdot (343/700) = 0.832$$

$$P(R=I \mid T=A) = 0.78$$

analogously for $T=B$

Conditioning ≠ Intervening

Example Calculation:

$$P(R=I \mid \text{do}(T=A)) = 0.93 \cdot (357/700) + 0.73 \cdot (343/700) = 0.832$$

$$P(R=I \mid T=A) = 0.78$$

analogously for $T=B$

$$P(R=I \mid \text{do}(T=A)) - P(R=I \mid \text{do}(T=B)) = 0.832 - 0.782$$

$$P(R=I \mid T=A) - P(R=I \mid T=B) = 0.78 - 0.83$$

Density View

$$P(R | do(T)) =$$

$$\sum_T P(R | T, Z) \cdot P(Z)$$

Density View

$$P(R | do(T)) =$$

$$\sum_T P(R | T, Z) \cdot P(Z) \not\equiv \sum_T P(R | T, Z) \cdot P(Z | T)$$

$$= P(R | T)$$

Special Sets that yield Identification

Theorem I, Adjustment Set:

For any pair (X, Y) with $Y \notin pa(X)$ the set Z will satisfy the equality

$$P(Y | do(X)) = \sum_Z P(Y | X, Z) \cdot P(Z), \text{ if either:}$$

Special Sets that yield Identification

Theorem I, Adjustment Set:

For any pair (X, Y) with $Y \notin pa(X)$ the set Z will satisfy the equality

$P(Y | do(X)) = \sum_Z P(Y | X, Z) \cdot P(Z)$, if either:

(i) $Z := pa(X)$,

Special Sets that yield Identification

Theorem I, Adjustment Set:

For any pair (X, Y) with $Y \notin \text{pa}(X)$ the set Z will satisfy the equality

$P(Y | \text{do}(X)) = \sum_Z P(Y | X, Z) \cdot P(Z)$, if either:

(i) $Z := \text{pa}(X)$,

(ii) $Z := W$ where W contains no descendants of X and blocks all backdoor paths.,

Special Sets that yield Identification

Theorem I, Adjustment Set:

For any pair (X, Y) with $Y \notin \text{pa}(X)$ the set Z will satisfy the equality

$P(Y | \text{do}(X)) = \sum_Z P(Y | X, Z) \cdot P(Z)$, if either:

(i) $Z := \text{pa}(X)$,

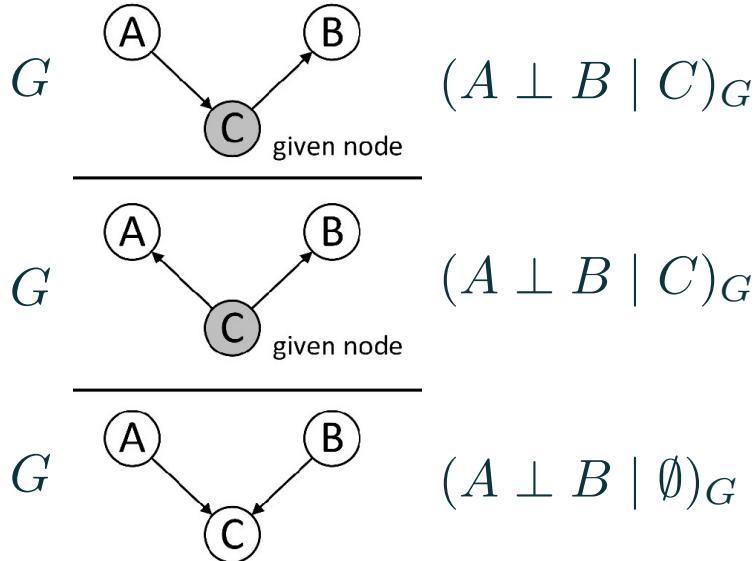
(ii) $Z := W$ where W contains no descendants of X and blocks all backdoor paths,

or (iii) $Z := W$ where W contains no descendant of any node

on a directed path from X to Y and blocks all non-directed paths from X to Y .

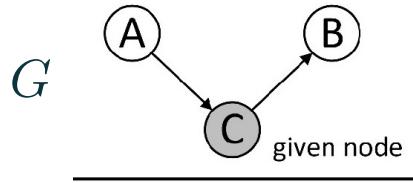
d-Separation & Conditional Independence

Graphical Tools



d-Separation & Conditional Independence

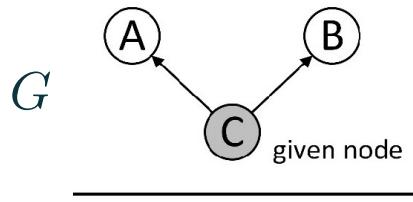
Graphical Tools



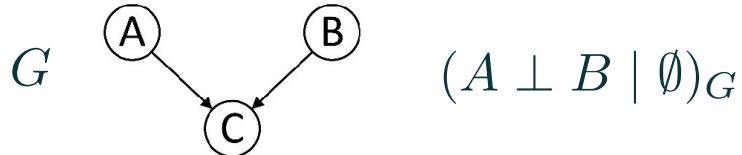
$$(A \perp B \mid C)_G$$

global Markov property

$$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_G \implies (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_P$$



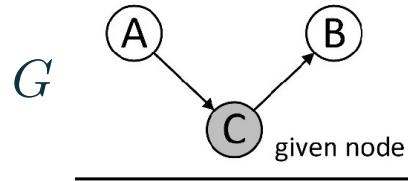
$$(A \perp B \mid C)_G$$



$$(A \perp B \mid \emptyset)_G$$

d-Separation & Conditional Independence

Graphical Tools

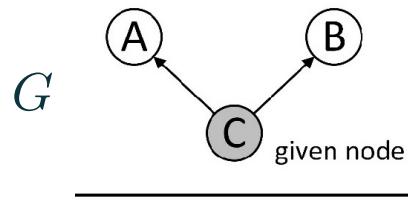


$$(A \perp B \mid C)_G$$

global Markov property

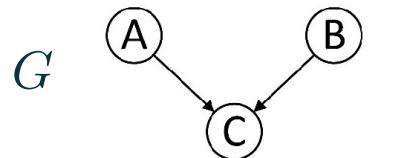
$$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_G \implies (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_P$$

faithfulness



$$(A \perp B \mid C)_G$$

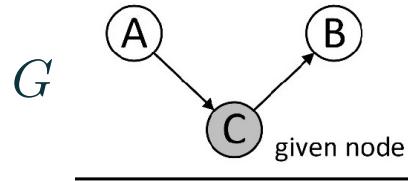
$$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_P \implies (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_G$$



$$(A \perp B \mid \emptyset)_G$$

d-Separation & Conditional Independence

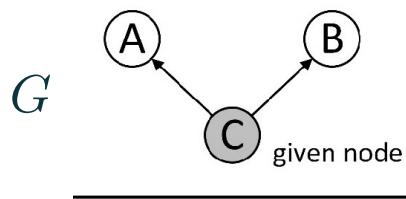
Graphical Tools



$$(A \perp B \mid C)_G$$

global Markov property

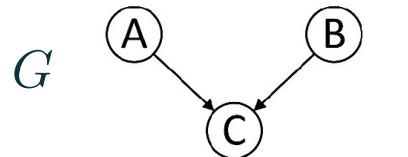
$$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_G \implies (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_P$$



$$(A \perp B \mid C)_G$$

faithfulness

$$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_P \implies (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})_G$$



$$(A \perp B \mid \emptyset)_G$$

local factorization

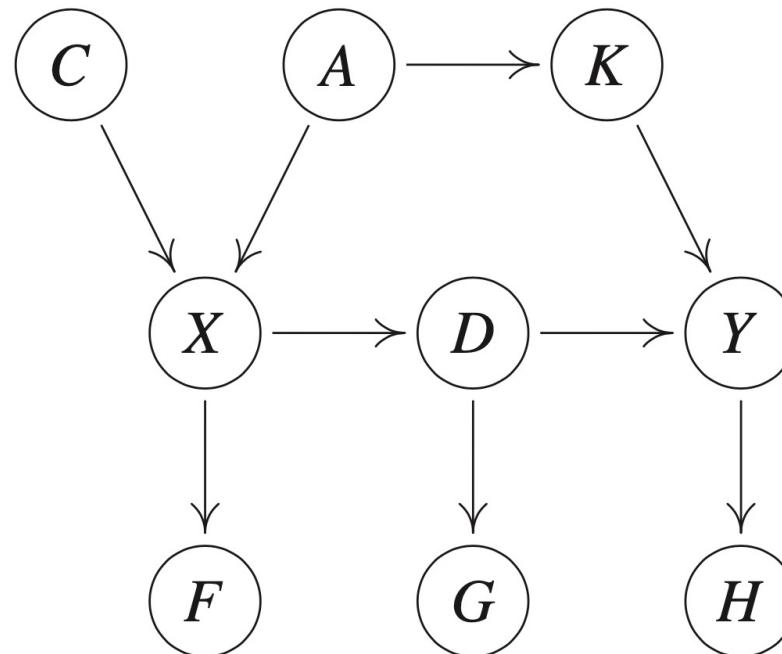
$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_i P(v_i \mid \text{pa}_i, u_i)$$



Special Sets that yield Identification

Theorem I gives us backdoor

adjustment set $Z = \{K\}$



Special Sets that yield Identification

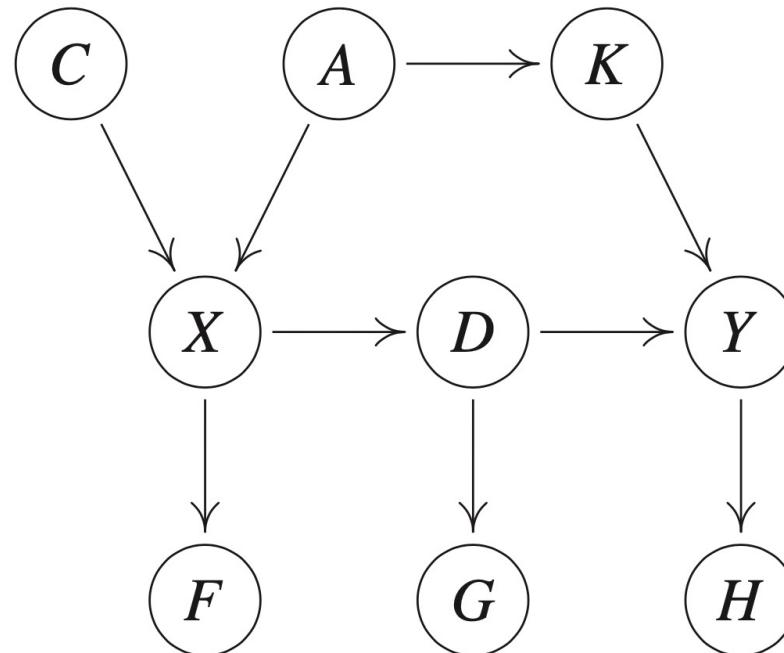
Theorem I gives us backdoor

adjustment set $Z = \{K\}$,

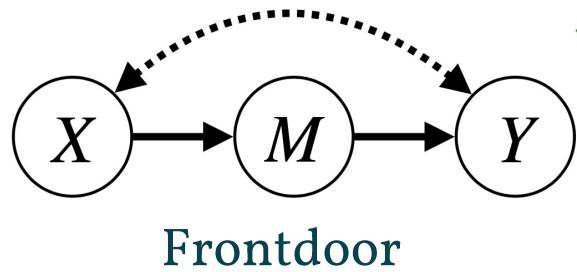
alternate non-backdoor

valid adjustment set is

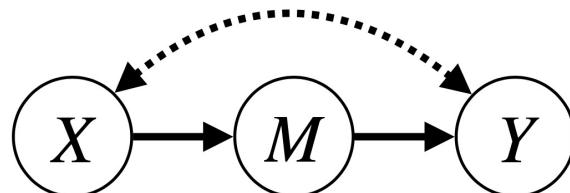
$$Z = \{F, C, K\}$$



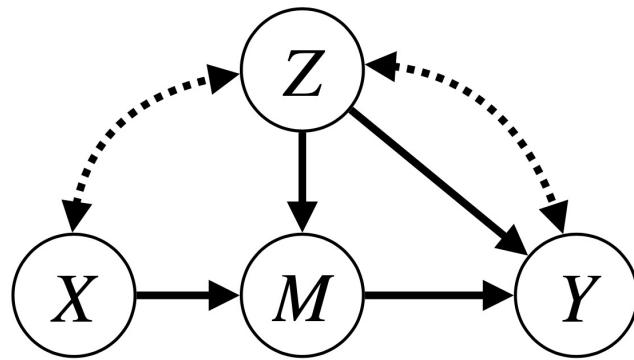
Plenty of ‘Special’ Graphs with corresponding Criteria



