IDSIA SUPSI

Structural Causal Models are Credal Networks?

joint work with Marco Zaffalon (IDSIA) and Rafael Cabañas (UAL)

Alessandro Antonucci (alessandro@idsia.ch)
Senior Lecturer Researcher IDSIA USI-SUPSI

slides

SIPTA Seminar, Virtual Event, Nov 28, 2023

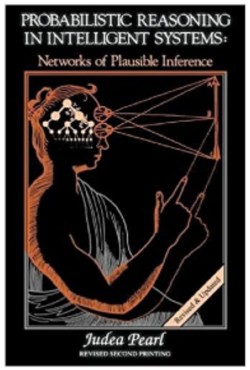
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)

USI/SUPSI
Istituto
Dalle Molle
di studi
sull'intelligenza
artificiale
IDSIA

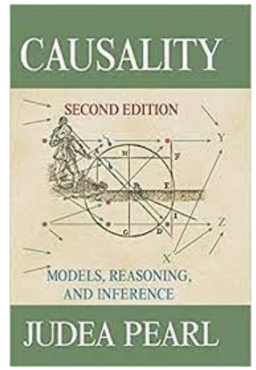
- World-class Research Institute on AI founded in 1988 in Lugano
- Affiliated with both University of Lugano (USI) and University of Applied Sciences of Southern Switzerland (SUPSI)
- Staff ~100 people + 50 PhD

Angelo Dalle Molle (1908 - 2002)

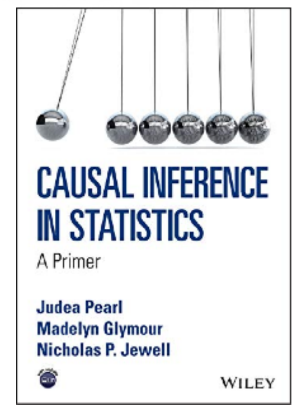
Framing the Topic

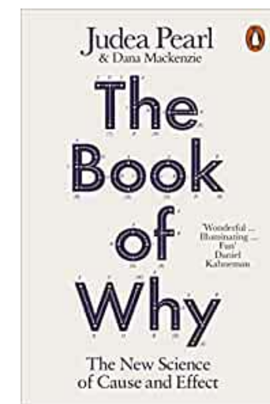


Bayesian Nets
(≤ 1988)



Do Calculus
(≤ 2000)



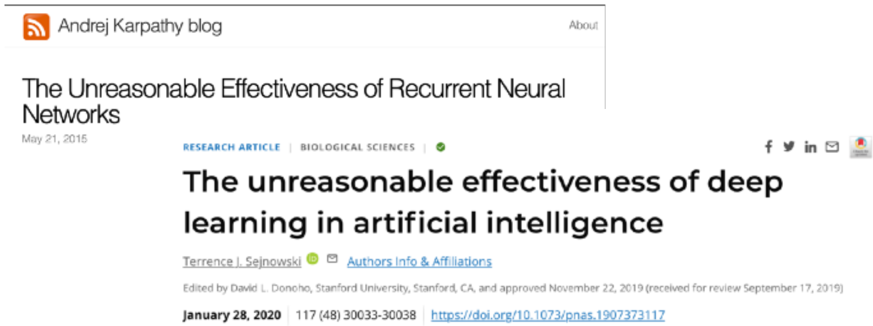


Structural Causal Models
(≤ 2016)

This Talk

Credal
Nets
(≤ 2000)

(Science >) AI > Deep Learning


Andrej Karpathy blog
About
The Unreasonable Effectiveness of Recurrent Neural Networks
May 21, 2015

RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

The unreasonable effectiveness of deep learning in artificial intelligence

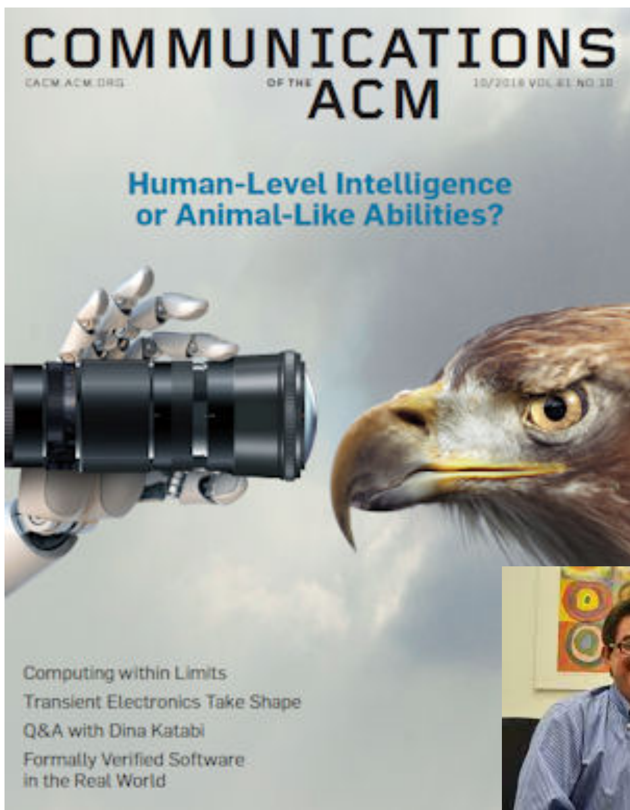
Terrence J. Sejnowski Authors Info & Affiliations

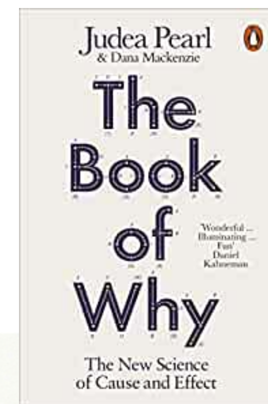
Edited by David L. Donoho, Stanford University, Stanford, CA, and approved November 22, 2019 (received for review September 17, 2019)

January 28, 2020 | 117 (48) 30033-30038 | https://doi.org/10.1073/pnas.1907373117

"*Deep learning has instead given us machines with truly* ***impressive abilities but no intelligence***.

The difference is profound and lies in the ***absence of a model of reality***."





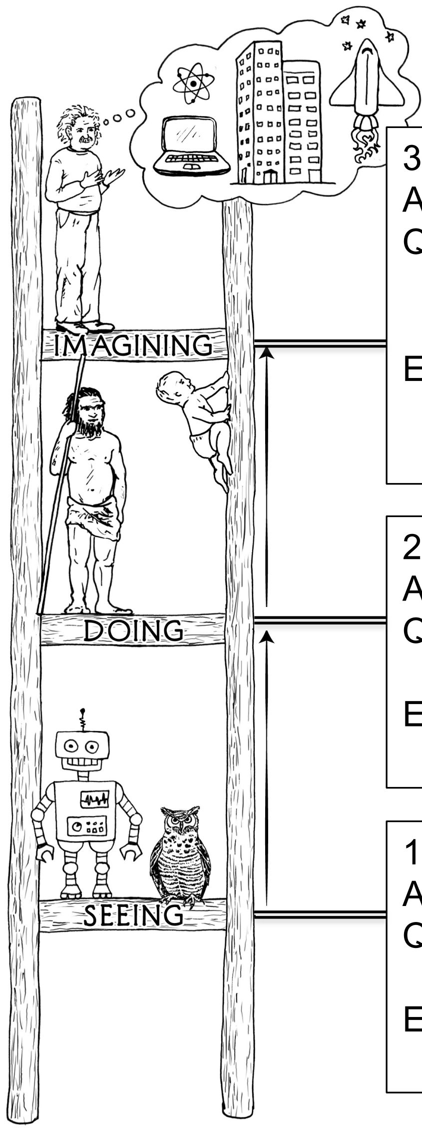
Pearl's Ladder of Causation and the Need for a Causal AI

3-LEVEL HIERARCHY

(Causal) AI?

RL

ML/DL



3. COUNTERFACTUALS
ACTIVITY: Imagining, Retrospection, Understanding
QUESTIONS: *What if I had done . . . ? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
EXAMPLES: Was it the aspirin that stopped my headache?
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

2. INTERVENTION
ACTIVITY: Doing, Intervening
QUESTIONS: *What if I do . . . ? How?*
(What would Y be if I do X?)
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION
ACTIVITY: Seeing, Observing
QUESTIONS: *What if I see . . . ?*
(How would seeing X change my belief in Y?)
EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

Source: The Book of Why, Pearl & Mc Kenzie

Pearl's Ladder of Causation and the Need for a Causal AI

credal nets

as a tool

to climb the **top**

(= counterfactuals)

of the ladder

(Cau

A

s?

R

ML/DL

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

Source: The Book of Why, Pearl & Mc Kenzie

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$

Boolean X

$P(X=0)=p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U

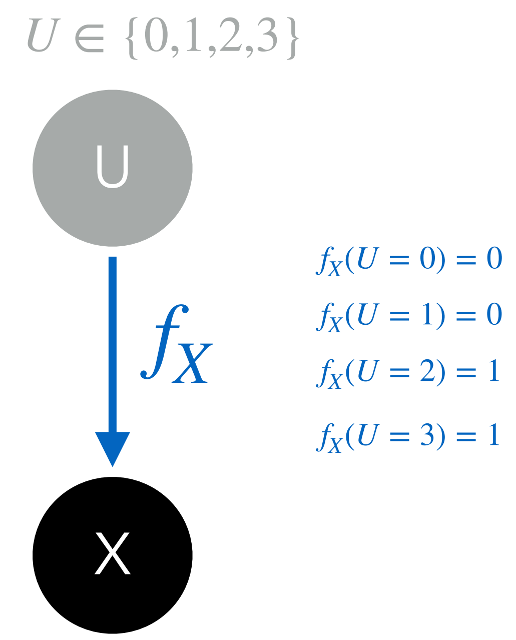
$U \in \{0,1,2,3\}$

Boolean X

$P(X=0)=p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U
- States of U determines those of X through a **structural equation** f_X
  f_X surjective but not invertible

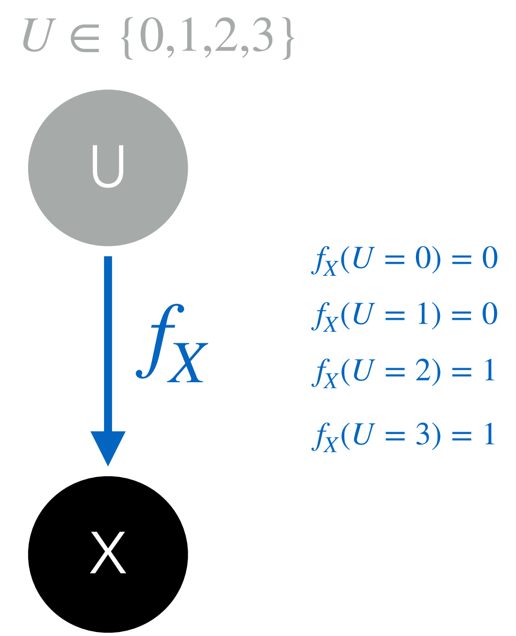


Boolean X

$P(X=0)=p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U
- States of U determines those of X through a **structural equation** f_X
  f_X surjective but not invertible
- $P(x) = \sum_x P(x \quad u)P(u) = \sum_u \delta_{f(u),x}P(u)$



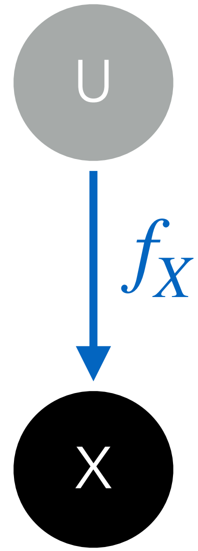
Boolean X
$P(X=0)=p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U
- States of U determines those of X through a **structural equation** f_X
  f_X surjective but not invertible
- $P(x) = \sum_x P(x \;\; u)P(u) = \sum_u \delta_{f(u),x} P(u)$
- A $P(U)$ giving $P(X)$? More than one!

$P(U) = \left[\frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2}\right]$

$U \in \{0,1,2,3\}$



$f_X(U=0)=0$

$f_X(U=1)=0$

$f_X(U=2)=1$

$f_X(U=3)=1$

Boolean X

$P(X=0)=p$

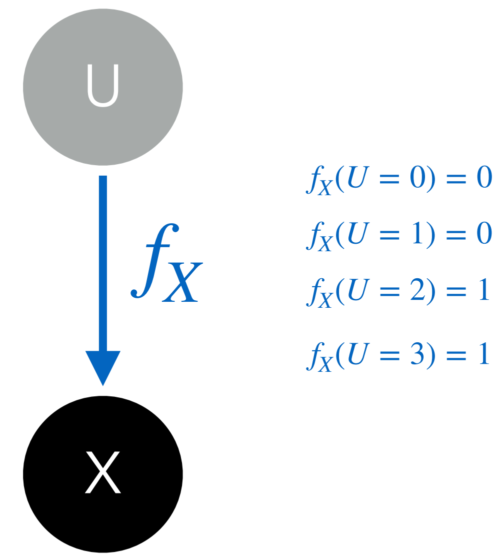
Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U
- States of U determines those of X through a **structural equation** f_X
  f_X surjective but not invertible
- $$P(x) = \sum_x P(x \quad u)P(u) = \sum_u \delta_{f(u),x} P(u)$$
- $P(U)$ giving $P(X)$? More than one!
- Causal inference to be based on the credal set $K(U)$ compatible with $P(X)$

$K(U) = \{P(U) : P(U=0) + P(U=1) = p\}$

$P(U) = \left[\frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2}\right]$

$U \in \{0,1,2,3\}$



Boolean X

$P(X=0) = p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A lat
- State
  throu
  f_X su
- $P(x) = \sum_x P(x \; u)P(u) = \sum_u \delta_{f(u),x}P(u)$
- $P(U)$ giving $P(X)$? More than one!
- Causal inference to be based on the credal set $K(U)$ compatible with $P(X)$

This is a (minimalistic) structural causal model

$K(U) = \{P(U) : P(U=0) + P(U=1) = p\}$

$P(U) = \left[\frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2}\right]$

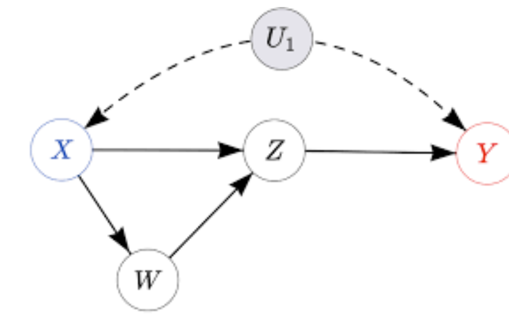
$U \in \{0,1,2,3\}$

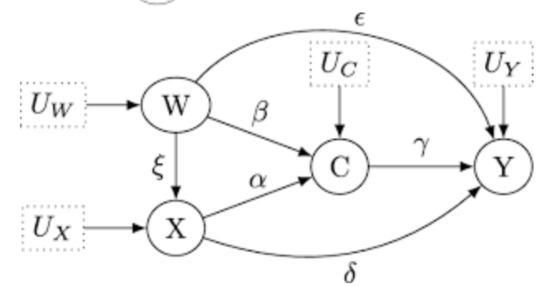
Boolean X

$P(X=0) = p$

Structural Causal Models (General Definition)