Plenty of ‘Special’ Graphs with corresponding Criteria

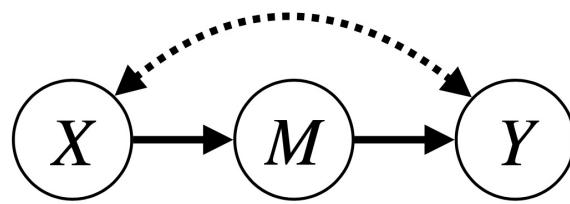


Frontdoor

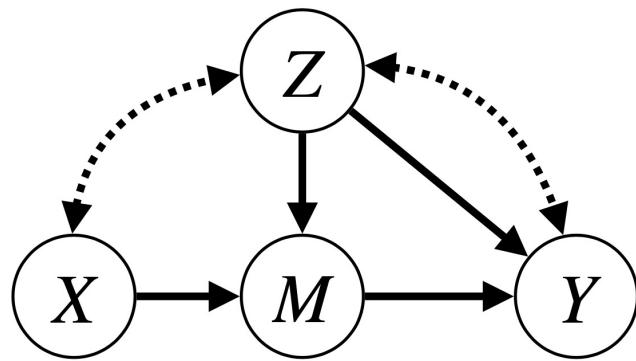


Conditional Frontdoor

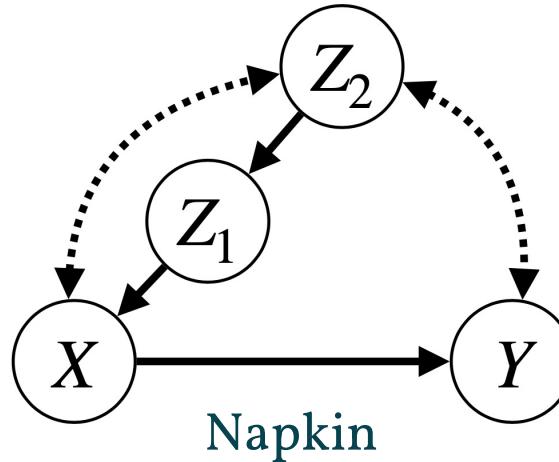
Plenty of ‘Special’ Graphs with corresponding Criteria



Frontdoor



Conditional Frontdoor



Napkin

Switching the approach to identification next

Adjustment Sets,
Truncated Factorization and Pearl's do-Calculus



Machines Climbing Pearl's Ladder of Causation

‘Hard’ interventions

Definition 5, Hard Intervention:

Given $x=g(z)$ for some $Z \subset V$ in $\text{do}(X=x)$,

‘Hard’ interventions

Definition 5, Hard Intervention:

Given $x=g(z)$ for some $Z \subset V$ in $\text{do}(X=x)$,

if x cuts off the dependency of X to $\text{pa}(X)$,

then $\text{do}(X=x)$ is a hard intervention.

‘Hard’ interventions

Definition 5, Hard Intervention:

Given $x=g(z)$ for some $Z \subset V$ in $\text{do}(X=x)$,

if x cuts off the dependency of X to $\text{pa}(X)$,

then $\text{do}(X=x)$ is a hard intervention.

Ex.: $\text{do}(X=4)$ fixes the value of X to 4

Identifying Markovian SCM

Theorem 2, Truncated Factorization^{*}:

* also known G-computation or manipulation theorem

Identifying Markovian SCM

Theorem 2, Truncated Factorization*:

* also known G-computation or manipulation theorem

Let V denote all variables,

then $P(v \mid do(x)) = \prod_{V_i \in V \setminus x} P(v_i \mid pa(v_i))$ for all v consistent with x .

Identifying Markovian SCM

Theorem 2, Truncated Factorization*:

* also known G-computation or manipulation theorem

Let V denote all variables,

then $P(v \mid \text{do}(x)) = \prod_{V_i \in V \setminus X} P(v_i \mid \text{pa}(v_i))$ for all v consistent with x .

If $Y \subset V$, then $P(y \mid \text{do}(x)) = \sum_{V \setminus (Y \cup X)} \prod_{V_i \in V \setminus X} P(v_i \mid \text{pa}(v_i))$.

Identifying Markovian SCM

Theorem 2, Truncated Factorization*:

* also known G-computation or manipulation theorem

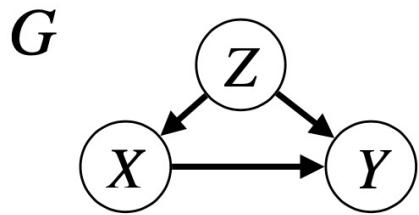
Let V denote all variables,

then $P(v | \text{do}(x)) = \prod_{V_i \in V \setminus X} P(v_i | \text{pa}(v_i))$ for all v consistent with x .

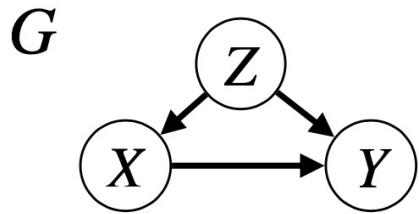
If $Y \subset V$, then $P(y | \text{do}(x)) = \sum_{V \setminus (Y \cup X)} \prod_{V_i \in V \setminus X} P(v_i | \text{pa}(v_i))$.

Ex.: $v := (x=0)$ is inconsistent with $\text{do}(X=1)$

Intuition: Graphical Decomposition



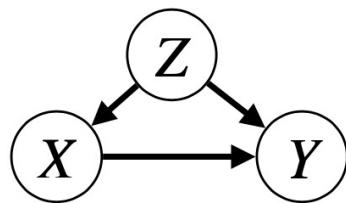
Intuition: Graphical Decomposition



$$P(x, y, z) = P(z)P(x | z)P(y | x, z)$$

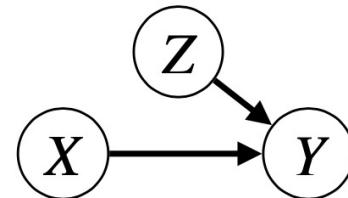
Intuition: Graphical Decomposition

G



$do(X = x)$

$G_{\bar{X}}$

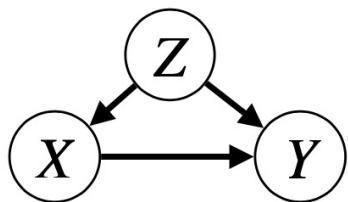


$$P(x, y, z) = P(z)P(x | z)P(y | x, z)$$

$$P(y, z | do(x)) = P(z)P(y | x, z)$$

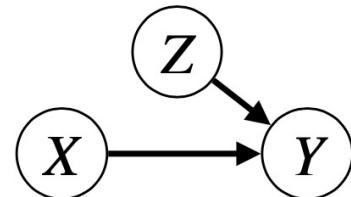
Intuition: Graphical Decomposition

G



$do(X = x)$

$G_{\bar{X}}$



$$P(x, y, z) = P(z)P(x | z)P(y | x, z)$$

$$P(y, z | do(x)) = P(z)P(y | x, z)$$

Theorem 2

gives

$$P(y | do(x)) = \sum_z P(z) \cdot P(y | x, z)$$

Switching the approach to identification again

Adjustment Sets,
Truncated Factorization and Pearl's do-Calculus



Machines Climbing Pearl's Ladder of Causation

Teaching the Causal Calculus on Twitter



Andrew Heiss, geriatric millennial @andrewheiss · Apr 20, 2020

Naive question: are backdoor and front door adjustment in DAG world just specific applications of the 3 rules of do-calculus, or are they a separate thing?

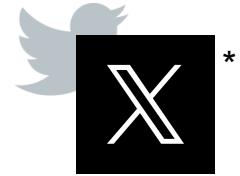
[Show this thread](#)



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Teaching the Causal Calculus on Twitter



Judea Pearl ✅
@yudapearl

...

Naive questions are everyone's questions. Ans. The backdoor and frontdoor are logical consequences of do-calculus. So why do we decorate them with names? Because they are easily recognizable in the DAG, so we store them explicitly in our arsenal of tools, so skip re-deriving them



Andrew Heiss, geriatric millennial @andrewheiss · Apr 20, 2020

Naive question: are backdoor and front door adjustment in DAG world just specific applications of the 3 rules of do-calculus, or are they a separate thing?

Show this thread

* Off-topic:
Elon probably somewhere
“put that bird back in its cage”



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Identifying Anything Identifiable

Theorem 3, do-Calculus: For disjoint variable sets X, Y, Z, W we have

Identifying Anything Identifiable

Theorem 3, do-Calculus: For disjoint variable sets X, Y, Z, W we have

Rule 1: Invariance to Observation

$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_W$$

G^*_W Graph with incoming arrows deleted, G_W^* vice versa for outgoing

Identifying Anything Identifiable

Theorem 3, do-Calculus: For disjoint variable sets X, Y, Z, W we have

Rule 1: Invariance to Observation

$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_W$$

Rule 2: Invariance to Intervention

$$P(Y | \text{do}(W), Z, \text{do}(X)) = P(Y | \text{do}(W), Z, X) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_{W, X^*}$$

G^*_W Graph with incoming arrows deleted, G_{W^*} vice versa for outgoing

Identifying Anything Identifiable

Theorem 3, do-Calculus: For disjoint variable sets X, Y, Z, W we have

Rule 1: Invariance to Observation

$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_W$$

Rule 2: Invariance to Intervention

$$P(Y | \text{do}(W), Z, \text{do}(X)) = P(Y | \text{do}(W), Z, X) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_{W, X^*}$$

Rule 3: Equivalence between Obs.-Interv.

$$P(Y | \text{do}(W), Z, \text{do}(X)) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G^*_{W, X(Z)}$$

G^*_W Graph with incoming arrows deleted, G_{W^*} vice versa for outgoing

Example Derivation: Confounded Frontdoor Graph

$$P(Y | \text{do}(X))$$

= ...

Example Derivation: Confounded Frontdoor Graph

$$P(Y | \text{do}(X))$$

$$= \sum_M P(Y | \text{do}(X), M) \cdot P(M | \text{do}(X))$$

axioms of prob.