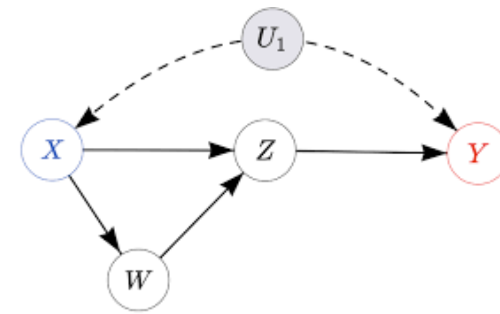
- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} := (U_1, \ldots, U_m)$ (exogenous variables)
- Directed graph $\mathcal{G}$ assumed to be semi-Markovian = root in $\mathbf{U}$, non-root in $\mathbf{X}$
- Equation $X = f_X(\mathrm{Pa}_X)$ for each $X \in \mathbf{X}$
- Marginal $P(U)$ for $U \in \mathbf{U}$ (assessed if possible)

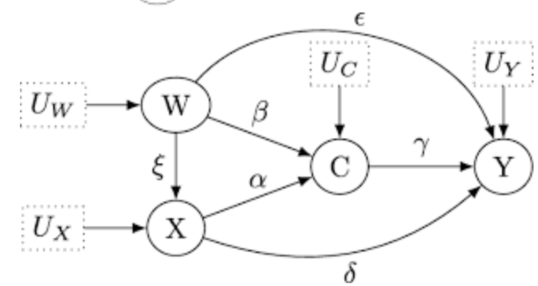




Structural Causal Models (General Definition)

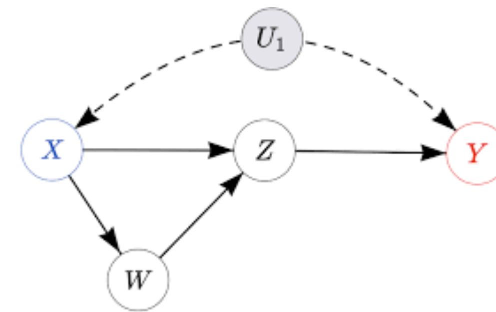
- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} := (U_1, \ldots, U_m)$ (exogenous variables)
- Directed graph $\mathscr{G}$ assumed to be semi-Markovian = root in $\mathbf{U}$, non-root in $\mathbf{X}$
- Equation $X = f_X(\mathrm{Pa}_X)$ for each $X \in \mathbf{X}$
- Marginal $P(U)$ for $U \in \mathbf{U}$ (assessed if possible)
- SCM = BN with CPTs $P(X \;\; \mathrm{Pa}_X) = \delta_{X, f_X(\mathrm{Pa}_X)}$
- Joint PMF $P(\mathbf{x}, \mathbf{u}) = \prod_{U \in \mathbf{U}} P(u) \prod_{X \in \mathbf{X}} \delta_{f_X(\mathrm{pa})_X, x}$

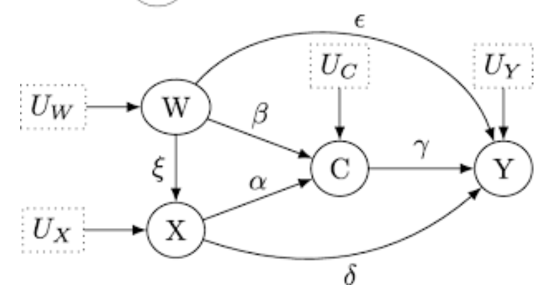


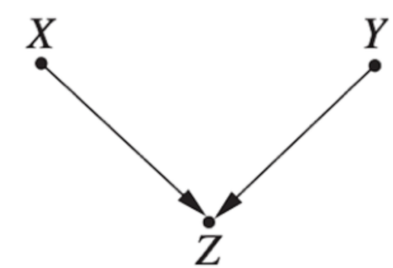


Structural Causal Models (General Definition)

- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} := (U_1, \ldots, U_m)$ (exogenous variables)
- Directed graph $\mathscr{G}$ assumed to be semi-Markovian = root in $\mathbf{U}$, non-root in $\mathbf{X}$
- Equation $X = f_X(\mathrm{Pa}_X)$ for each $X \in \mathbf{X}$
- Marginal $P(U)$ for $U \in \mathbf{U}$ (assessed if possible)
- SCM = BN with CPTs $P(X \ \mathrm{Pa}_X) = \delta_{X, f_X(\mathrm{Pa}_X)}$
- Joint PMF $P(\mathbf{x}, \mathbf{u}) = \prod_{U \in \mathbf{U}} P(u) \prod_{X \in \mathbf{X}} \delta_{f_X(\mathrm{pa})_X, x}$
- Here discrete vars, continuous case analogous







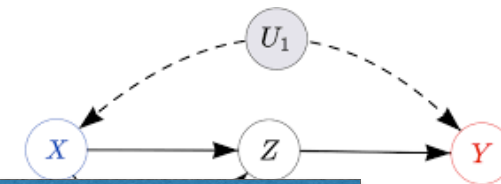
$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$

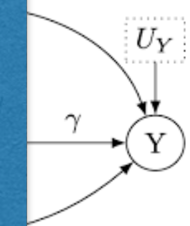
$f_Z : Z = 2X + 3Y$

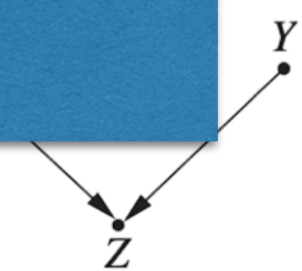
Structural Causal Models (General Definition)

- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} :=$
- Direc
  semi
- Equa
- Marg
- SCM
- Joint PMF $P(\mathbf{x}, \mathbf{u}) = \prod_{U \in \mathbf{U}} P(u) \prod_{X \in \mathbf{X}} \delta_{f_X(\text{pa})_X, x}$
- Here discrete vars, continuous case analogous

SCMs as (one of) the most powerful tools for causal analyses







$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$
$f_Z : Z = 2X + 3Y$

Headache Example (Staying on the First Rung)

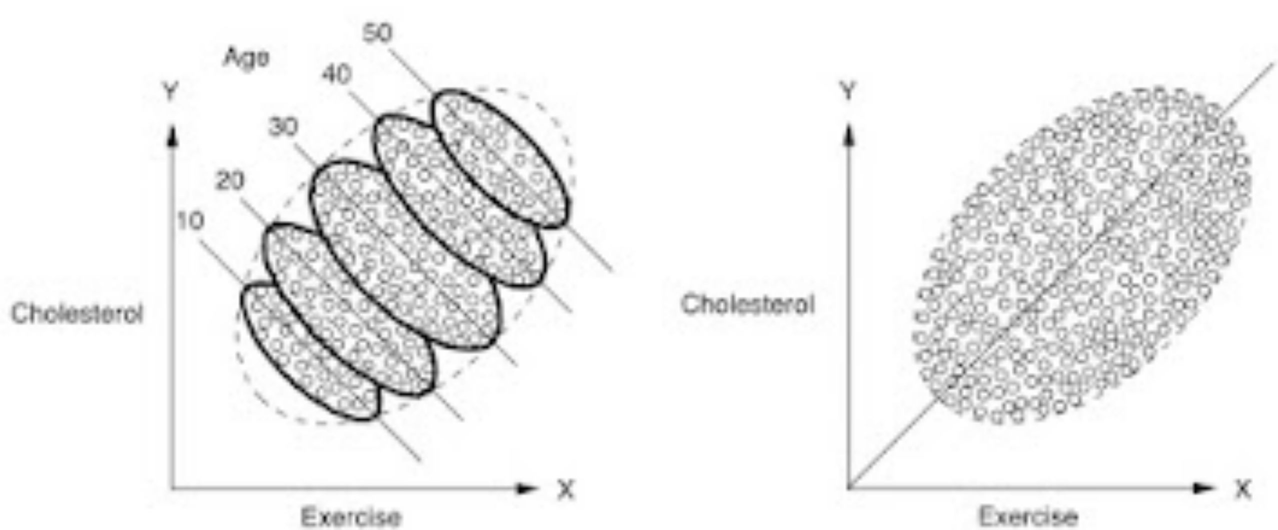
- You take aspirin ($X = 1$) and headache vanishes ($Y = 1$)
- Probability that this has been due to aspirin?
- Observational data $\mathscr{D}$ about the two variables available
- From $\mathscr{D}$, $P(Y = 0 \;\; X = 0) = 0.5 > P(Y = 0 \;\; X = 1) = 0.1$

X	Y	n
0	0	...
0	1	....
1	0	....
1	1	....

Headache Example (Staying on the First Rung)

- You take aspirin ($X = 1$) and headache vanishes ($Y = 1$)
- Probability that this has been due to aspirin?
- Observational data $\mathscr{D}$ about the two variables available
- From $\mathscr{D}$, $P(Y = 0 \;\; X = 0) = 0.5 > P(Y = 0 \;\; X = 1) = 0.1$
- Not genuine causal analysis: adding further covariates might give contradictory results (Simpson's paradox)

X	Y	n
0	0	...
0	1	....
1	0	....
1	1	....

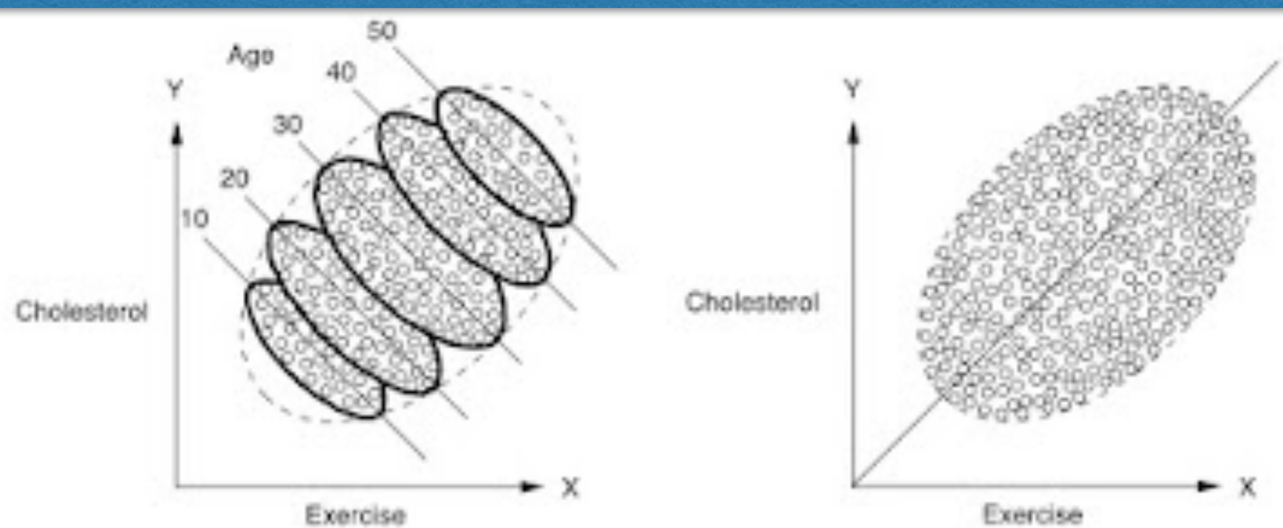


Headache Example (Staying on the First Rung)

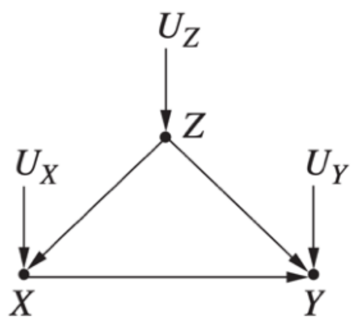
- You take aspirin ($X = 1$) and headache vanishes ($Y = 1$)
- Probability that this has been due to aspirin?
- Obs
- Fron
- Not
  migh

X	Y	n
	0	...
	1	....
	0	....
	1	....

Time to climb up the ladder

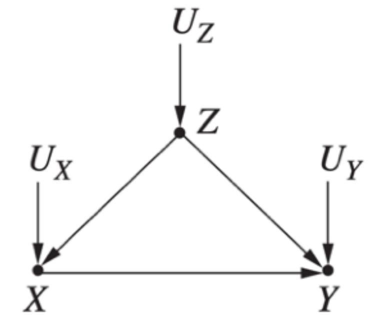


Take the Aspirin! (Interventions = Second Rung)

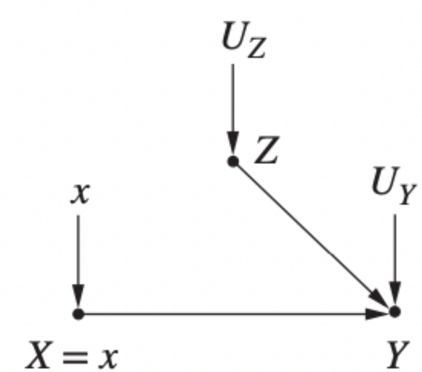


- Gender Z as an additional (endogenous) variable
- Markovian $\mathscr{G}$ (one exo parent for each endo)
- Force people to take aspirin = **intervention** $\text{do}(X = 1)$

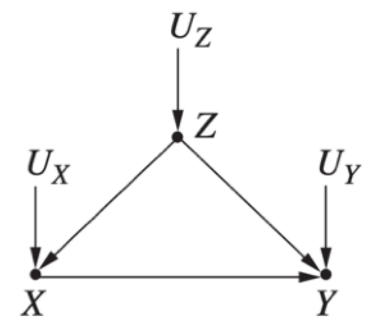
Take the Aspirin! (Interventions = Second Rung)



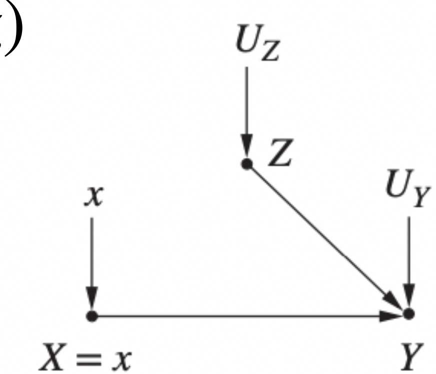
- Gender Z as an additional (endogenous) variable
- Markovian $\mathscr{G}$ (one exo parent for each endo)
- Force people to take aspirin = **intervention** $\mathrm{do}(X = 1)$
- f_X should be modified (constant output), after a **surgery** on $\mathscr{G}$ (incoming arcs removed) intervention = observation



Take the Aspirin! (Interventions = Second Rung)

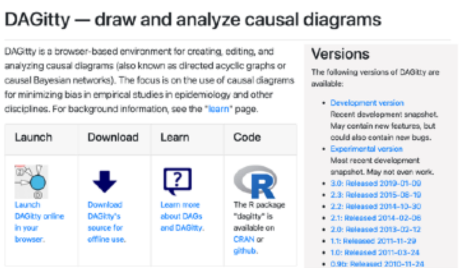


- Gender Z as an additional (endogenous) variable
- Markovian $\mathscr{G}$ (one exo parent for each endo)
- Force people to take aspirin = **intervention** $\text{do}(X = 1)$
- f_X should be modified (constant output), after a **surgery** on $\mathscr{G}$ (incoming arcs removed) intervention = observation
- Pearl's **do calculus** allows to reduce interventional queries to observational ones (solved by BN inference)
- E.g., backdoor $P(y\ \ \text{do}(X = x)) = \sum_z P(y\ \ x, z) \cdot P(z)$
- Do calculus only needs $\mathscr{G}$ (and not the SCM)!