Example Derivation: Confounded Frontdoor Graph

$$P(Y | \text{do}(X))$$

$$= \sum_M P(Y | \text{do}(X), M) \cdot P(M | \text{do}(X)) \quad \text{axioms of prob.}$$

$$= \sum_M P(Y | \text{do}(X), \text{do}(M)) \cdot P(M | \text{do}(X)) \quad \text{Rule 2}$$

Example Derivation: Confounded Frontdoor Graph

$$P(Y | \text{do}(X))$$

$$= \sum_M P(Y | \text{do}(X), M) \cdot P(M | \text{do}(X)) \quad \text{axioms of prob.}$$

$$= \sum_M P(Y | \text{do}(X), \text{do}(M)) \cdot P(M | \text{do}(X)) \quad \text{Rule 2}$$

$$= \sum_M P(Y | \text{do}(X), \text{do}(M)) \cdot P(M | X) \quad \text{Rule 2}$$

Example Derivation: Confounded Frontdoor Graph

$$P(Y | \text{do}(X))$$

$$= \sum_M P(Y | \text{do}(X), M) \cdot P(M | \text{do}(X)) \quad \text{axioms of prob.}$$

$$= \sum_M P(Y | \text{do}(X), \text{do}(M)) \cdot P(M | \text{do}(X)) \quad \text{Rule 2}$$

$$= \sum_M P(Y | \text{do}(X), \text{do}(M)) \cdot P(M | X) \quad \text{Rule 2}$$

$$= \sum_M P(Y | \text{do}(M)) \cdot P(M | X) \quad \text{Rule 3}$$

$$= \dots = \sum_{X'} \sum_M P(Y | M, X') \cdot P(X' | M) \cdot P(M | X) \quad (\dots)$$

Intuition: Graphical Example for Rule I

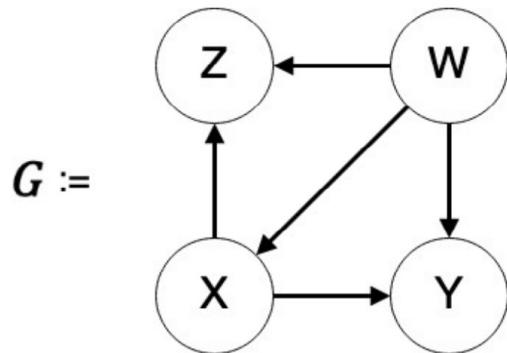
Rule I: Invariance to Observation

$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G_{*W}$$

Intuition: Graphical Example for Rule 1

Rule 1: Invariance to Observation

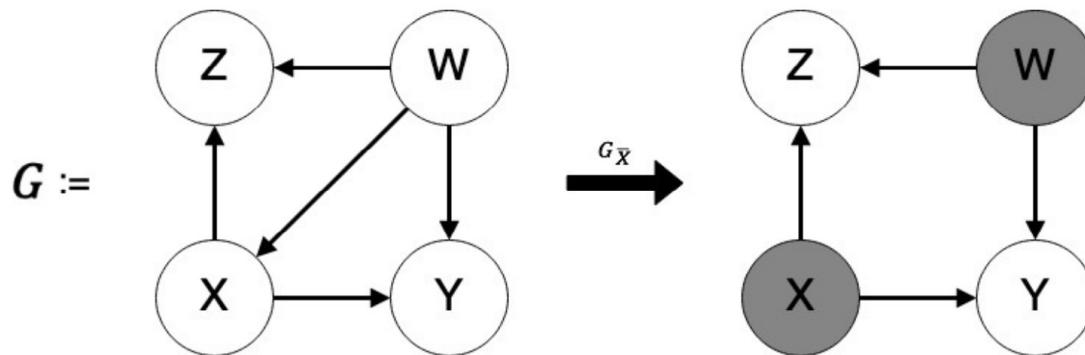
$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G_{*W}$$



Intuition: Graphical Example for Rule I

Rule I: Invariance to Observation

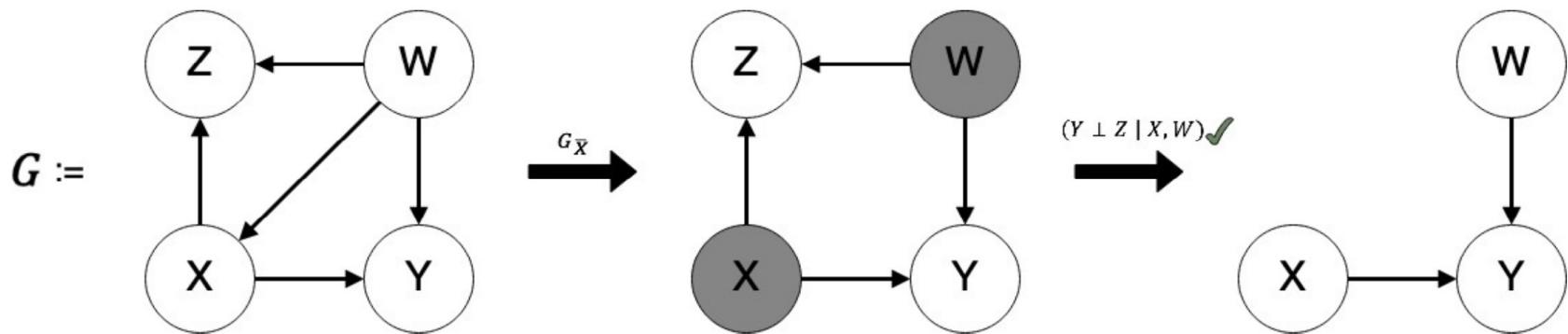
$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G_{*W}$$



Intuition: Graphical Example for Rule 1

Rule 1: Invariance to Observation

$$P(Y | \text{do}(W), Z, X) = P(Y | \text{do}(W), Z) \quad \text{if } (Y \perp X | Z, W) \text{ in } G_{*W}$$



Switching the approach to **partial** identification

Adjustment Sets,
Truncated Factorization and Pearl's do-Calculus

Motivation behind Partial Identification

- Regular identification (a-c) requires conditions such as **sufficiency**

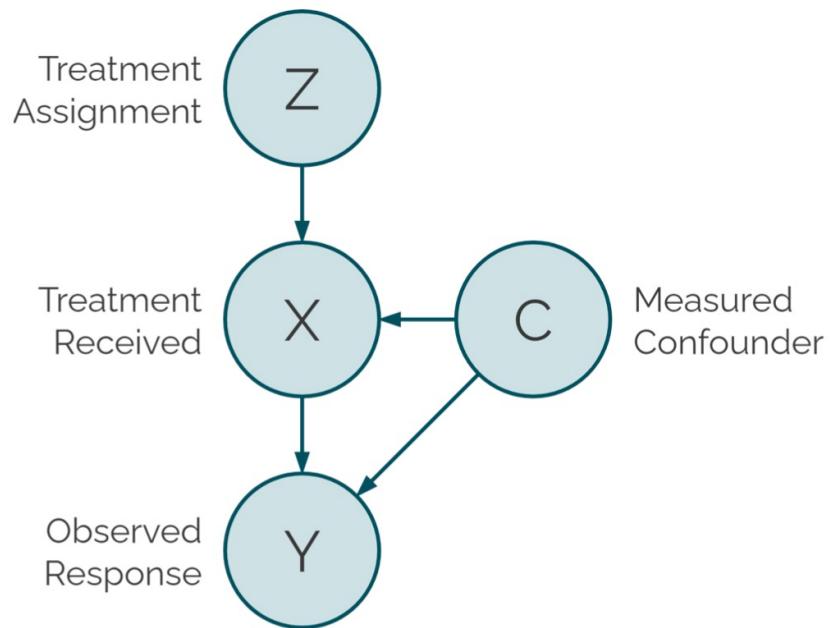
Motivation behind Partial Identification

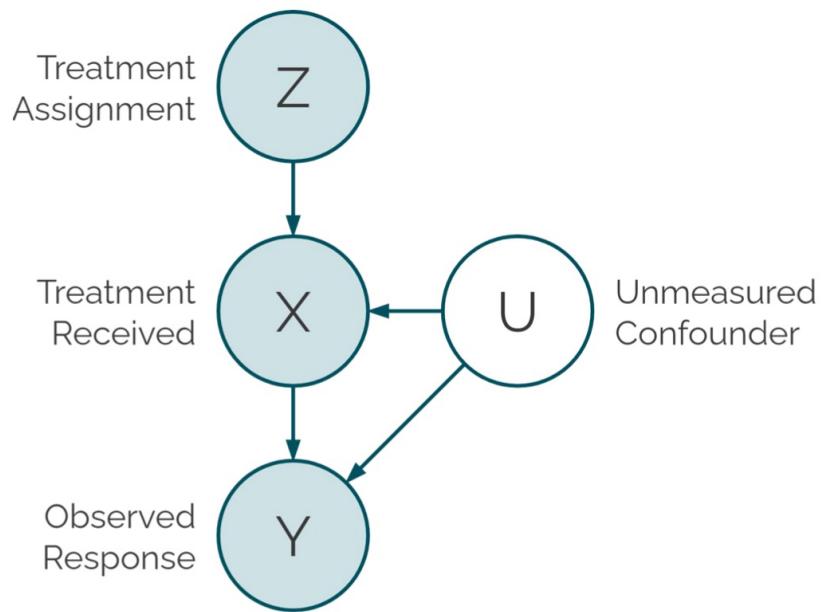
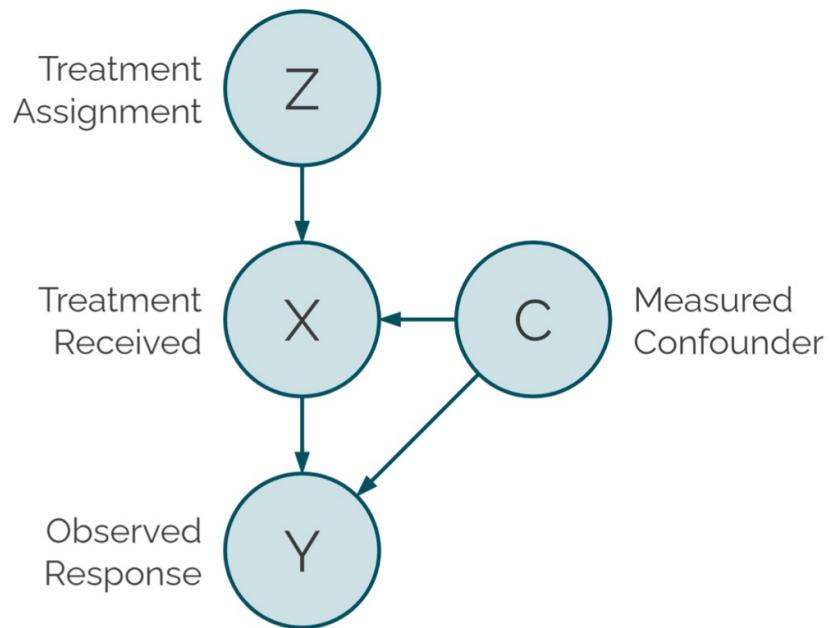
- Regular identification (a-c) requires conditions such as **sufficiency**
- Example (Medical): M.D. is already happy when knowing that the causal effect (e.g. ATE) of a particular drug is positive

Motivation behind Partial Identification

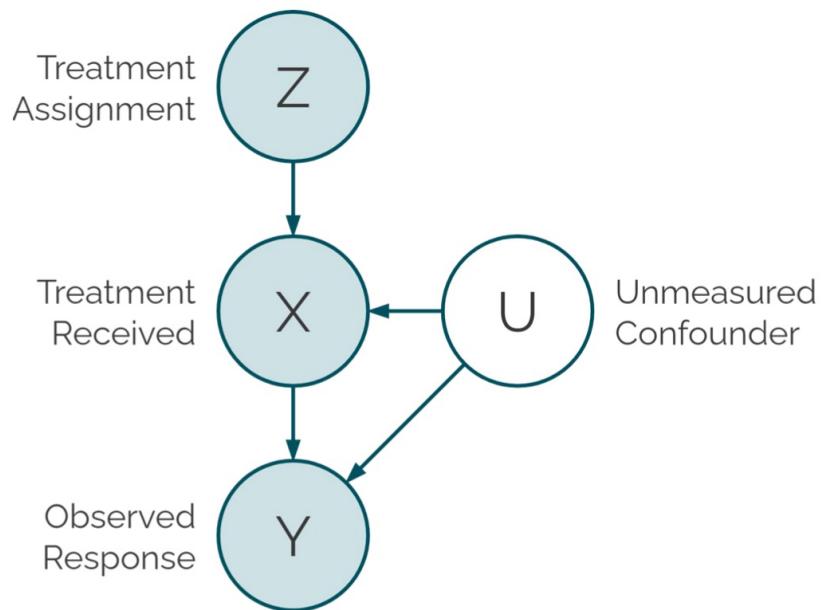
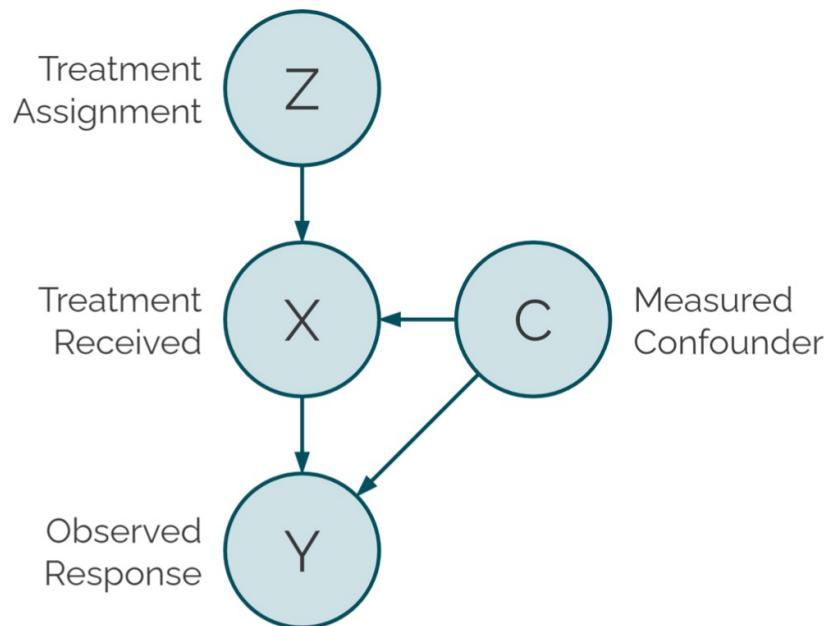
- Regular identification (a-c) requires conditions such as **sufficiency**
- Example (Medical): M.D. is already happy when knowing that the causal effect (e.g. ATE) of a particular drug is positive

Qualitative vs. Quantitative insight





the 'standard' real-world situation



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Causal Relations as Enumerable Function Sets

Definition 6, Response Function Variable:

$\text{Dom}(R_V) := \{I, \dots, N_v\}$ where $N_v := |V|^{|\text{pa}(V)|}$, then R_V is called a RFV.

Causal Relations as Enumerable Function Sets

Definition 6, Response Function Variable:

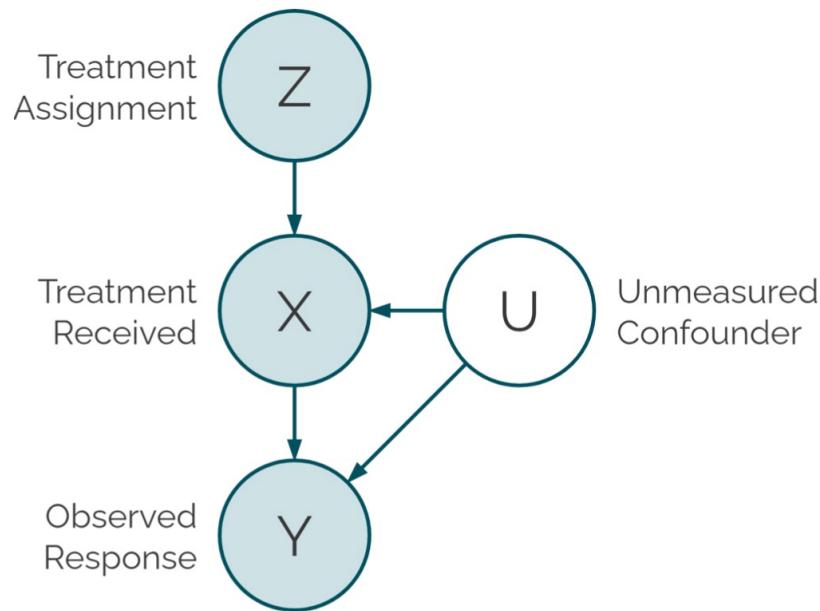
$\text{Dom}(R_V) := \{I, \dots, N_v\}$ where $N_v := |V|^{|\text{pa}(V)|}$, then R_V is called a RFV.

Theorem 4, RFV Distribution:

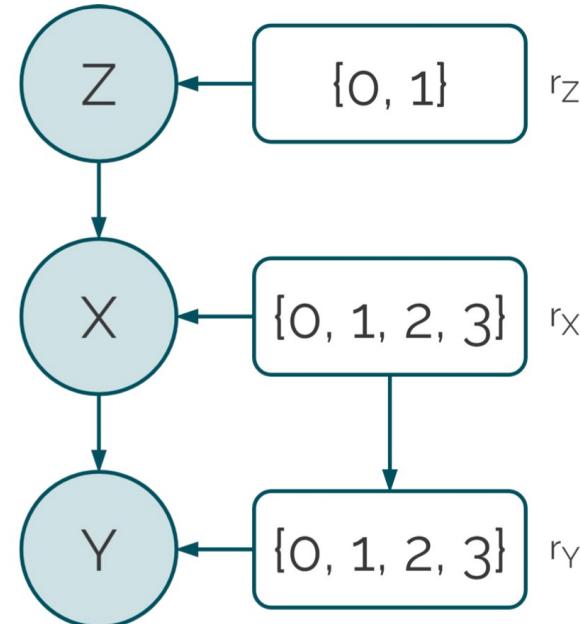
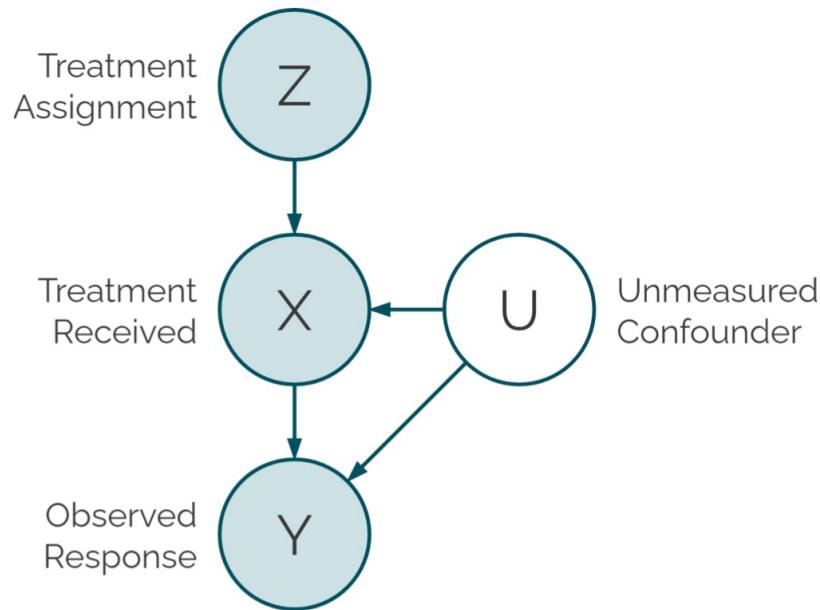
Furthermore, the observational distribution $P(V)$ can be

rewritten as: $\sum_r P(r) \cdot \prod_{V_i \in V} I(V_i, \text{pa}(V_i), r_{V_i})$.

Converting the Exogenous Terms

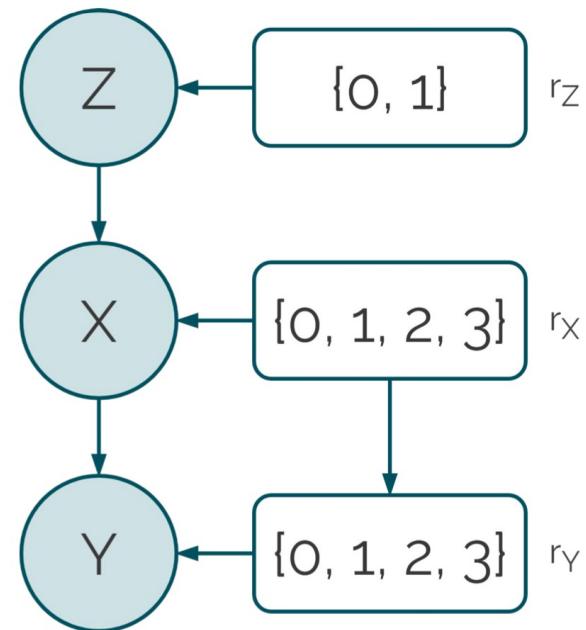


Converting the Exogenous Terms to RFVs



$P(Z, X, Y)$ written as $P(R_Z, R_X, R_Y)$

$P(\mathbf{v})$
Observed data $P(\mathbf{v})$

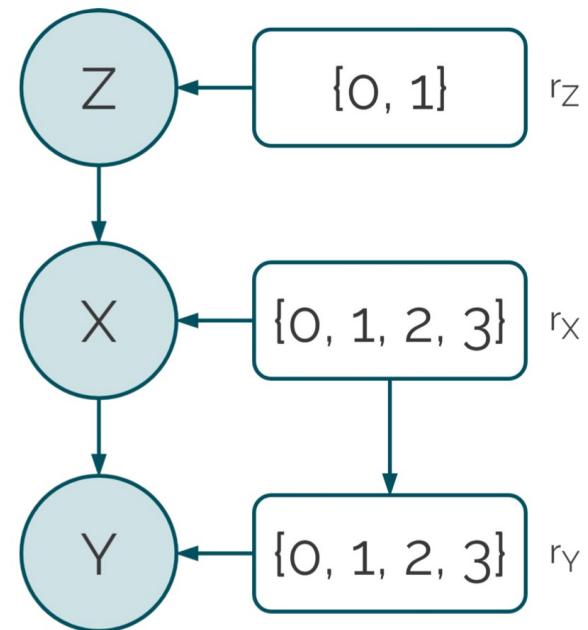


$P(Z, X, Y)$ written as $P(R_Z, R_X, R_Y)$

$$P(\mathbf{v}) = \sum_{\mathbf{r}} P(\mathbf{r}) \prod_{V \in \mathbf{v}} \mathbb{I}(V; \text{pa}_V, r_V)$$