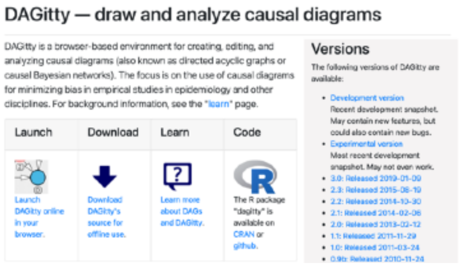
Identifiability of Causal Queries

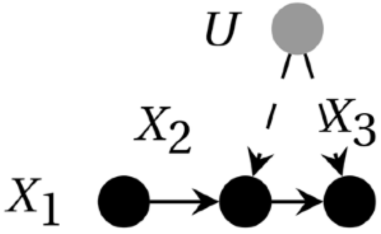
- Do calculus reduces interventional to observational queries by exploiting d-separation in SCMs
- Sound and complete (graph-theoretic) algorithm + inference in the empirical joint PMF
- Alternatively: surgery and inference in the SCM …



Identifiability of Causal Queries

- Do calculus reduces interventional to observational queries by exploiting d-separation in SCMs
- Sound and complete (graph-theoretic) algorithm + inference in the empirical joint PMF
- Alternatively: surgery and inference in the SCM …
- Not all queries can be computed by do calculus. If not we call the query **unidentifiable**
- Emerging idea: unidentifiable queries are only partially identifiable (bounds can be estimated!)
- Recent works in this field by various groups: sampling (Bareinboim), poly programming (Shpitser)



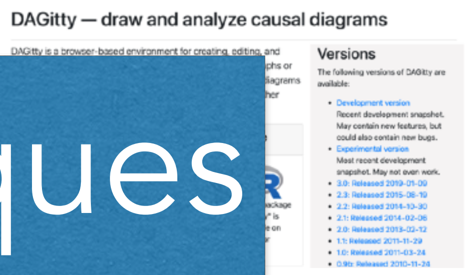


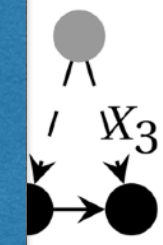
$P(x_3 \;\; do(x_2)) \in [l, u]$

Identifiability of Causal Queries

- Do calculus reduces interventional to observational quer
- Sour + inf
- Alter
- Not If no
- Emerging idea: unidentifiable queries are only partially identifiable (bounds can be estimated!)
- Recent works in this field by various groups: sampling (Bareinboim), poly programming (Shpitser)

Optimisation techniques for CNs to be used for partial identifiability

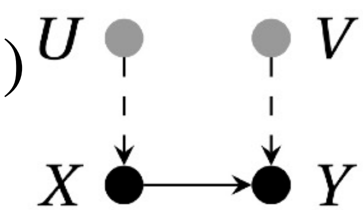




$P(x_3 \quad do(x_2)) \in [l, u]$

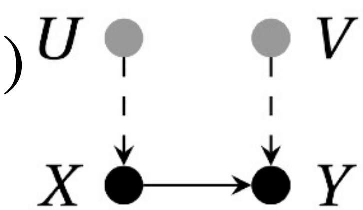
Back to Headache (Moving to the Third Rung)

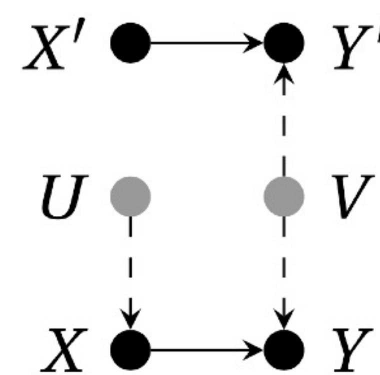
- **What if** I had not taken the aspirin, would have headache stayed?
- An intervention contrasting the current observation …
- This is a **counterfactual** query $P(Y_{X=0} = 0 \quad X = 1, Y = 1)$ (called probability of necessity, PN, sub denote do)



Back to Headache (Moving to the Third Rung)

- **What if** I had not taken the aspirin, would have headache stayed?
- An intervention contrasting the current observation ...
- This is a **counterfactual** query $P(Y_{X=0} = 0 \;\; X = 1, Y = 1)$ (called probability of necessity, PN, sub denote do)
- We need the complete SCM: $\mathcal{G} + \{f_X\}_{X \in \mathbf{X}} + \{P(U)\}_{U \in \mathbf{U}}$
- With complete SCM, an augmented model called **twin network** with duplicated endogenous variables is used for counterfactual analysis after surgery
- (Non-trivial) **counterfactuals are unidentifiable**!





To Compute Counterfactuals ...

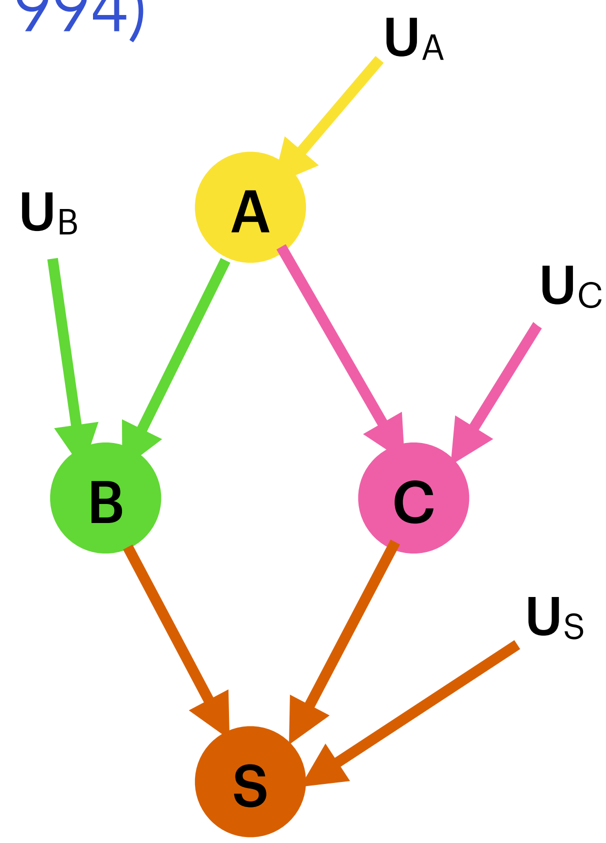
- We need a fully specified SCM, i.e.,
  1. Graph $\mathscr{G}$ over $(\mathbf{X}, \mathbf{U})$ (often available by domain expert or Markovian assumption)
  2. Endogenous equations $\{f_X\}_{X \in \mathbf{X}}$ (available or obtained by complete enumeration)
  3. Exogenous marginals $\{P(U)\}_{U \in \mathbf{U}}$ (rarely available)

To Compute Counterfactuals ...

- We need a fully specified SCM, i.e.,
  1. Graph $\mathscr{G}$ over $(\mathbf{X}, \mathbf{U})$
     (often available by domain expert or Markovian assumption)
  2. Endogenous equations $\{f_X\}_{X \in \mathbf{X}}$
     (available or obtained by complete enumeration)
  3. Exogenous marginals $\{P(U)\}_{U \in \mathbf{U}}$ (rarely available)
- Latent $P(\mathbf{U}) = \prod P(U)$ unavailable? We have data $\mathscr{D}$ about $\mathbf{X}$
- Compute counterfactual = Compute $\{P(U)\}_{U \in \mathbf{U}}$ from $\mathscr{D}$
- Not a new problem: LP approach for special cases already in Balke and Pearl (1994), but do-calculus reduced attention to CFs

Causal Analysis at the Party (Balke & Pearl 1994)

Ann sometimes goes to parties
Bob is not a party guy,
but he likes Ann
and he might be there
Carl broke up with Ann,
he tries to avoid Ann,
but he likes parties
Carl and Bob hate each other,
they might have a Scuffle
if both at the party



besides such knowledge assume
we have observations $\mathscr{D}$ corresponding
to a joint mass function $P(A, B, C, S)$
(e.g., in the form of a BN)

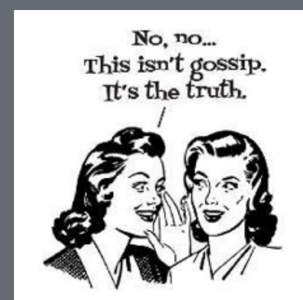
Causal Analysis at the Party (Balke & Pearl 1994)

CAUSAL GOSSIP

INTERVENTIONAL

"**A**nn must not be at the party, or **B**ob would be there instead of home"

$$P(B\ \mathrm{do}(\overline{a})) = ?$$



COUNTERFACTUAL

"If **B**ob were at the party, then Bob and Carl would surely **S**cuffle"

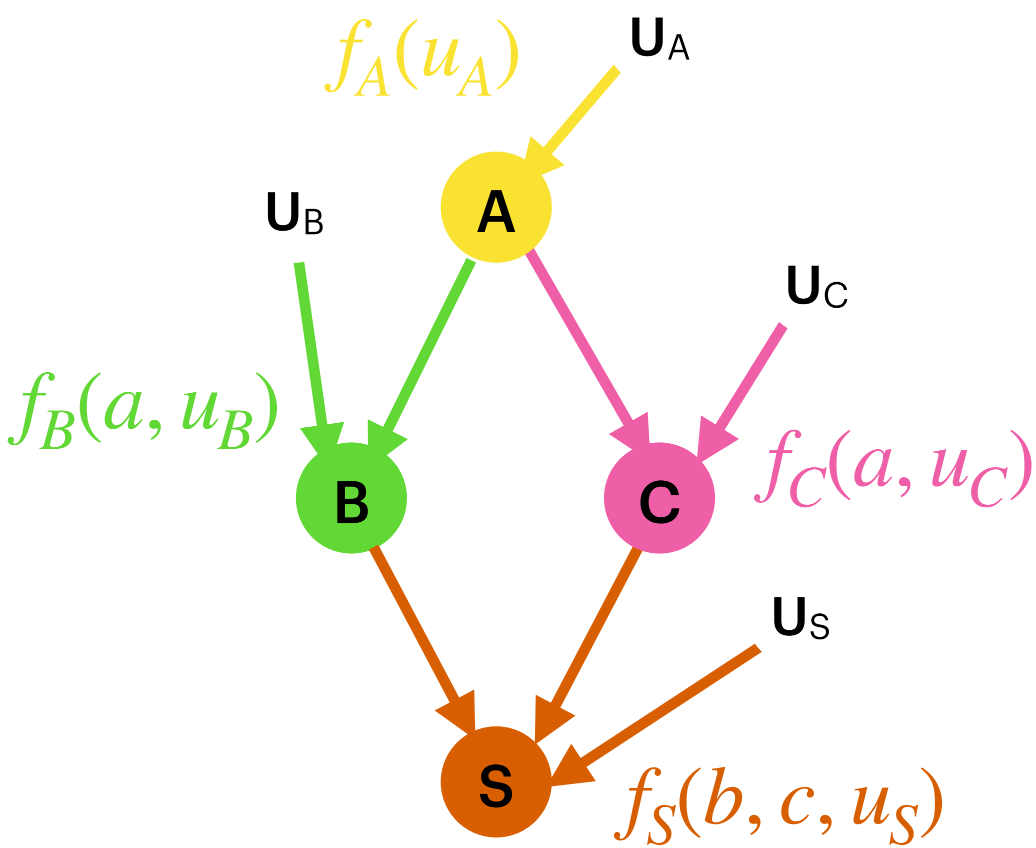
$$P(S_b\ \overline{b}) = ?$$

a (fully specified) SCM can answer these questions

Let's (Eventually) Use IPs!

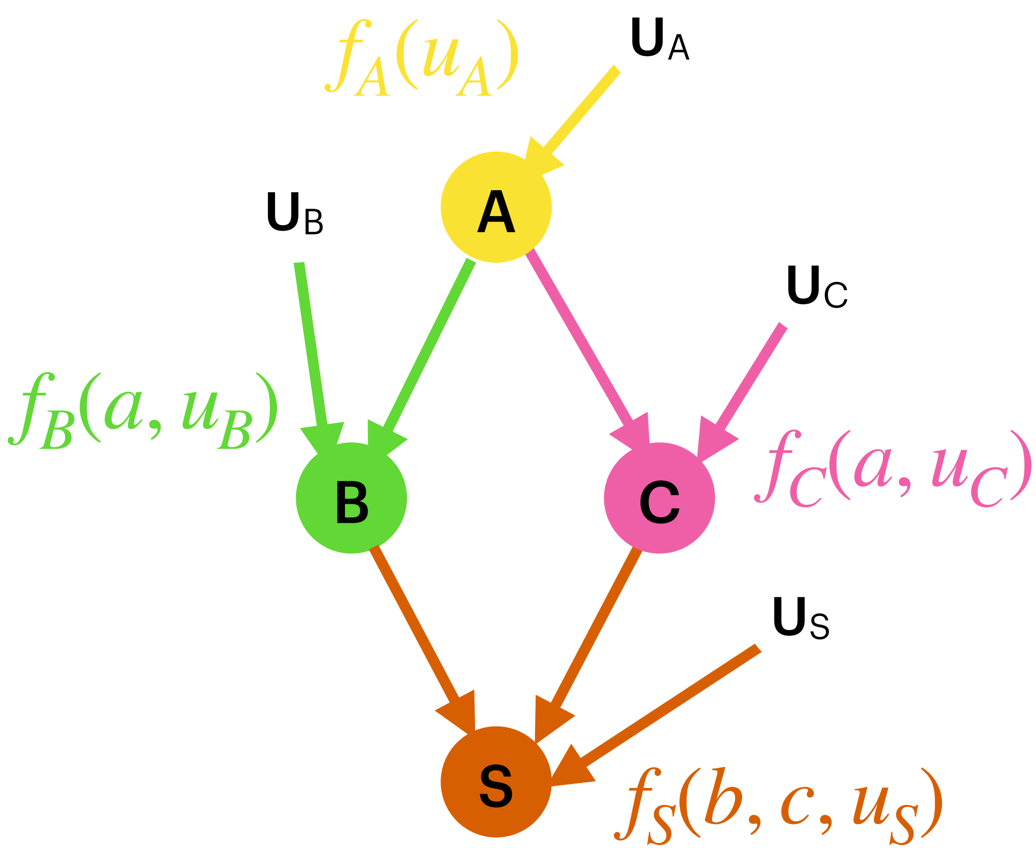
- Find the exogenous marginals?