Observed data $P(\mathbf{v})$

Latent model $P(\mathbf{r})$



$P(Z, X, Y)$ written as $P(R_Z, R_X, R_Y)$

$$P(\mathbf{v}) = \sum_{\mathbf{r}} P(\mathbf{r}) \prod_{V \in \mathbf{v}} \mathbb{I}(V; \mathbf{pa}_V, r_V)$$

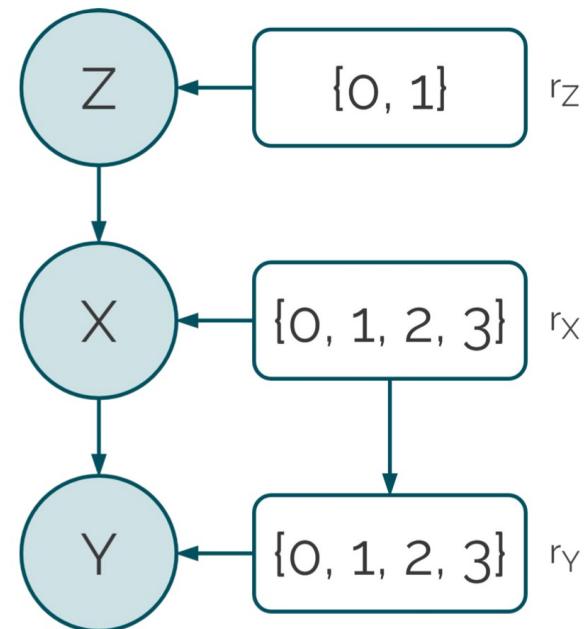
Observed data $P(\mathbf{v})$

Latent model $P(\mathbf{r})$

E.g.:

$$P(\mathbf{v} = (0, 1, 1)) = P(\mathbf{r} = (0, 1, 0)) + P(\mathbf{r} = (0, 3, 1)) + P(\mathbf{r} = (0, 1, 0)) + P(\mathbf{r} = (0, 3, 1))$$

Condition accounts for **consistent states**



Bound on Causal Effect via Min-Max of LP

min/max

Objective function (e.g. Average Causal Effect (ACE))

Bound on Causal Effect via Min-Max of LP

min/max

Objective function (e.g. Average Causal Effect (ACE))

s.t.

Observed Data $P(v) = \text{Latent Model } P(r)$

$$\sum_r P(r) = 1, \quad P(r) \geq 0$$

Bound on Causal Effect via Min-Max of LP

min/max

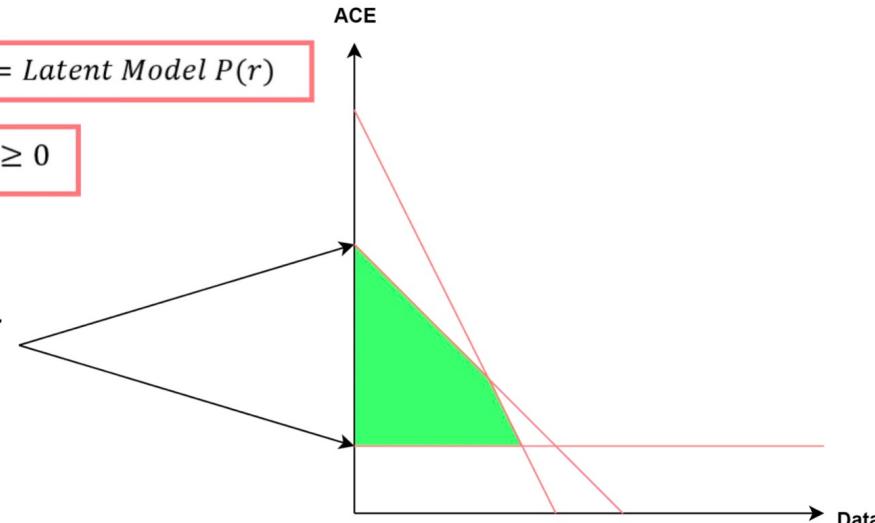
Objective function (e.g. Average Causal Effect (ACE))

s.t.

Observed Data $P(v) = \text{Latent Model } P(r)$

$\sum_r P(r) = 1, \quad P(r) \geq 0$

Higher and Lower
Bound

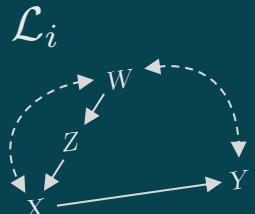




4

Estimation using Machine Learning

Two examples: iSPN & NCM



Machines Climbing Pearl's Ladder of Causation

Individuals (Medical Example: Patient Records)



Numerical Representation

A	(62	18	24	...	21	...)
F		48	60	20	...	32	...	
H		34	90	40	...	64	...	
M		39	75	37	...	66	...	
										

$A = \text{Age}$

$F = \text{Food Habits (or Nutrition)}$

$H = \text{Health}$

$M = \text{Mobility}$

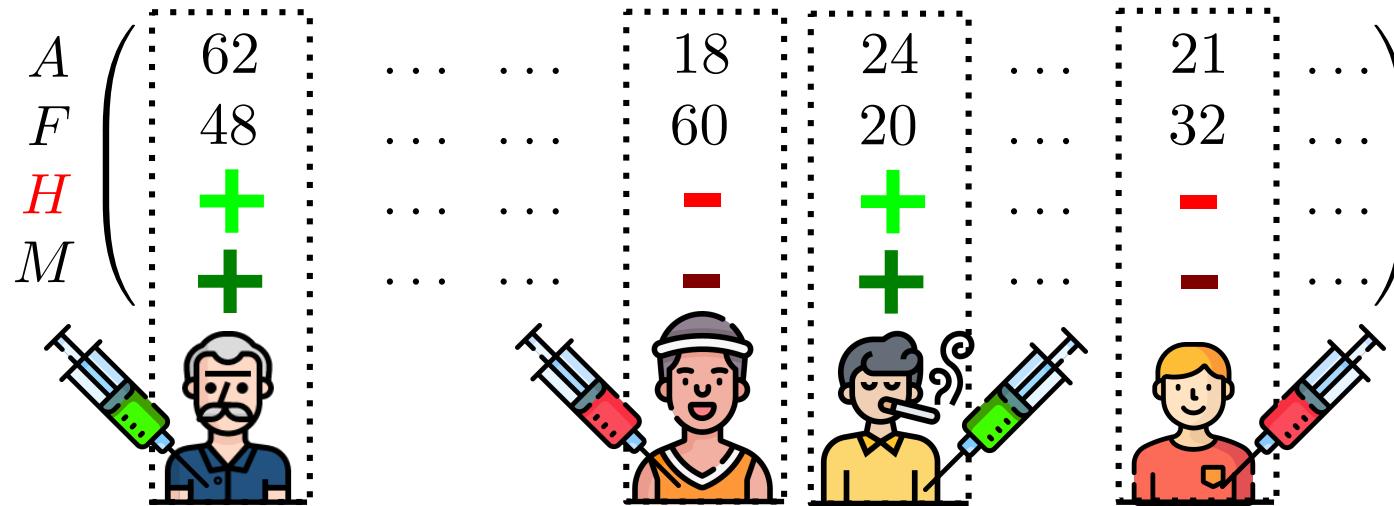


TECHNISCHE
UNIVERSITÄT
DARMSTADT



THE UNIVERSITY
OF TEXAS AT DALLAS

Intervention (e.g. a Vaccine)



$A = \text{Age}$

$F = \text{Food Habits (or Nutrition)}$

$H = \text{Health}$

$M = \text{Mobility}$

Structural Causal Model

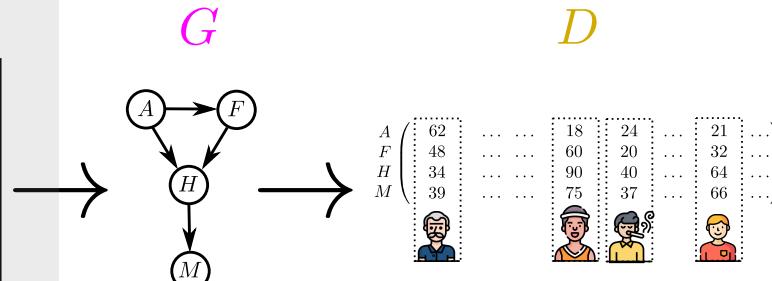
$$A = U(0, 100)$$

$$F = \frac{1}{2}A + \mathcal{N}(10, 10)$$

$$H = \frac{1}{100}(100 - A^2) + \frac{1}{2}F + \mathcal{N}(40, 30)$$

$$M = \frac{1}{2}H + \mathcal{N}(20, 10)$$

(A)ge
(F)ood Habits
(H)ealth
(M)obility



$A = \text{Age}$ $F = \text{Food Habits (or Nutrition)}$ $H = \text{Health}$ $M = \text{Mobility}$

Structural Causal Model

$$A = U(0, 100)$$

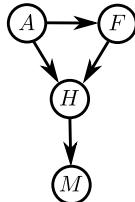
$$F = \frac{1}{2}A + \mathcal{N}(10, 10)$$

$$H = \frac{1}{100}(100 - A^2) + \frac{1}{2}F + \mathcal{N}(40, 30)$$

$$M = \frac{1}{2}H + \mathcal{N}(20, 10)$$

(A)ge
(F)ood Habits
(H)ealth
(M)obility

G



D

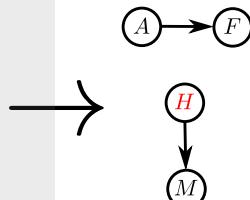
A	62	...	18	24	...	21	...
F	48	...	60	20	...	32	...
H	34	...	90	40	...	64	...
M	39	...	75	37	...	66	...

$$A = U(0, 100)$$

$$F = \frac{1}{2}A + \mathcal{N}(10, 10)$$

$$H = U(0, 100)$$

$$M = \frac{1}{2}H + \mathcal{N}(20, 10)$$



A	62	...	18	24	...	21	...
F	48	...	60	20	...	32	...
H	+	+	-	+	-	+	-
M	62	...	18	24	...	21	...

A = Age

F = Food Habits (or Nutrition)

H = Health

M = Mobility



TECHNISCHE
UNIVERSITÄT
DARMSTADT



THE UNIVERSITY
OF TEXAS AT DALLAS

Interventional Sum-Product Network

Definition 7, iSPN:

$$\text{A model } M(G, D) = g_{\Theta}(D, f_{\Psi}(G))$$

Interventional Sum-Product Network

Definition 7, iSPN:

A model $M(G, D) = g_\Theta(D, f_\Psi(G))$ where D is a collection of data sets containing experimental distributions

Interventional Sum-Product Network

Definition 7, iSPN:

A model $M(G, D) = g_\Theta(D, f_\Psi(G))$ where D is a collection of data sets containing experimental distributions & sub-model g_Θ an SPN whose leaf and sum nodes are parameterized by MLP f_Ψ which takes causal graph G as input

Interventional Sum-Product Network

Definition 7, iSPN:

A model $M(G, D) = g_\Theta(D, f_\Psi(G))$ where D is a collection of data sets containing experimental distributions & sub-model g_Θ an SPN whose leaf and sum nodes are parameterized by MLP f_Ψ , which takes causal graph G as input, is called an iSPN.

Interventional Sum-Product Network

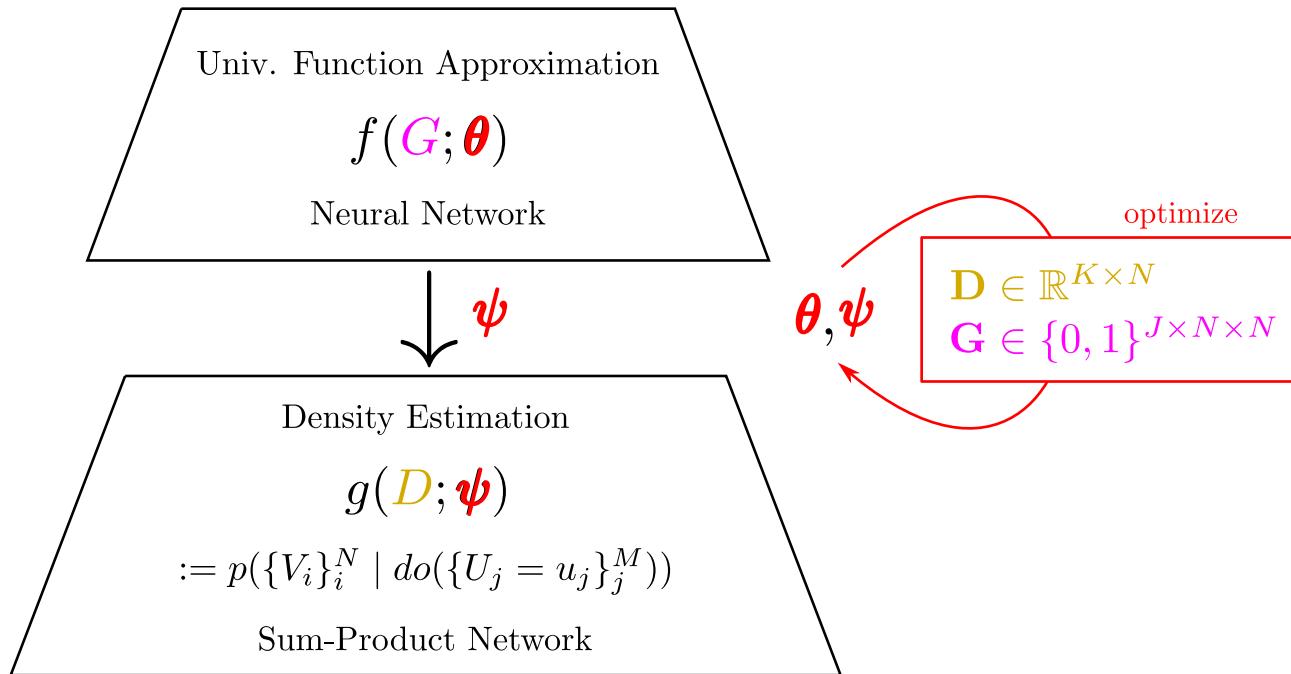
Definition 7, iSPN:

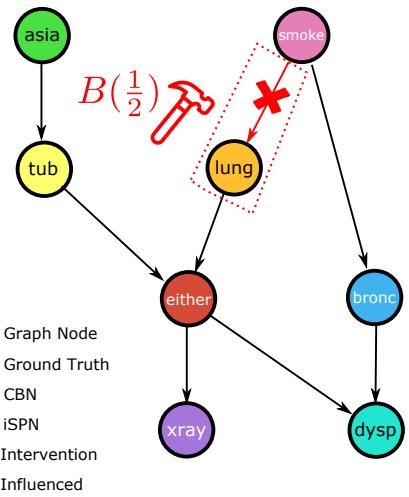
A model $M(G, D) = g_\Theta(D, f_\Psi(G))$ where D is a collection of data sets containing experimental distributions & sub-model g_Θ an SPN whose leaf and sum nodes are parameterized by MLP f_Ψ , which takes causal graph G as input, is called an iSPN.

To learn about SPN (without the ‘i’), check out the other ESSAI course:

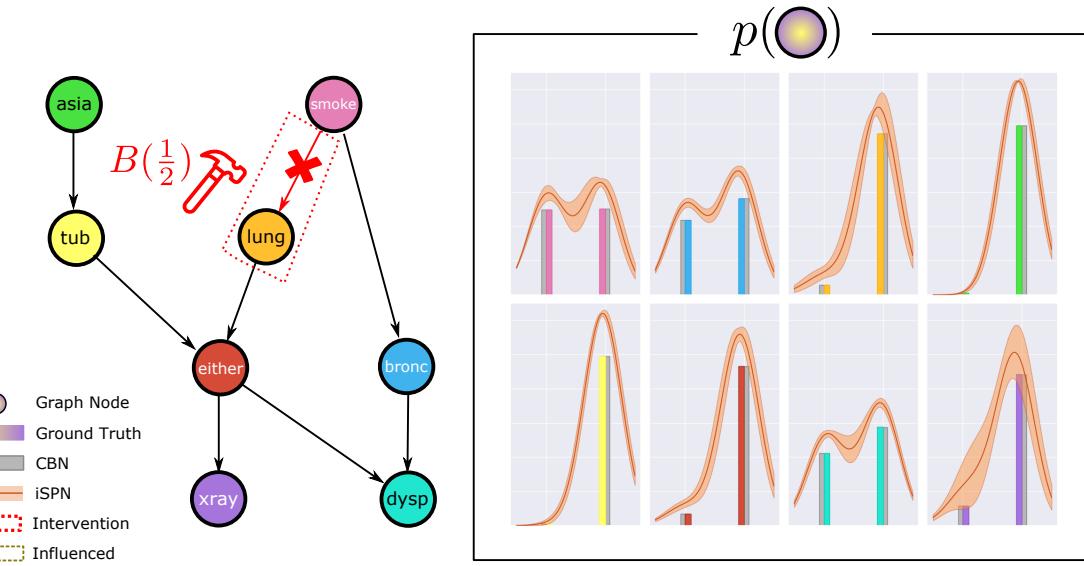
Probabilistic Circuits: Deep Probabilistic Models with Reliable Reasoning
by Robert Peharz and Antonio Vergari

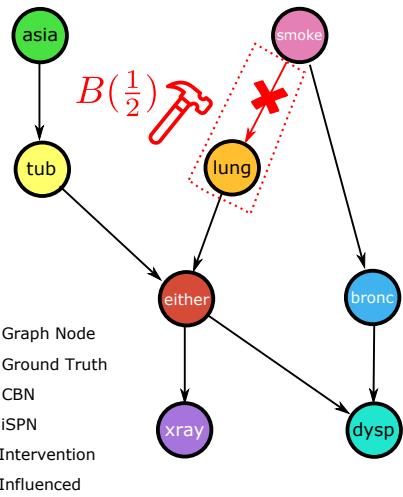
Schematic of the iSPN, a Causal Density Estimator



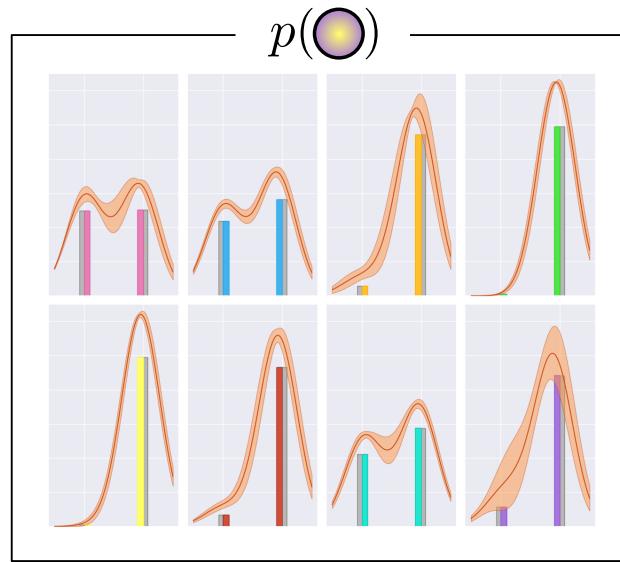


Observational Data

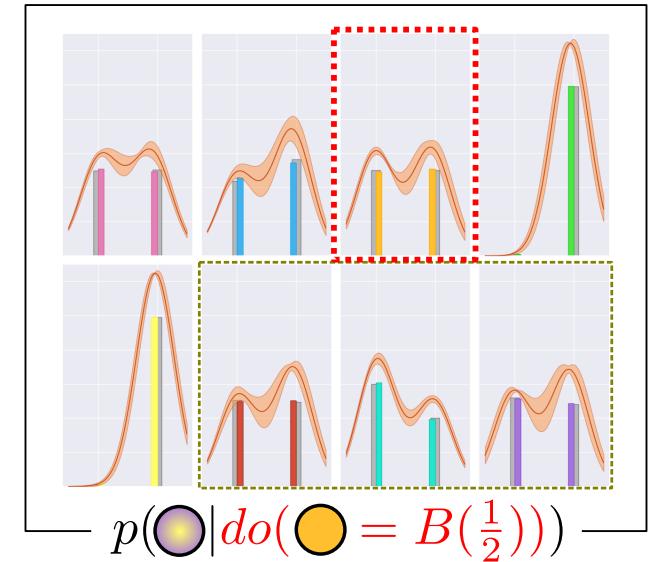




Observational Data



Intervention on 'lung'