  $P(U_A)P(U_B)P(U_C)P(U_S)$



Let's (Eventually) Use IPs!

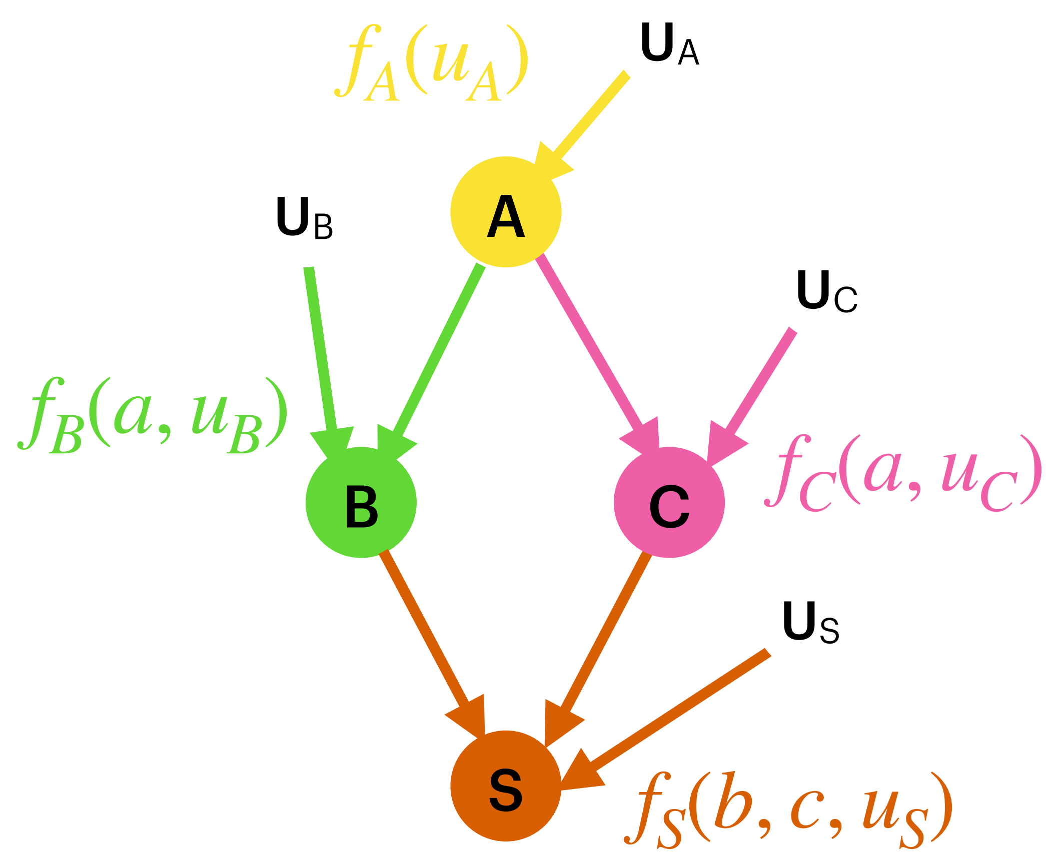
- Find the exogenous marginals?
  $P(U_A)P(U_B)P(U_C)P(U_S)$
- **Endogenous** (= with $\mathscr{D}$) **consistency**



$$\sum_{u_A, u_B, u_C, u_D} \left[p(u_A) \cdot \delta_{a, f_A(u_A)} \cdot p(u_B) \cdot \delta_{b, f_B(a, u_B)} \cdot p(u_C) \cdot \delta_{c, f_C(a, u_c)} \cdot p(u_S) \cdot \delta_{s, f_S(b, c, u_S)} \right] = \tilde{p}(a, b, c, s)$$

Let's (Eventually) Use IPs!

- Find the exogenous marginals?
  $P(U_A)P(U_B)P(U_C)P(U_S)$
- **Endogenous** (= with $\mathscr{D}$) **consistency**
- This induces global non-linear (so-called Verma) constraints

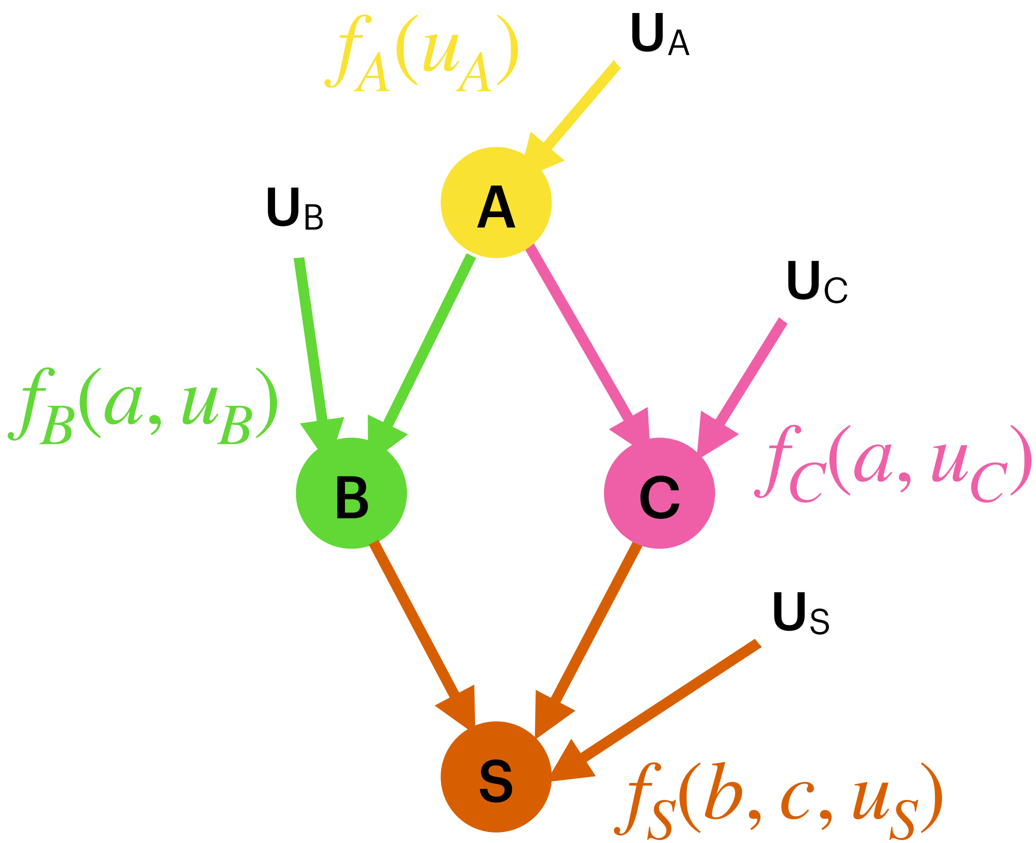


Unknown / Unknown / Unknown / Unknown / Empirical, known

$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A) \cdot \delta_{a,f_A(u_A)} \cdot p(u_B) \cdot \delta_{b,f_B(a,u_B)} \cdot p(u_C) \cdot \delta_{c,f_C(a,u_c)} \cdot p(u_S) \cdot \delta_{s,f_S(b,c,u_S)}\right] = \tilde{p}(a,b,c,s)$$

Let's (Eventually) Use IPs!

- Find the exogenous marginals?
  $P(U_A)P(U_B)P(U_C)P(U_S)$
- **Endogenous** (= with $\mathscr{D}$) **consistency**
- This induces global non-linear (so-called Verma) constraints
- Constraints became local and linear ones by marginalisation and conditioning (Zaffalon et al., 2020)

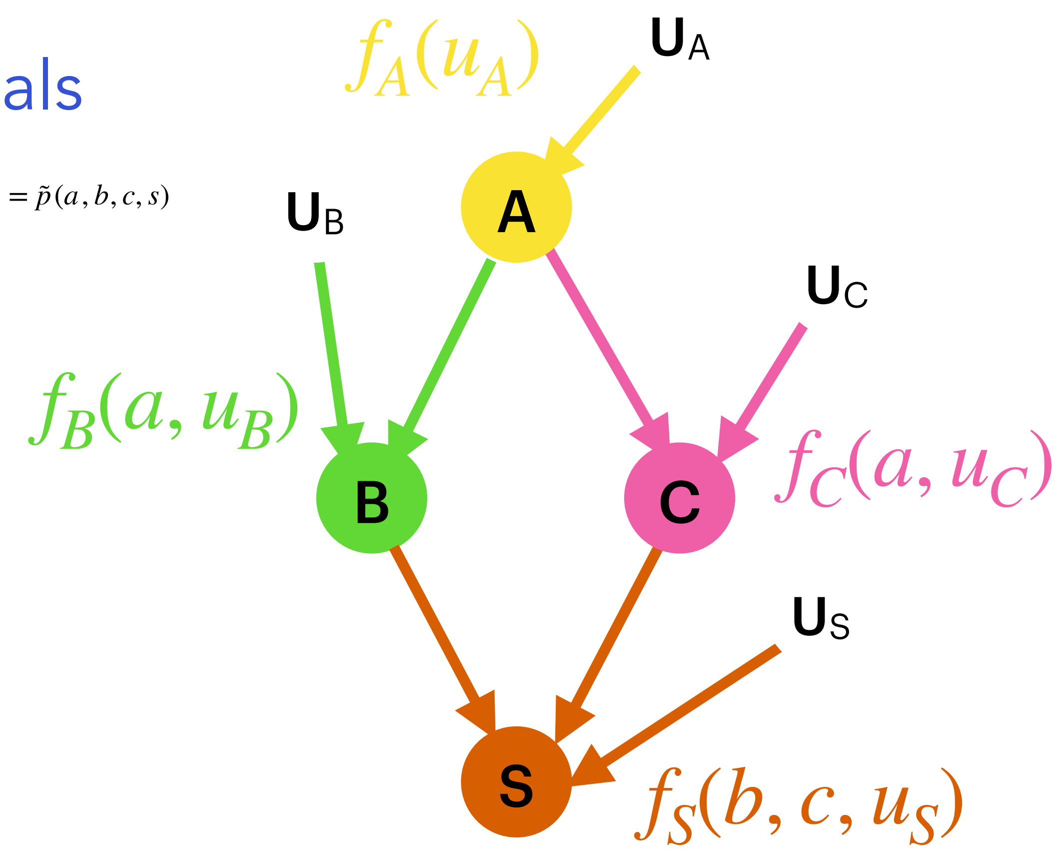


Unknown / Unknown / Unknown / Unknown / Empirical, known

$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A) \cdot \delta_{a,f_A(u_A)} \cdot p(u_B) \cdot \delta_{b,f_B(a,u_B)} \cdot p(u_C) \cdot \delta_{c,f_C(a,u_c)} \cdot p(u_S) \cdot \delta_{s,f_S(b,c,u_S)} \right] = \tilde{p}(a,b,c,s)$$

Constraining Exogenous Marginals

$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A) \cdot \delta_{a,f_A(u_A)} \cdot p(u_B) \cdot \delta_{b,f_B(a,u_B)} \cdot p(u_C) \cdot \delta_{c,f_C(a,u_C)} \cdot p(u_S) \cdot \delta_{s,f_S(b,c,u_S)} \right] = \tilde{p}(a,b,c,s)$$



Constraining Exogenous Marginals

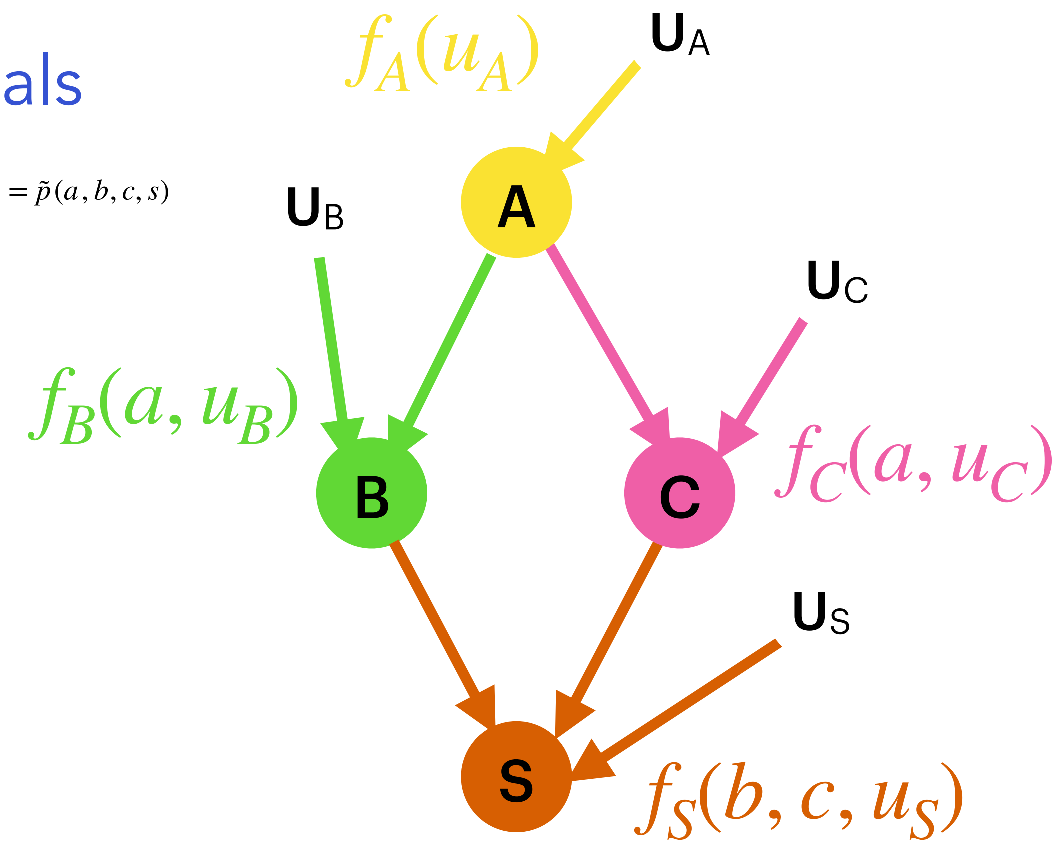
$$\sum_{u_A, u_B, u_C, u_D} \left[p(u_A) \cdot \delta_{a, f_A(u_A)} \cdot p(u_B) \cdot \delta_{b, f_B(a, u_B)} \cdot p(u_C) \cdot \delta_{c, f_C(a, u_C)} \cdot p(u_S) \cdot \delta_{s, f_S(b,c,u_S)} \right] = \tilde{p}(a, b, c, s)$$