Age

Food Habits

Health

Mobility



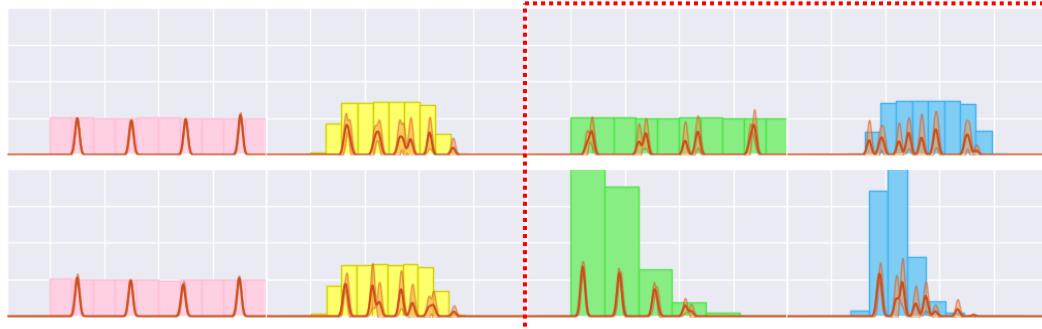
Age

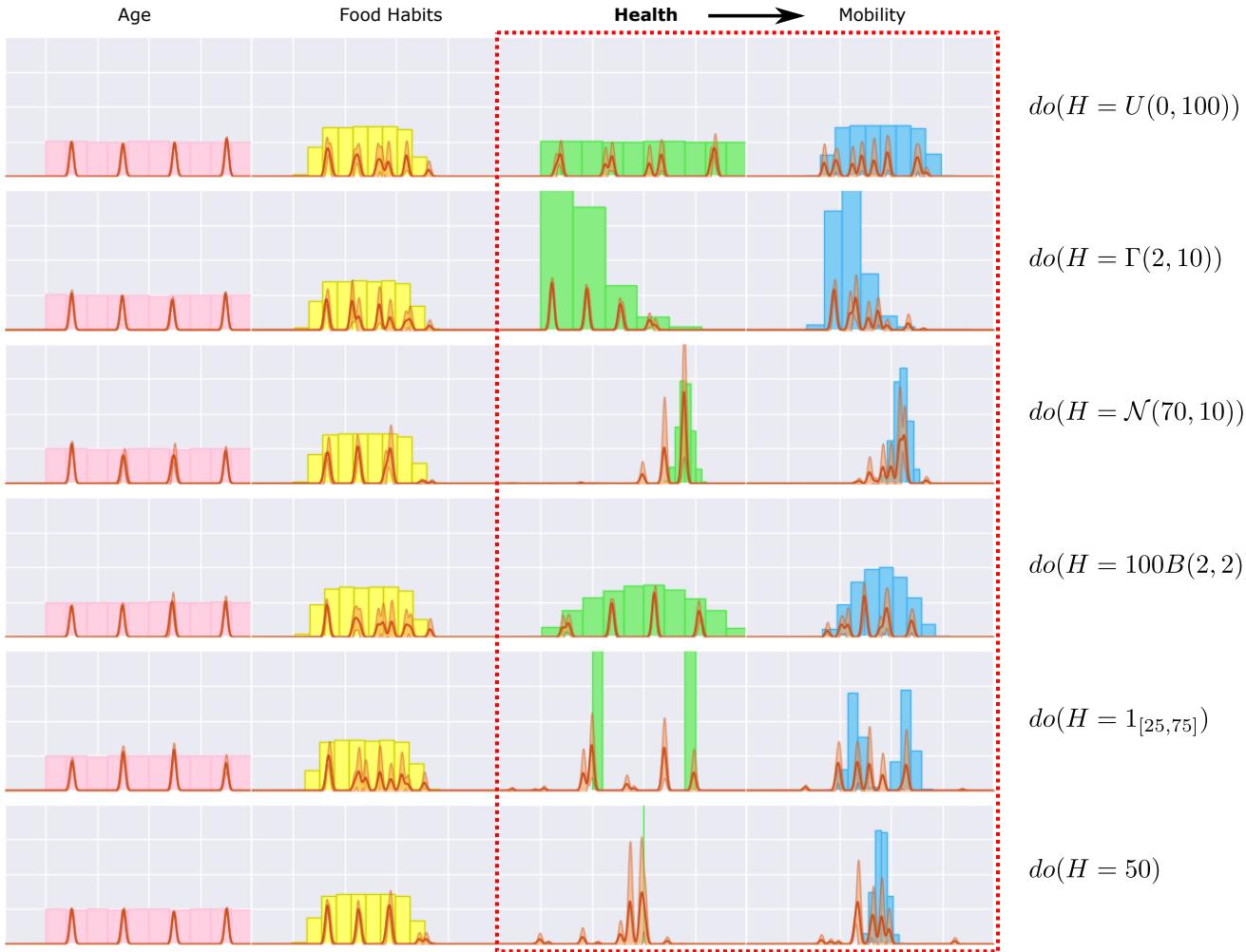
Food Habits

Health



Mobility





Dealing
with
different
types of
interven-

Tractable Inference, a Plus for iSPN

Tractable Inference, a Plus for iSPN

- Cooper 1990, Roth 1996:
Marginal (and conditional) inference
in Probabilistic Graphical Models (PGM),
for instance a Bayesian Network (BN),
is *intractable (#P-hard)*

Tractable Inference, a Plus for iSPN

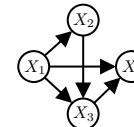
- Cooper 1990, Roth 1996:
Marginal (and conditional) inference
in Probabilistic Graphical Models (PGM),
for instance a Bayesian Network (BN),
is *intractable (#P-hard)*

- Poon & Domingos 2011:
...
in SPN is *tractable* (P)
– linear in the size of the network.

Tractable Inference, a Plus for iSPN

- Cooper 1990, Roth 1996:
Marginal (and conditional) inference
in Probabilistic Graphical Models (PGM),
for instance a Bayesian Network (BN),
is *intractable (#P-hard)*

Semantic



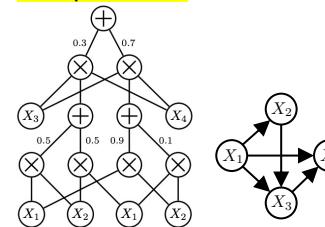
n	$O(2^n)$
1	2
2	4
3	8
4	16
...	...
100	$1,26 \cdot 10^{30}$

- Poon & Domingos 2011:
...
in SPN is *tractable* (P)
– linear in the size of the network.

Tractable Inference, a Plus for iSPN

- Cooper 1990, Roth 1996:
Marginal (and conditional) inference
in Probabilistic Graphical Models (PGM),
for instance a Bayesian Network (BN),
is *intractable (#P-hard)*

Computational Semantic



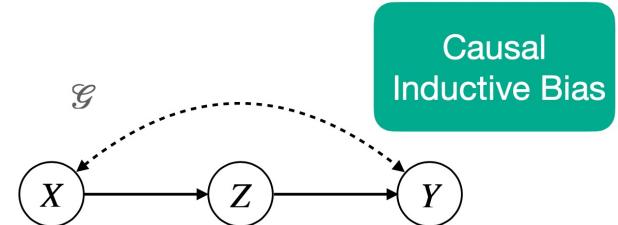
n	$O(n)$	$O(2^n)$
1	1	2
2	2	4
3	3	8
4	4	16
...		
100	100	$1,26 \cdot 10^{30}$

Negative: no implicit identification!

Next: an alternative without linear time inference but identification

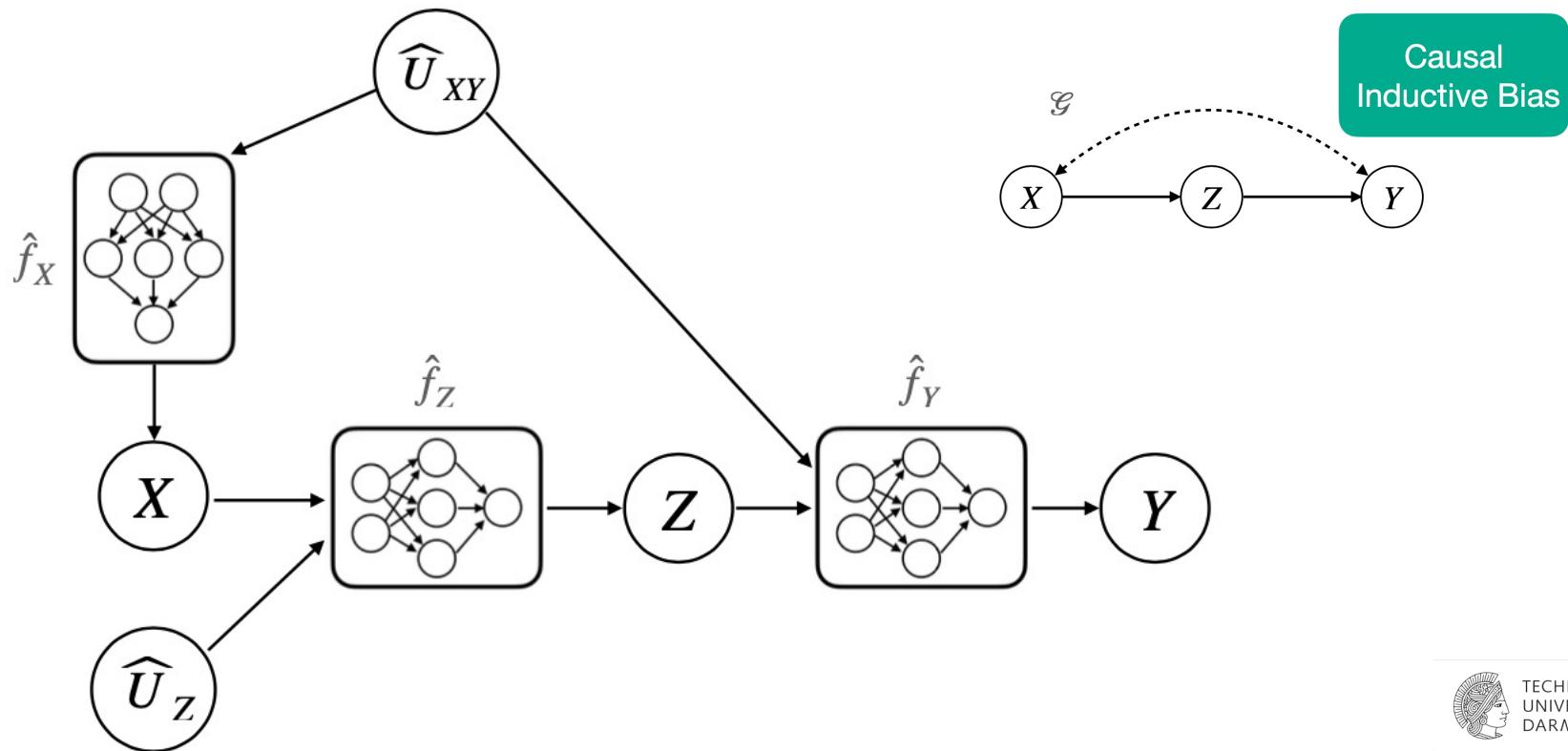
=

+ Causal Graph



Neural Causal Model = Neural Nets + Causal Graph

= parameterized SCM with neural nets mechanisms



Algorithm 1: Identifying/estimating queries with NCMs.

Input : causal query $Q = P(\mathbf{y} \mid do(\mathbf{x}))$, L_1 data $P(\mathbf{v})$, and causal diagram \mathcal{G}

Output : $P^{\mathcal{M}^*}(\mathbf{y} \mid do(\mathbf{x}))$ if identifiable, FAIL otherwise.

```
1  $\widehat{M} \leftarrow \text{NCM}(\mathbf{V}, \mathcal{G})$                                 // from Def. 7
2  $\theta_{\min}^* \leftarrow \arg \min_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
3  $\theta_{\max}^* \leftarrow \arg \max_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
4 if  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x})) \neq P^{\widehat{M}(\theta_{\max}^*)}(\mathbf{y} \mid do(\mathbf{x}))$  then
5   return FAIL
6 else
7   return  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x}))$     // choose min or max
          arbitrarily
```

Algorithm 1: Identifying/estimating queries with NCMs.

Input : causal query $Q = P(\mathbf{y} \mid do(\mathbf{x}))$, L_1 data $P(\mathbf{v})$, and causal diagram \mathcal{G}

Output: $P^{\mathcal{M}^*}(\mathbf{y} \mid do(\mathbf{x}))$ if identifiable, FAIL otherwise.

- 1 $\widehat{M} \leftarrow \text{NCM}(\mathbf{V}, \mathcal{G})$ // from Def. 7
- 2 $\theta_{\min}^* \leftarrow \arg \min_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$ s.t. $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$
- 3 $\theta_{\max}^* \leftarrow \arg \max_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$ s.t. $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$
- 4 **if** $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x})) \neq P^{\widehat{M}(\theta_{\max}^*)}(\mathbf{y} \mid do(\mathbf{x}))$ **then**
- 5 **return** FAIL
- 6 **else**
- 7 **return** $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x}))$ // choose min or max arbitrarily

Min-Max
Optimization subject
to **equality** on
observational
distribution

Algorithm 1: Identifying/estimating queries with NCMs.