$$P(a) = \sum_{u_A} P(a \;\; u_A) \cdot P(u_A)$$

$$P(b \;\; a) = \sum_{u_B} P(b \;\; a, u_B) \cdot P(u_B)$$

$$P(c \;\; a) = \sum_{u_C} P(c \;\; a, u_C) \cdot P(u_C)$$

$$P(s \;\; b, c) = \sum_{u_S} P(s \;\; b, c, u_S) \cdot P(u_S)$$



Constraining Exogenous Marginals

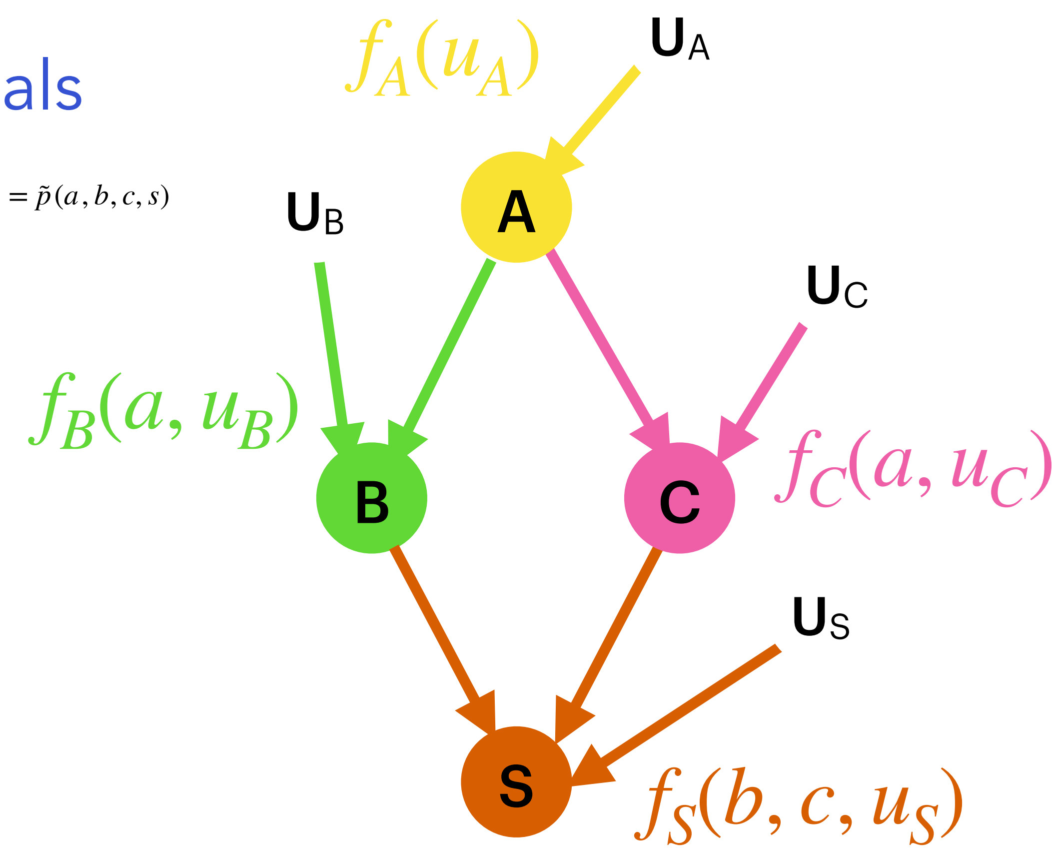
$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A)\cdot\delta_{a,f_A(u_A)}\cdot p(u_B)\cdot\delta_{b,f_B(a,u_B)}\cdot p(u_C)\cdot\delta_{c,f_C(a,u_C)}\cdot p(u_S)\cdot\delta_{s,f_S(b,c,u_S)}\right] = \tilde{p}(a,b,c,s)$$

$$P(a) = \sum_{u_A} P(a \quad u_A)\cdot P(u_A)$$

$$P(b \quad a) = \sum_{u_B} P(b \quad a, u_B)\cdot P(u_B)$$

$$P(c \quad a) = \sum_{u_C} P(c \quad a, u_C)\cdot P(u_C)$$

$$P(s \quad b,c) = \sum_{u_S} P(s \quad b,c,u_S)\cdot P(u_S)$$



- Linear constraints on marginal exogenous probabilities leading to the credal sets specification $K(U_A)$, $K(U_B)$, $K(U_C)$, $K(U_S)$
- Structural equations (= endogenous CPTS) remain unaffected

Constraining Exogenous Marginals

$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A)\cdot\delta_{a,f_A(u_A)}\cdot p(u_B)\cdot\delta_{b,f_B(a,u_B)}\cdot p(u_C)\cdot\delta_{c,f_C(a,u_c)}\cdot p(u_S)\cdot\delta_{s,f_S(b,c,u_S)}\right] = \tilde{p}(a,b,c,s)$$

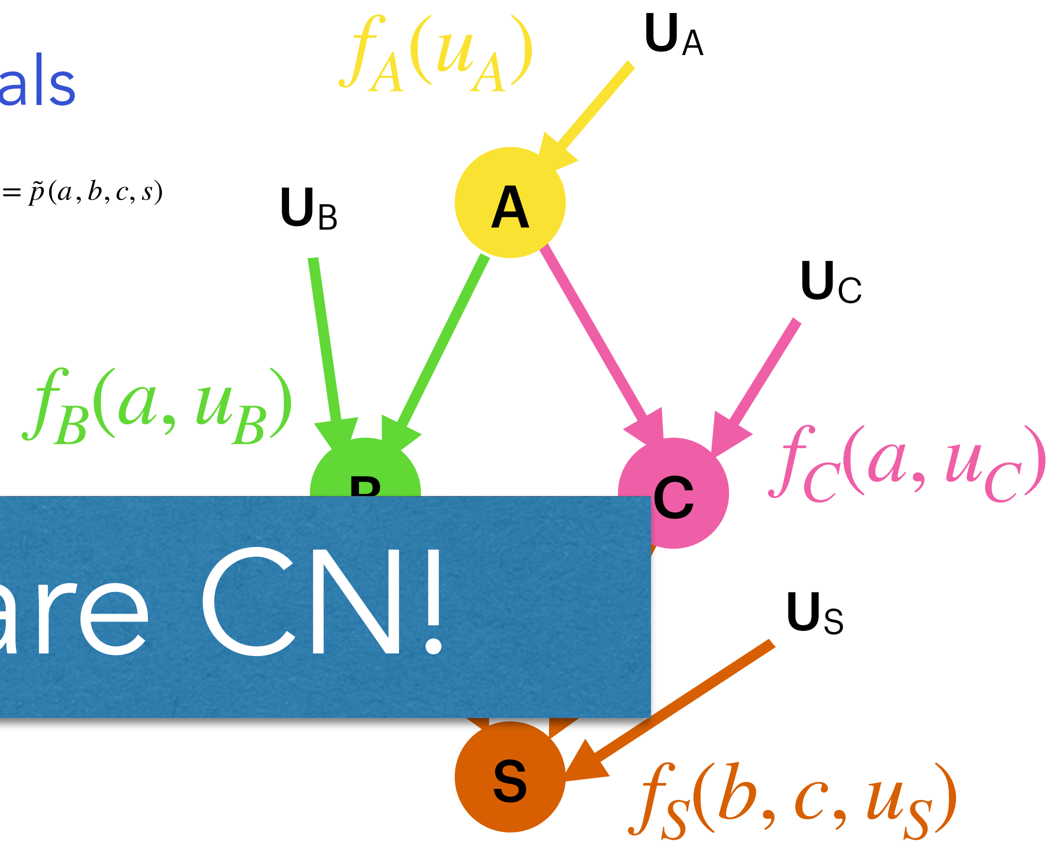
$$P(a) = \sum_{u_A} P(a \;\; u_A)\cdot P(u_A)$$

$$P(b \;\; a) = \sum_{u_B}$$

$$P(c \;\; a) = \sum_{u_C}$$

$$P(s \;\; b,c) = \sum_{u_S} P(s \;\; b,c,u_S)\cdot P(u_S)$$

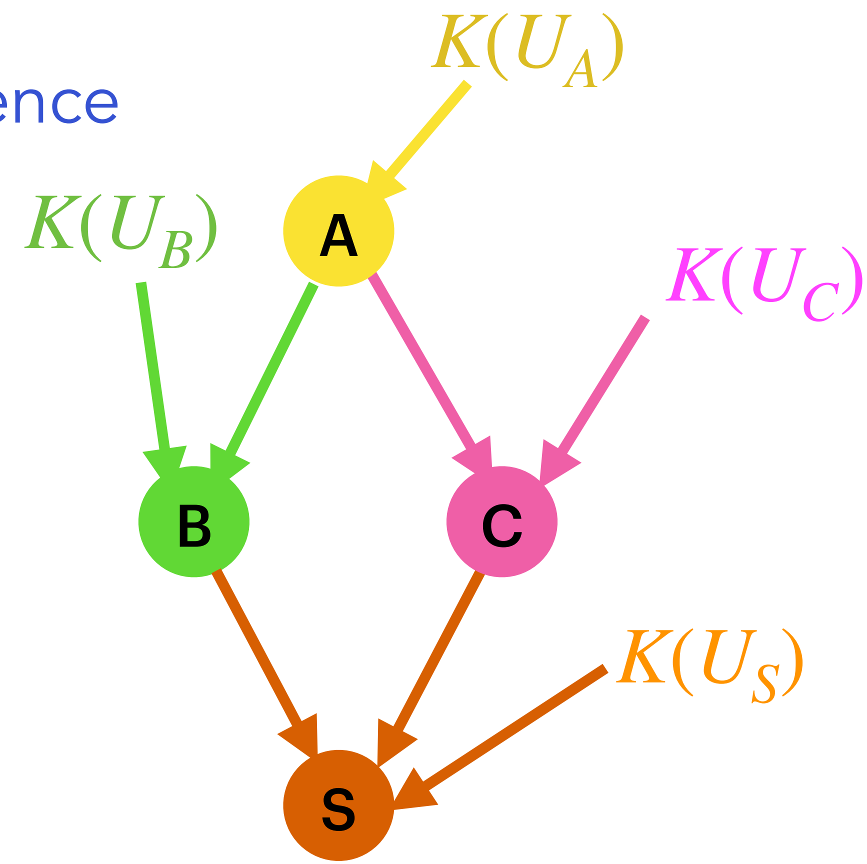
SCMs are CN!



- Linear constraints on marginal exogenous probabilities leading to the credal sets specification $K(U_A)$, $K(U_B)$, $K(U_C)$, $K(U_S)$
- Structural equations (= endogenous CPTS) remain unaffected

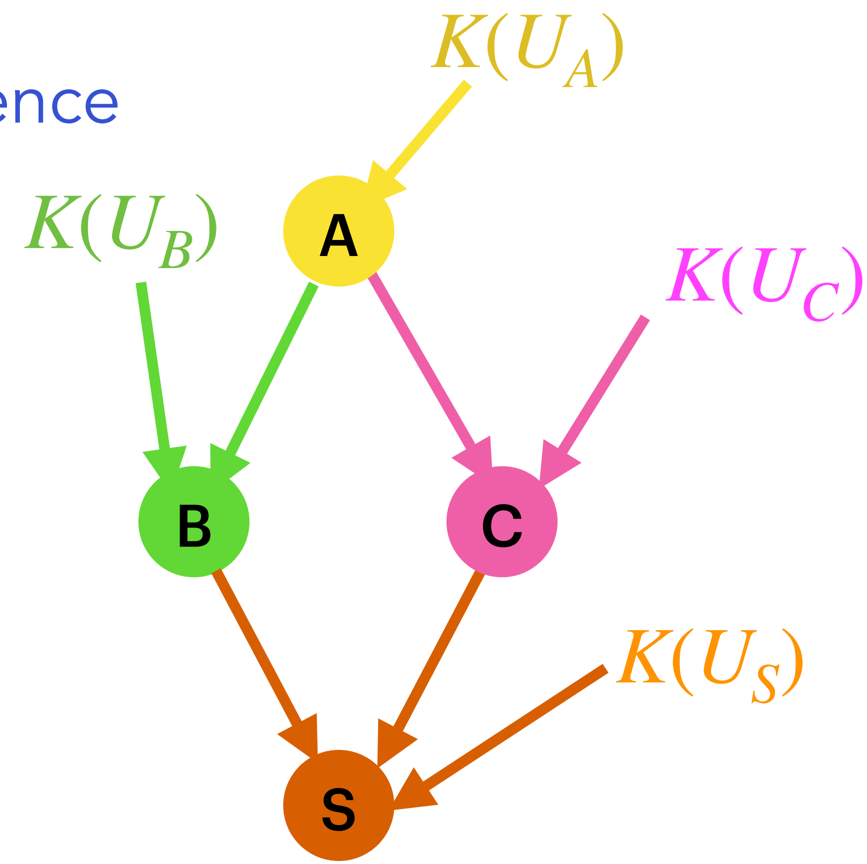
Reducing Causal Queries to CN Inference

- Consistent SCMs as a single CN



Reducing Causal Queries to CN Inference

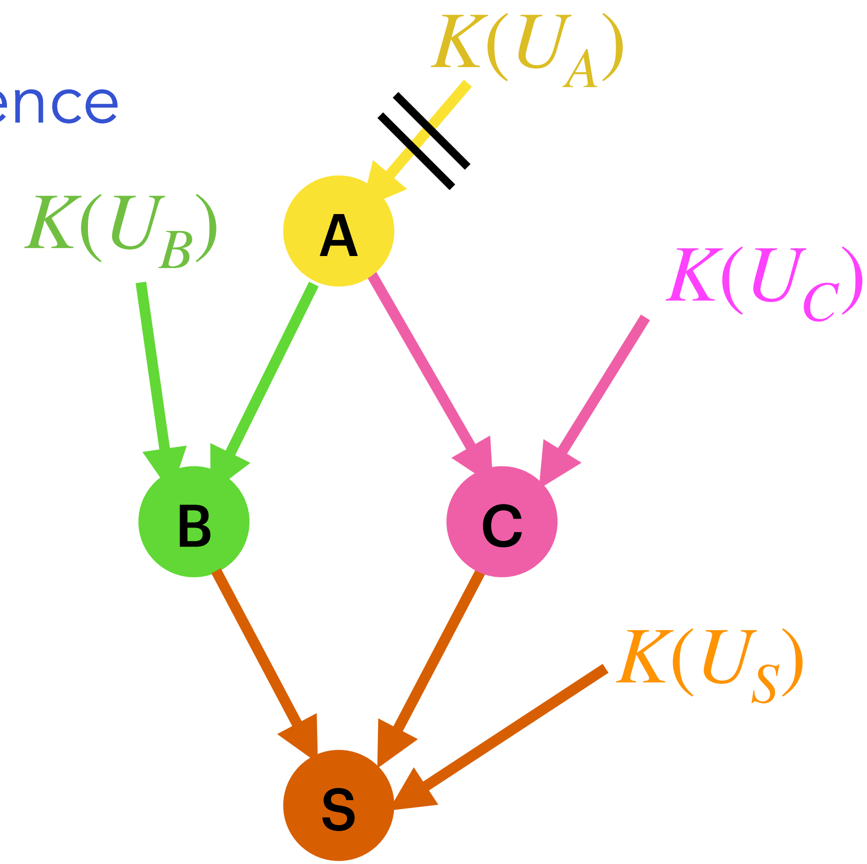
- Consistent SCMs as a single CN
- d-separation holds for CNs, we can do surgery à la Pearl
- CN algs to compute bounds!



Reducing Causal Queries to CN Inference

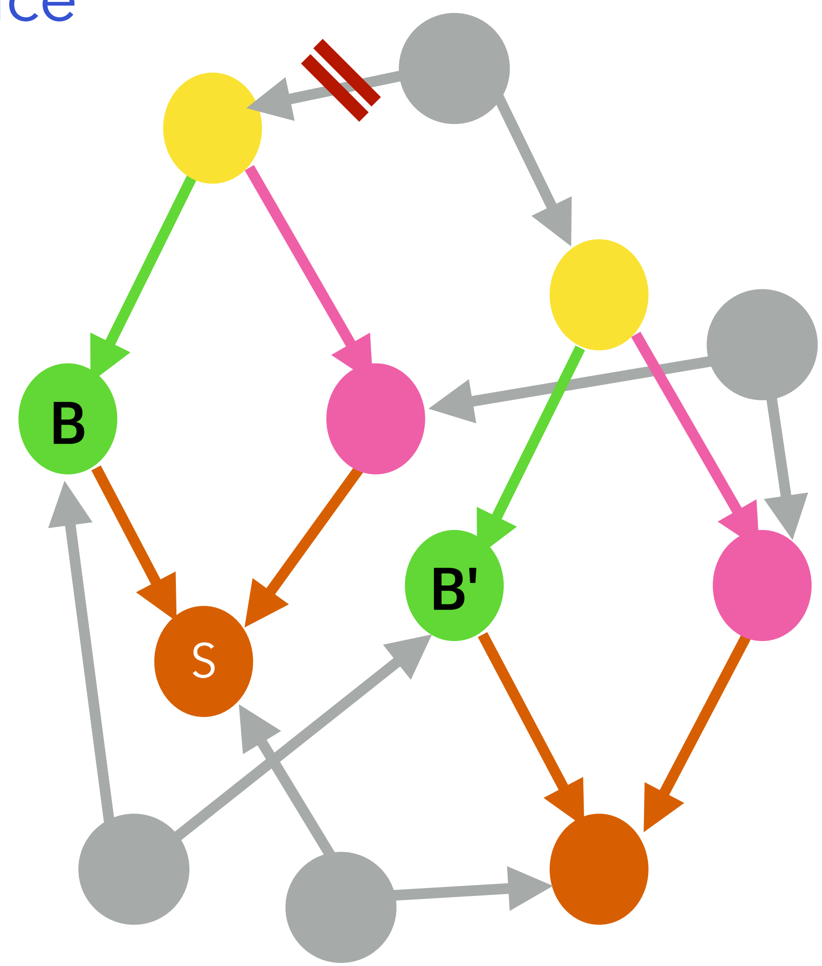
- Consistent SCMs as a single CN
- d-separation holds for CNs,
  we can do surgery à la Pearl
- CN algs to compute bounds!
- Interventions are straightforward

$$P(B \;\; \mathrm{do}(\bar{a})) \in [\underline{P}'(B \;\; \bar{a}), \overline{P}'(B \;\; \bar{a})]$$



Reducing Causal Queries to CN Inference

- Consistent SCMs as a single CN
- d-separation holds for CNs, we can do surgery à la Pearl
- CN algs to compute bounds!
- Interventions are straightforward
  $P(B \;\; \mathrm{do}(\bar{a})) \in [\underline{P}'(B \;\; \bar{a}), \overline{P}'(B \;\; \bar{a})]$
- Counterfactuals require twin nets
  $P(S_b \;\; \bar{b}) \in [\underline{P}(S \;\; b, \bar{b}'), \overline{P}(S \;\; b, \bar{b}')]$
- Identifiable? $\underline{P} = \overline{P}$

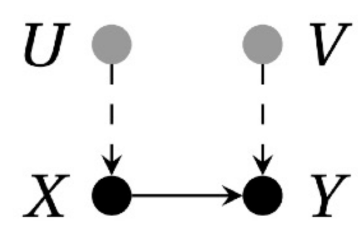


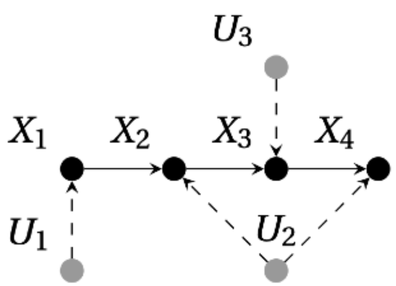
Markovian and Quasi-Markovian SCMs as CNs

Algorithm 1 Given an SCM M and a PMF $\tilde{P}(\boldsymbol{X})$, return CSs $\{K(U)\}_{U \in \boldsymbol{U}}$

for $X \in \boldsymbol{X}$ **do**
  $U \leftarrow \mathrm{Pa}(X) \cap \boldsymbol{U}$ // U as the unique exogenous parent of X
  $\underline{\mathrm{Pa}}(X) \leftarrow \mathrm{Pa}(X) \setminus \{U\}$ // Endogenous parents of X
  if $\underline{\mathrm{Pa}}(X) = \emptyset$ **then**
    $K(U) \leftarrow \{P'(U) : \sum_{u \in f_X^{-1}} P'(u) = \tilde{P}(x), \forall x \in \Omega_X\}$ // Eq. (4)
  else
    $K(U) \leftarrow \{P'(U) : \sum_{u \in f^{-1}_{X|\underline{\mathrm{pa}}(X)}(x)} P'(u) = \tilde{P}(x|\underline{\mathrm{pa}}(X)), \forall x \in \Omega_X, \forall \underline{\mathrm{pa}}(X) \in \Omega_{\underline{\mathrm{Pa}}_X}\}$ // Eq. (6)
  end if
end for

Markovian Models





Quasi-Markovian Models

Algorithm 2 Given an SCM M and a PMF $\tilde{P}(\boldsymbol{X})$, return CSs $\{K(U)\}_{U \in \boldsymbol{U}}$

for $U \in \boldsymbol{U}$ **do**
  $\{X_U^k\}_{k=1}^{n_U} \leftarrow \mathrm{Sort}[X \in \boldsymbol{X} : U \in \mathrm{Pa}(X)]$ // Children of U in topological order
  $\gamma \leftarrow \emptyset$
  for $(x_U^1, \ldots, x_U^{n_U}) \in \times_{k=1}^{n_U} \Omega_{X_U^k}$ **do**
    for $(\underline{\mathrm{pa}}(X_U^1), \ldots, \underline{\mathrm{pa}}(X_U^{n_U})) \in \times_{k=1}^{n_U} \Omega_{\underline{\mathrm{Pa}}(X_U^k)}$ **do**
      $\Omega'_U \leftarrow \cap_{k=1}^{n_U} f^{-1}_{X_U^k|\underline{\mathrm{pa}}(X_U^k)}(x_U^k)$
      $\gamma \leftarrow \gamma \cup \left\{\sum_{u \in \Omega'_U} P(u) = \prod_{k=1}^{n_U} \tilde{P}(x_U^k | x_U^1, \ldots, x_U^{k-1}, \underline{\mathrm{pa}}(X_U^1)), \ldots, \underline{\mathrm{pa}}(X_U^k))\right\}$
    end for
  end for
  $K(U) \leftarrow \{P(U) : \gamma\}$ // CS by linear constraints on $P(U)$
end for

GitHub

Software and Experiments

Crema
Credal Models Algorithms

Java library for CNs

Credici
Credal Inference for Causal Inference

Java library for Causal Inference built on the top of CREMA

GitHub

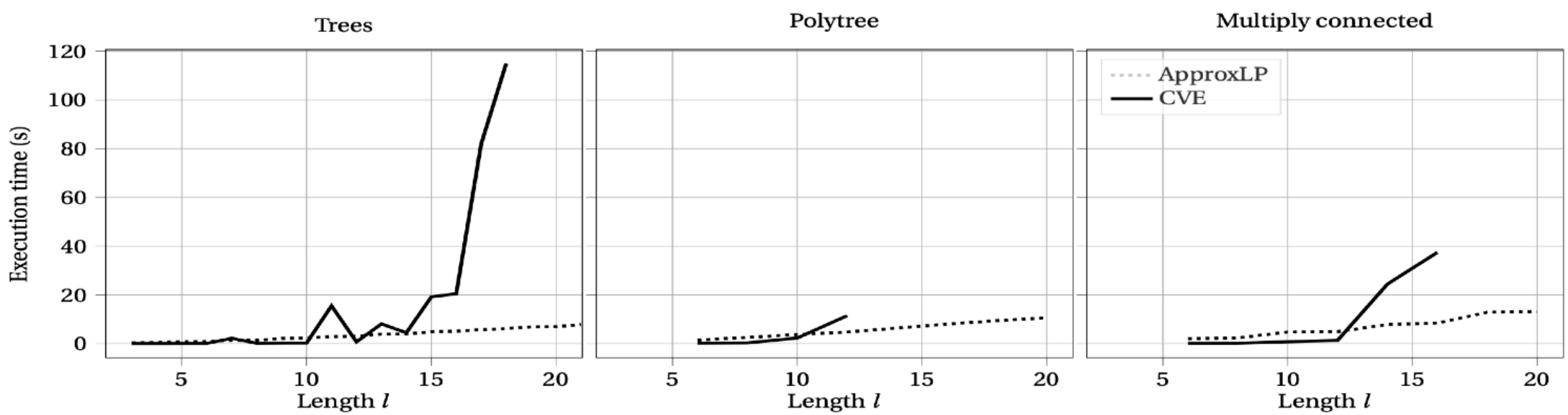
Software and Experiments

Crema
Credal Models Algorithms

Java library for CNs

Credici
Credal Inference for Causal Inference

Java library for Causal Inference built on the top of CREMA



Exact inference by credal variable elimination only for small models

ApproxLP (Antonucci et al., 2014) allows to process larger models

RMSE always <0.7%

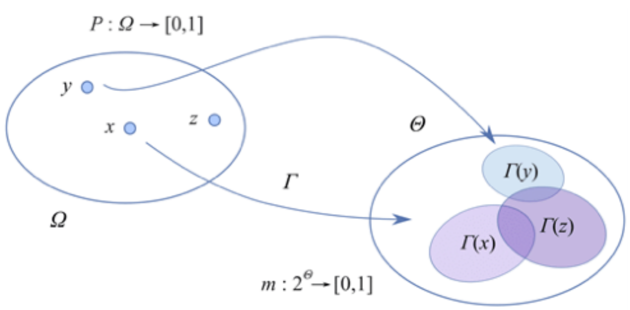
Intermezzo: Belief Functions (as Credal Sets)

- Linear constraints for CN induced by SCM have a peculiar form
- These are CS corresponding to **belief functions** (Dempster '68, Shafer '76)
- Class of generalised probabilistic models
- PMF distributes mass over the singletons, BF over (poss. overlapping) sets
- Dempster's **multi-valued mapping**, in SCMs $\mathbf{U} = f^{-1}(\mathbf{X})$, $\mathrm{BF}(\mathbf{U}) := f^{-1}[P(\mathbf{X})]$
- Dedicated conditioning/combination rules

$$\sum_{u\,:\,\text{condition}} P(u) = \text{const}$$

$$\sum_{u\in\Omega'_U} P(u) = \prod_{k=1}^{n_U} \tilde{P}(x_U^k|x_U^1,\ldots,x_U^{k-1},\underline{\mathrm{pa}}(X_U^1)),\ldots,\underline{\mathrm{pa}}(X_U^k))$$

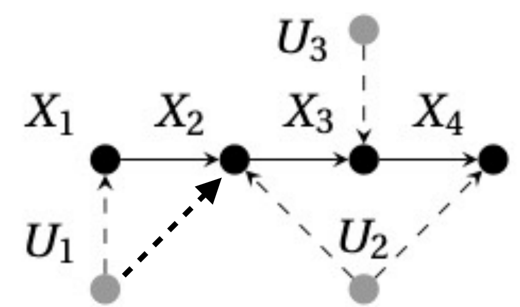
$$\{P'(U) : \sum_{u\in f^{-1}_{X|\underline{\mathrm{pa}}(X)}(x)} P'(u) = \tilde{P}(x|\underline{\mathrm{pa}}(X)), \forall x \in \Omega_X, \forall \underline{\mathrm{pa}}(X) \in \Omega_{\underline{\mathrm{Pa}}_X}\}$$



Credits: Fabio Cuzzolin

Back to SCM2CN: Non Quasi-Markovian Case

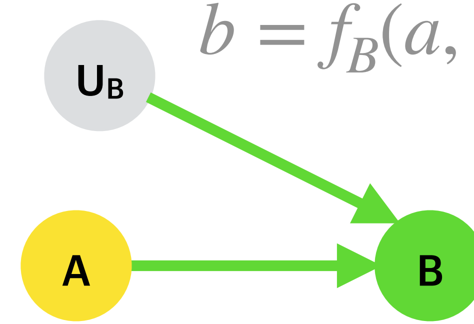
- Non Quasi-Markovian? Non-Linear constraint
- E.g., $\sum P(u_1) \cdot P(u_2) = \ldots$
- Merge exogenous variables $U := (U_1, U_2)$
- Independence constraints can be disregarded (but higher exogenous dimensionality)
- Again CN approximate inference to solve causal queries
- State space dimensionality affects complexity
- We might have very large latent spaces ...