Input : causal query $Q = P(\mathbf{y} \mid do(\mathbf{x}))$, L_1 data $P(\mathbf{v})$, and causal diagram \mathcal{G}

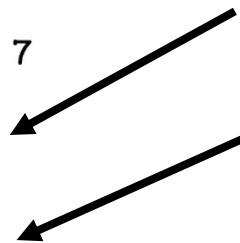
Output: $P^{\mathcal{M}^*}(\mathbf{y} \mid do(\mathbf{x}))$ if identifiable, FAIL otherwise.

```
1  $\widehat{M} \leftarrow \text{NCM}(\mathbf{V}, \mathcal{G})$                                 // from Def. 7
2  $\theta_{\min}^* \leftarrow \arg \min_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
3  $\theta_{\max}^* \leftarrow \arg \max_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
4 if  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x})) \neq P^{\widehat{M}(\theta_{\max}^*)}(\mathbf{y} \mid do(\mathbf{x}))$  then
5   return FAIL
6 else
7   return  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x}))$     // choose min or max
          arbitrarily
```

Min-Max

Optimization subject to **equality** on observational distribution

Query is identifiable iff. the optimizations agree



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Full Stop!

Now: Summarizing what've learned

Recap of ‘New’ Formalizations & Results this Lecture

- Definition 1: Causal Effect Identifiability
- Definition 2: Markovian SCM
- Definition 3: Causal Effect
- Definition 4: Confounding
- Definition 5: Hard Intervention
- Theorem 1: Adjustment Sets

Recap of ‘New’ Formalizations & Results this Lecture

- Theorem 1: Adjustment Sets
- Theorem 2: Truncated Factorization
- Theorem 3: do-Calculus
- Definition 6: Response Function Variable
- Theorem 4: RFV Distribution
- Definition 7: iSPN

Recap of ‘New’ Formalizations & Results this Lecture

- And a lot more informally about

Markovian Identifiability,

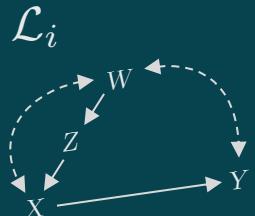
Tractability of Inference,

NCM / Neural Identification,

etc.



Z | Announcements



Want to get a picture of who
the scientists are that do
causality research?

Genealogy of Causality

Access via genealogy.causality.link

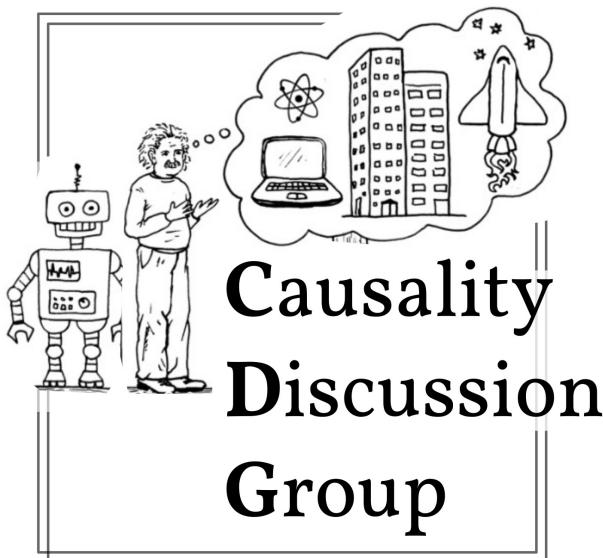


Name	Institution	Supervisor	Location	Previous Positions
UCLA				
Judea Pearl	UCLA	?	US	Rutgers, Technion, New
Wesley Salmon	UCLA	Hans Reichenbach	US	?
Hans Reichenbach	UCLA	Paul Hensel, Max Noeth	US	Berlin, Istanbul, Erlange
John Hopkins				
Ilya Shpitser	John Hopkins		US	UCLA, Judea Pearl
Oregon State University				
Karthika Mohan	Oregon State University	Judea Pearl	US	
CMU				
Kun Zhang	CMU		Pittsburgh, US	MPI Tübingen
Clark Glymour	CMU	Wesley Salmon	Pittsburgh, US	
Peter Spirtes	CMU		Pittsburgh, US	
ETH Zürich				
Peter Bühlmann	ETH		Zürich	?
Marloes Maathuis	ETH		Zürich	?
Nicolai Meinshausen	ETH			
LMU Munich				
Stephan Hartmann	LMU		Munich, Germany	
MPI Tübingen				
Bernhard Schölkopf	MPI Tübingen	Vladimir Vapnik	Tübingen, Germ	TU Berlin
Ulrike von Luxburg	MPI Tübingen		Tübingen, Germany	
Michel Besserve				

Want to discuss more
following this talk?

Every Week with Paper Authors

→ **Discuss LIVE**



530 members
on Slack

Join the community
via
discuss.causality.link

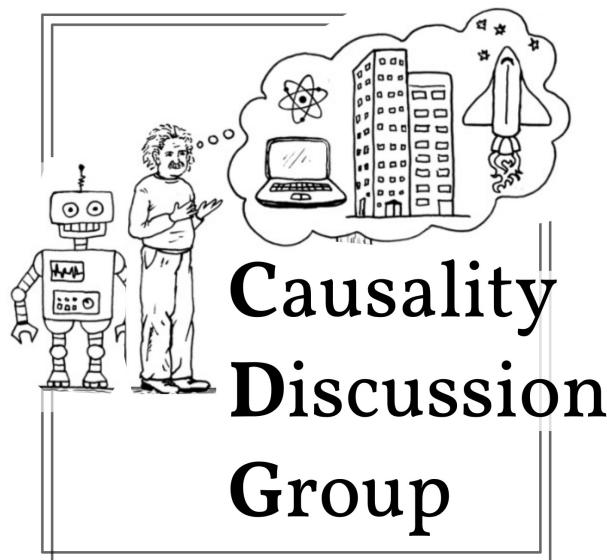
Past Sessions: [Password: Causality, Direct Access Link]

- ▷ Session 01.03.2023 | **Deep Counterfactual Estimation with Categorical Background Variables** | Discussant: Edward De Brouwer
- ▷ Session 22.02.2023 | **Information-Theoretic Causal Discovery and Intervention Detection over Multiple Environments** | Discussant: Osman Ali Mian
- ▷ Session 08.02.2023 | **CLEAR: Generative Counterfactual Explanations on Graphs** | Discussants: Jing Ma, Ruocheng Guo
- ▷ Session 01.02.2023 | **Causal Transformer for Estimating Counterfactual Outcomes** | Discussant: Valentyn Melnychuk
- ▷ Session 25.01.2023 | **Abstracting Causal Models** | Discussant: Sander Beckers
- ▷ Session 18.01.2023 | **Desiderata for Representation Learning: A Causal Perspective** | Discussant: Yixin Wang
- ▷ Session 11.01.2023 | **Causal Feature Selection via Orthogonal Search** | Discussant: Ashkan Soleymani
- ▷ Session 14.11.2022 | **Rewind 2022** | Final session of 2022 to simply rewind on what we experienced throughout the year
- ▷ Session 07.12.2022 | **Causal Inference Through the Structural Causal Marginal Problem** | Discussant: Luigi Gresele
- ▷ Session 30.11.2022 | **Selecting Data Augmentation for Simulating Interventions** | Discussant: Maximilian Ilse
- ▷ Session 23.11.2022 | **On Disentangled Representations Learned from Correlated Data** | Discussant: Frederik Träuble
- ▷ Session 16.11.2022 | **Causal Curiosity: RL Agents Discovering Self-supervised Experiments for Causal Repr. Learning** | Discussant: Sumedh Sontakke
- ▷ Session 09.11.2022 | **Causal Machine Learning: A Survey and Open Problems** | Discussants: Jean Kaddour, Aengus Lynch
- ▷ Session 02.11.2022 | **A Critical Look at the Consistency of Causal Estimation with Deep Latent Variable Models** | Discussant: Severi Rissanen
- ▷ Session 26.10.2022 | **Nonlinear Invariant Risk Minimization: A Causal Approach** | Discussant: Chaochao Lu
- ▷ Session 19.10.2022 | **CausalVAE: Disentangled Representation Learning via Neural Structural Causal Models** | Discussant: Mengyue Yang
- ▷ Session 12.10.2022 | **Weakly Supervised Causal Representation Learning** | Discussant: Johann Brehmer
- ▷ Session 05.10.2022 | **Towards Causal Representation Learning** | Discussant: Anirudh Goyal
- ▷ Session 21.09.2022 | **Selection Collider Bias in Large Language Models** | Discussant: Emily McMillin
- ▷ Session 14.09.2022 | **The Causal-Neural Connection: Expressiveness, Learnability, and Inference** | Discussants: Kai-Zhan Lee, Kevin Xia
- ▷ Session 07.09.2022 | **Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style** | Discussant: Julius von Kügelgen

35+ Sessions
Completed
and
All Recorded

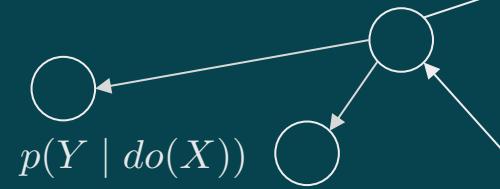
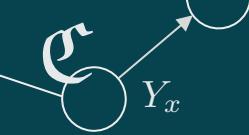


Join the community
via discuss.causality.link



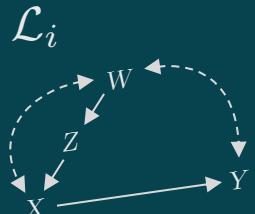
Access the genealogy
via genealogy.causality.link

Genealogy of Causality				
Name	University	Advisors	Current Position	Previous Positions
UCLA				Rutgers, Technion, New
Judea Pearl				?
Wesley Salmon	UCLA	Hans Reichenbach	US	Berlin, Istanbul, Erlange
Hans Reichenbach	UCLA	Paul Hensel, Max Noeth	US	
John Hopkins				
Ilya Shpitser	John Hopkins		US	UCLA, Judea Pearl
Oregon State University				
Karthika Mohan	Oregon State University	Judea Pearl	US	
CMU				
Kun Zhang	CMU		Pittsburgh, US	MPI Tübingen
Clark Glymour	CMU	Wesley Salmon	Pittsburgh, US	
Peter Spirtes	CMU		Pittsburgh, US	
ETH Zürich				
Peter Bühlmann	ETH		Zürich	?
Marloes Maathuis	ETH		Zürich	?
Nicolai Meinshausen	ETH			
LMU Munich				
Stephan Hartmann	LMU		Munich, Germany	
MPI Tübingen				
Bernhard Schölkopf	MPI Tübingen	Vladimir Vapnik	Tübingen, Germ	TU Berlin
Ulrike von Luxburg	MPI Tübingen		Tübingen,	Germany
Michel Besserve				

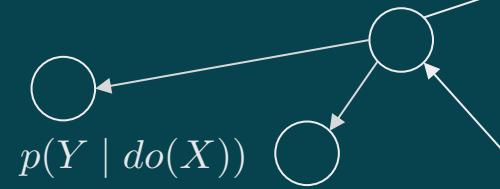
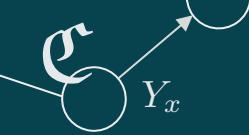


That's a wrap!

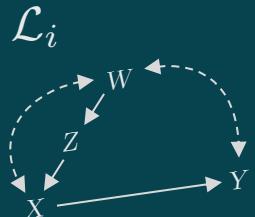
Feel free to reach out:
matej.zecevic@tu-darmstadt.de



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA
Machines Climbing Pearl's Ladder of Causation



B | Backup



Counterfactuals

A 3-step procedure

Abduction : Update belief in P_U given evidence E

Action : Change equations accordingly, $do(X = x)$

Prediction : Look at variable of interest $P(Y = y)$