Canonical Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B \quad A)$ should be a deterministic CPT

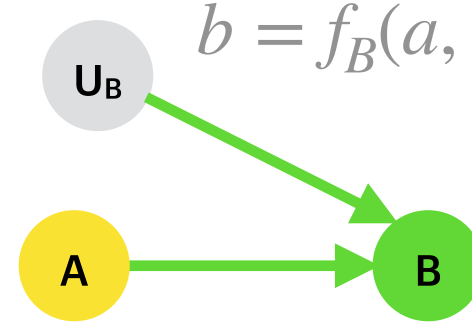
$$b = f_B(a, u_B)$$



Canonical Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B\ \ A)$ should be a deterministic CPT

$$b = f_B(a, u_B)$$



$$P(B\ \ A)$$

	A=0	A=1
B=0	1	1
B=1	0	0

$B = 0$

	A=0	A=1
B=0	1	0
B=1	0	1

$B = A$

	A=0	A=1
B=0	0	1
B=1	1	0

$B = \neg A$

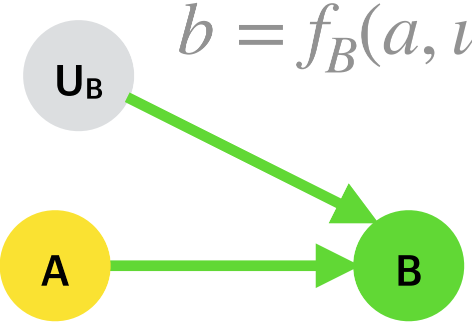
	A=0	A=1
B=0	0	0
B=1	1	1

$B = 1$

Canonical Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B\ \ A)$ should be a deterministic CPT
- U_B indexing all these deterministic CPTs

$$b = f_B(a, u_B)$$



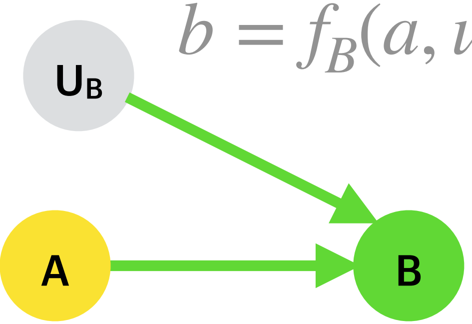
$$P(B\ \ A, U)$$

	A=0	A=1	A=0	A=1	A=0	A=1	A=0	A=1
B=0	1	1	1	0	0	1	0	0
B=1	0	0	0	1	1	0	1	1
	U=0		U=1		U=2		U=3	
	$B = 0$		$B = A$		$B = \neg A$		$B = 1$	

Canonical Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B \;\; A)$ should be a deterministic CPT
- U_B indexing all these deterministic CPTs
- Knowledge might discard some states (ex., Bob goes to the party if Ann does)

$$b = f_B(a, u_B)$$

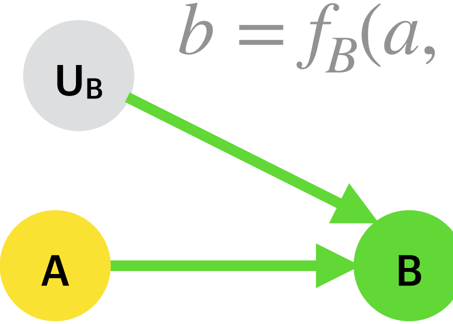


$$P(B \;\; A, U)$$

	A=0	A=1	A=0	A=1	A=0	A=1	A=0	A=1
B=0			1	0			0	0
B=1			0	1			1	1
	U=0		U=1		U=2		U=3	
	$B = 0$		$B = A$		$B = \neg A$		$B = 1$	

Canonical Specification of Structural Equations

$b = f_B(a, u_B)$



- Finding the equations given $\mathcal{G}$ only
- $P(B|A)$ should be a deterministic CPT
- U_B indexing all these deterministic CPTs
- Knowledge might discard some states (ex., Bob goes to the party if Ann does)
- With Boolean parent & child) $|U| = 4$ in general (exp size) :

$$|U| = |X|^{\prod_{Y \in \mathrm{Pa}_Y} |Y|}$$

even more challenging with multiple exogenous parents

$$P(B|A, U)$$

$B = 0$ $B = A$ $B = \neg A$ $B = 1$

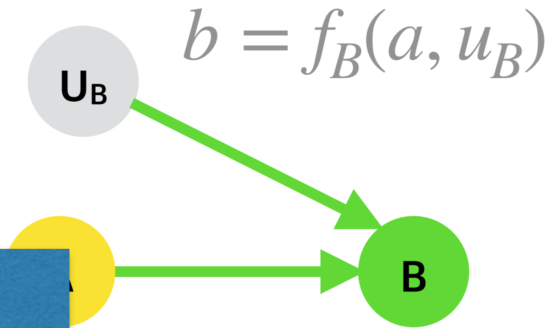
Canonical Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B \mid A)$ should be a deterministic CPT
- U_B indexing
- Knowledge r
  (ex., Bob goe
- With Boolea
  in general (exp size) :

$$U = X^{\prod_{Y \in \mathrm{Pa}_Y} Y}$$

even more challenging
with multiple exogenous parents

CFs based on $\mathscr{G}$ and $\mathscr{D}$ only



$B \mid A, U)$

	A=0	A=1	A=0	A=1	A=0	A=1	A=0	A=1
B=0			1	0			0	0
B=1			0	1			1	1
	U=0		U=1		U=2		U=3	
	$B = 0$		$B = A$		$B = \neg A$		$B = 1$	

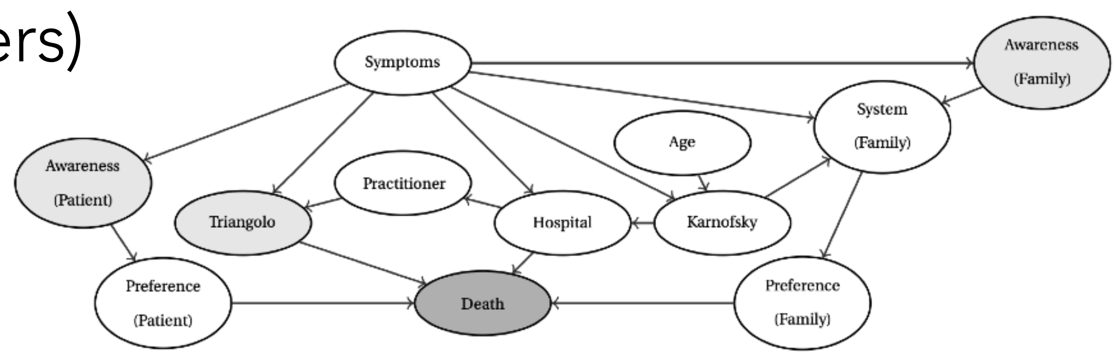
An Application: Counterfactual Analysis in Palliative Cares

- Study of terminally ill cancer patients' preferences wrt their place of death (home or hospital)
- $\mathcal{G}$ obtained by expert knowledge and data
- Exogenous variables?
- Markovian assumption (= no confounders)

ASSOCIAZIONE TRIANGOLO
qualitépalliative marchio di qualità "in cure palliative"
servizio cure palliative e domiciliari

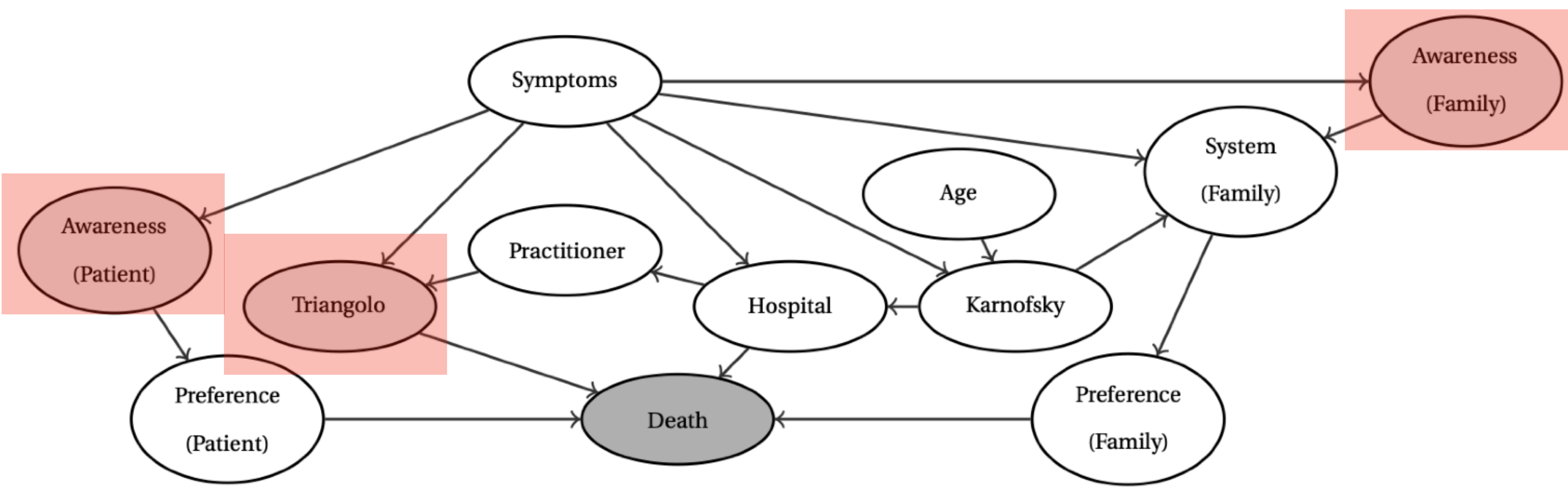

Impact on place of death in cancer patients: a causal exploration in southern Switzerland

Heidi Kern [1], Giorgio Corani [2], David Huber [2], Nicola Vermes [2], Marco Zaffalon [2], Marco Varini [3], Claudia Wenzel [4], André Fringer [5]



An Application: Counterfactual Analysis in Palliative Cares

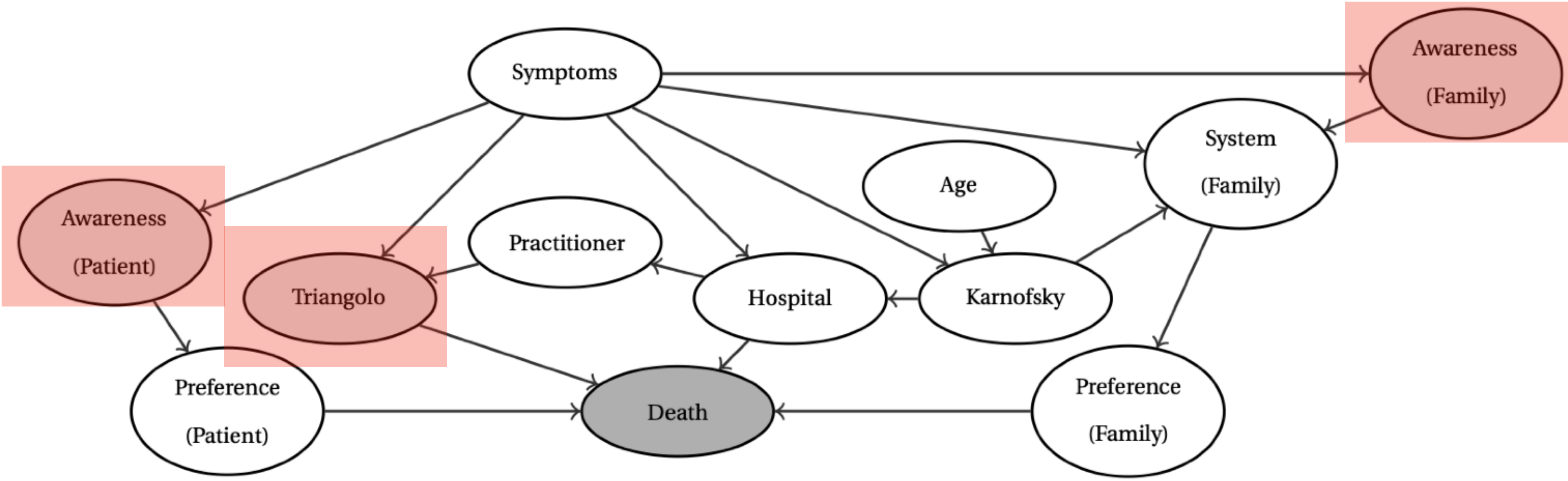
- Most patients prefer to die at home
- But a majority actually die in institutional settings
- Interventions by health care professionals can facilitate dying at home?



An Application: Counterfactual Analysis in Palliative Cares

- Importance of a variable?
- Probability of necessity and sufficiency

$$PNS := P(Y_{X=1} = 1, Y_{X=0} = 0)$$



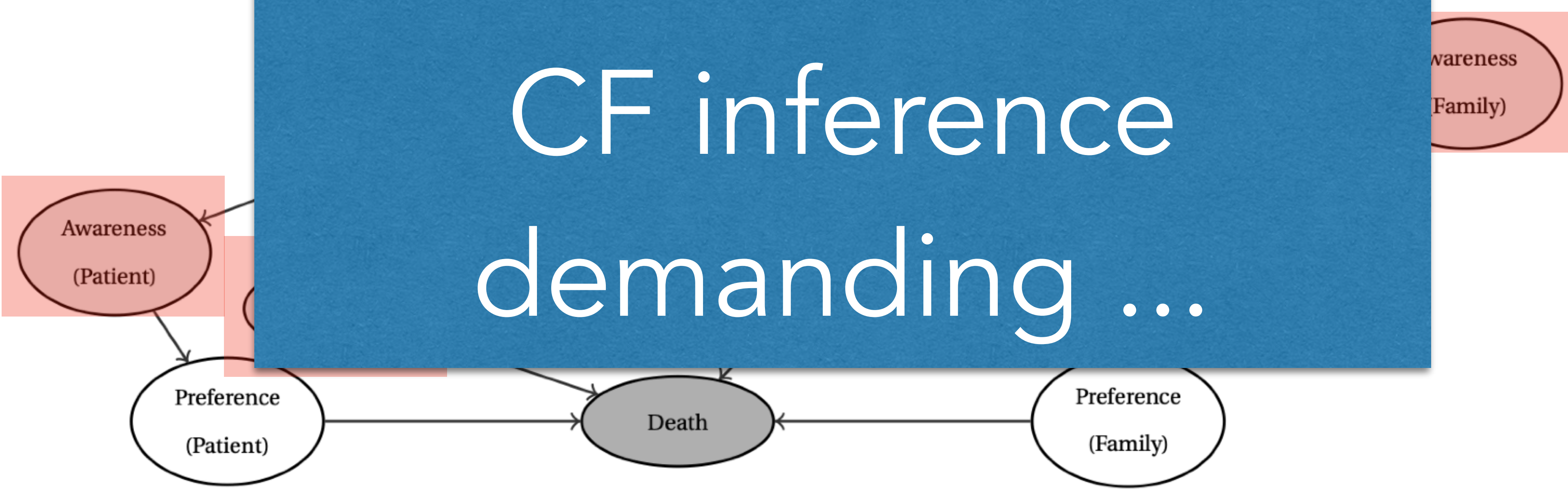
An Application: Counterfactual Analysis in Palliative Cares

- Importance of a variable?
- Probab

P

Small CN but large cardinalities

CF inference demanding ...



Causal Expectation Maximisation (Zaffalon et al., 2021)

- Exogenous variables are always missing (MAR, asystematic, way)
- Expectation Maximisation (Dempster 1977)
  - Random initialisation of P(U)
  - E-step: Missing data completion by expected (fractional) counts
  - M-step: "completed" data to retrain P(U)
  - Iterate until convergence
- EM goes to a (local/global) max of $\log P(\mathscr{D})$

U1	U2	X1	X2	n
*	*	0	0	...
*	*	0	1	...
*	*	1	0	...
*	*	1	1	...

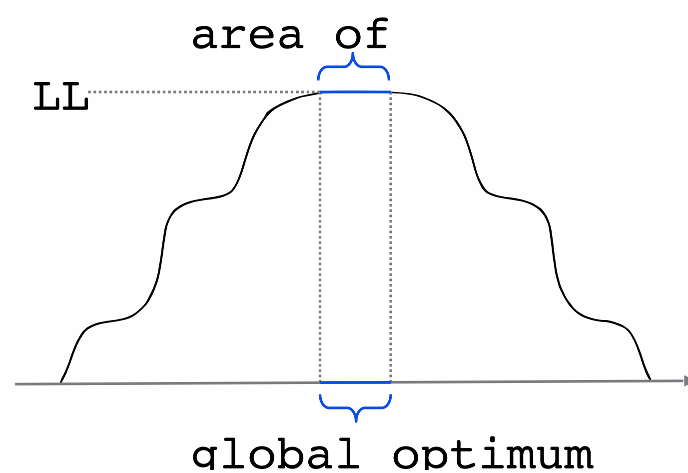
Causal EM: Likelihood Unimodality

- Causal EM reduce should converge to global maxima only the corresponding $P(U)$ belongs to credal set $K(U)$
- Sampling initialisations = sampling of $K(U)$
- For each sample we obtain an inner point

Theorem 1. *Let $\mathscr{K}$ denote the set of quantifications for $\{P(U)\}_{U\in\boldsymbol{U}}$ consistent with the following constraint to be satisfied for each $c \in \mathscr{C}$ and each $\boldsymbol{y}^{(c)}$:*

$$\sum_{\substack{\boldsymbol{u}^{(c)}:f_X(\mathrm{pa}_X)=x \\ \forall X\in\boldsymbol{X}^{(c)}}} \prod_{U\in\boldsymbol{U}^c} P(u) = \prod_{X\in\boldsymbol{X}^{(c)}} \hat{P}(x|\boldsymbol{y}_X^{(c)}), \tag{8}$$

where the values of u, x and $\boldsymbol{y}_X^{(c)}$ are those consistent with $\boldsymbol{u}^{(c)}$ and $\boldsymbol{y}^{(c)}$. If $\mathscr{K} \neq \emptyset$, the log-likelihood in Eq. (7) achieves its global maximum if and only if $\{P(U)\}_{U\in\boldsymbol{U}} \in \mathscr{K}$. If $\mathscr{K} = \emptyset$, the marginal log-likelihood in Eq. (7) can only take values strictly lower than the global maximum.



Causal EM: Guarantees?

- We first reduced causal queries to CN inference
- Causal EM reduces CN inference to (iterated) BN inference
- Identifiable queries? Each sample gives the same values (a numerical alternative to do-calculus)
- Unidentifiable? Each sample as an inner point
- Credible intervals can be derived

Theorem 5. *Let $[a^*, b^*]$ denote the exact probability bounds of a causal query. Say that $\rho := \{r_i\}_{i=1}^n$ are the outputs of n EMCC iterations, while $[a, b]$ is the interval induced by ρ, i.e., $a := \min_{i=1}^n r_i$ and $b := \max_{i=1}^n r_i$. By construction $a^* \leq a \leq b \leq b^*$. The following inequality holds:*

$$P\left(a - \varepsilon L \leq a^* \leq b^* \leq b + \varepsilon L \,\middle|\, \rho\right) = \frac{1 + (1+2\varepsilon)^{2-n} - 2(1+\varepsilon)^{2-n}}{(1-L^{n-2}) - (n-2)(1-L)L^{n-2}}, \tag{13}$$

where $L := (b-a)$ and $\varepsilon := \delta/(2L)$ is the relative error at each extreme of the interval obtained as a function of the absolute allowed error $\delta \in (0, L)$.

Causal EM: Guarantees?

- We first reduced causal queries to CN inference
- Causal EM reduces CN inference to (iterated) BN inference
- Identifiable queries? Each sample gives the same values (a numerical alternative to do-calculus)
- Unidentifiable? Each sample as an inner point
- Credible intervals can be derived

Theorem 5. *Let $[a^*, b^*]$ denote the*

a

a

$$P\left(a - \varepsilon L \leq a^* \leq b^*\right.$$

where $L := (b-a)$ and $\varepsilon := \delta/(2L)$ is
function of the absolute allowed err

Extension relaxing uniformity assumption

Theorem 9. *Assume that the EMCC runs in ρ are distributed as a four parameter beta distribution, i.e., $\pi_i \sim Beta(\alpha, \beta, a^*, b^*)$, for each $i = 1, \ldots, k$. The following*

$$P\left(a - \varepsilon L \leq a^* \leq b^* \leq b + \varepsilon L \,\middle|\, \rho\right) = \frac{\int_0^{\delta_b/2} \int_0^{\delta_a/2} P(x, y; L, \alpha, \beta, k)\, \mathrm{d}x\, \mathrm{d}y}{\int_0^{a+(1-b)} \int_0^{a+(1-b)-y} P(x, y; L, \alpha, \beta, k)\, \mathrm{d}x\, \mathrm{d}y}, \tag{24}$$

where $P(x, y; L, \alpha, \beta, k)$ is equal to

$$\left(\frac{(L+x)^{\alpha}\,{}_2F_1(\alpha, 1-\beta, \alpha+1, \frac{L+x}{L+x+y}) - x^{\alpha}\,{}_2F_1(\alpha, 1-\beta, \alpha+1, \frac{x}{L+x+y})}{\alpha (L+x+y)^{\alpha} B(\alpha, \beta)}\right)^k, \tag{25}$$

Casual EM: Guarantees?

- We first reduced causal queries to CN inference
- Cau
- Ide
  (a n
- Uni
- Cre

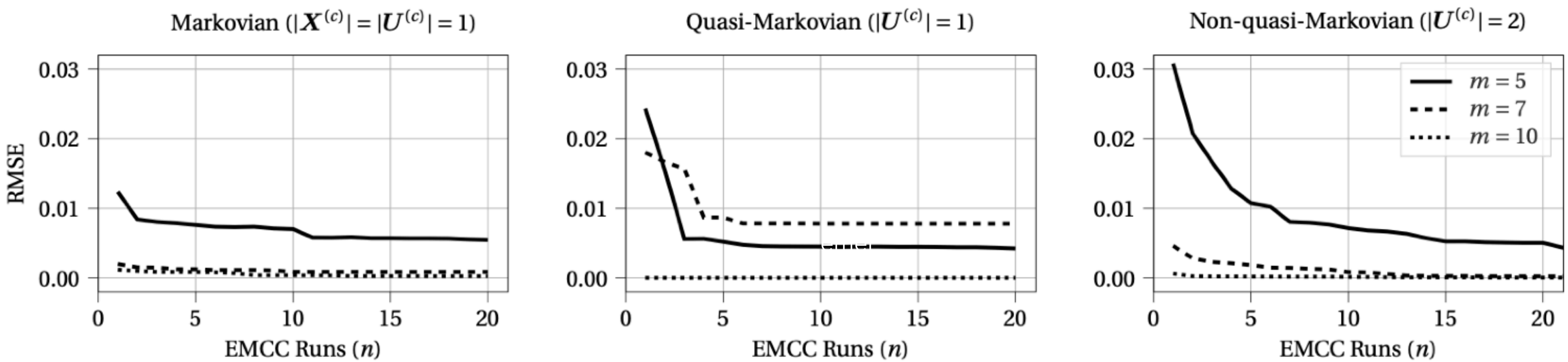
In practice?
20 EM runs to get close to the actual bounds with 95% credibility
For identifiable queries 9 runs to be sure with 99% credibility

Theore
are the
and $b := \max_{i=1} r_i$. *By construction* ... *The following inequality holds:*

$$P\left(a - \varepsilon L \leq a^* \leq b^* \leq b + \varepsilon L \,\middle|\, \rho\right) = \frac{1 + (1+2\varepsilon)^{2-n} - 2(1+\varepsilon)^{2-n}}{(1 - L^{n-2}) - (n-2)(1-L)L^{n-2}}, \quad (13)$$

where $L := (b - a)$ *and* $\varepsilon := \delta/(2L)$ *is the relative error at each extreme of the interval obtained as a function of the absolute allowed error* $\delta \in (0, L)$.

Causal EM: Experiments



PNS for artificial SMCs: quick convergence
(= much faster than direct CN approach)

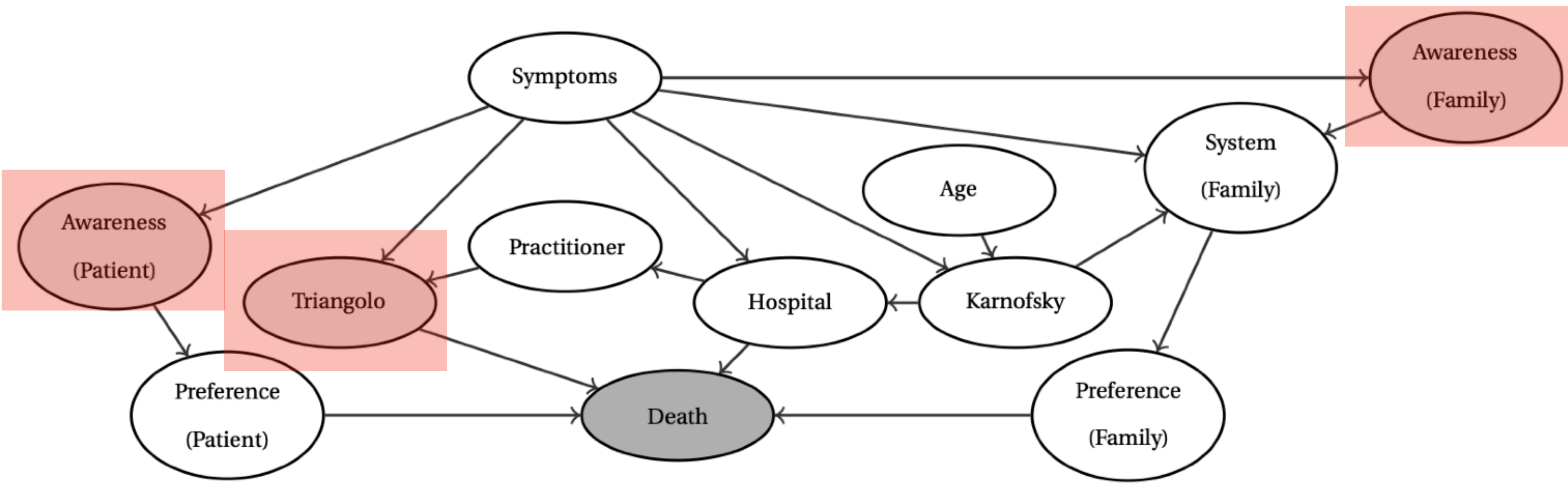
Counterfactual Analysis in Palliative Cares by Causal EM

- Importance of a variable?
- Probability of necessity and sufficiency

$$PNS := P(Y_{X=1} = 1, Y_{X=0} = 0)$$

- 15 EM runs before convergence

PNS(Family_Awareness) ∈ [0.06,0.10]



PNS(Patient_Awareness) ∈ [0.03,0.10]

PNS(Triangolo) ∈ [0.30,0.31]

Coun

- Im
- Pr
- 15

.06,0.10]

Aware (Patie

eness mily)

One should act on Triangolo first: for instance, by making Triangolo available to all patients, we should expect a reduction of people at the hospital by 30%

This would save money too, and would allow politicians to do economic considerations as to which amount it is even economically profitable to fund Triangolo, and have patients die at home, rather than spending more to have patients die at the hospital

PNS(Patient_Awareness) ∈ [0.03,0.10]

PNS(Triangolo) ∈ [0.30,0.31]

Causal Analysis from **Biased** Data

- Selective data acquisition
  (untreated M and treated F missing)

Treatment X	Recovery Y	Gender Z	counts
0	0	0	2
1	0	0	41
0	1	0	114
1	1	0	313
0	0	1	107
1	0	1	109
0	1	1	13
1	1	1	1

[Müeller et al., 2022]

Causal Analysis from **Biased** Data

- Selective data acquisition
  (untreated M and treated F missing)
- A (Boolean) **selector** variable $S \equiv (X \neq Z)$

Treat, X	Recover y	Gender Z	Selector S	counts
*	*	*	0	2
1	0	0	1	41
*	*	*	0	114
1	1	0	1	313
0	0	1	1	107
*	*	*	0	109
0	1	1	1	13
*	*	*	0	1

[Müeller et al., 2022]

Causal Analysis from **Biased** Data

- Selective data acquisition
  (untreated M and treated F missing)
- A (Boolean) **selector** variable $S \equiv (X \neq Z)$
- Assume we know $n(S=0) \propto P(S=0)$

Treat, X	Recovery Y	Gender Z	Selector S	counts
1	0	0	1	41
1	1	0	1	313
0	0	1	1	107
0	1	1	1	13
*	*	*	0	226

[Müeller et al., 2022]

Causal Analysis from **Biased** Data

- Selective data acquisition
  (untreated M and treated F missing)
- A (Boolean) **selector** variable $S \equiv (X \neq Z)$
- Assume we know $n(S=0) \propto P(S=0)$
- Interventional queries with bias?
- Do calculus for selection bias
  Barenboim & Tian (AAAI, 2015)

Treat, X	Recovery Y	Gender Z	Selector S	counts
1	0	0	1	41
1	1	0	1	313
0	0	1	1	107
0	1	1	1	13
*	*	*	0	226

[Müeller et al., 2022]

Recovering Causal Effects from Selection Bias

Elias Bareinboim*
Computer Science Department
University of California, Los Angeles
Los Angeles, CA. 90095
eb@cs.ucla.edu

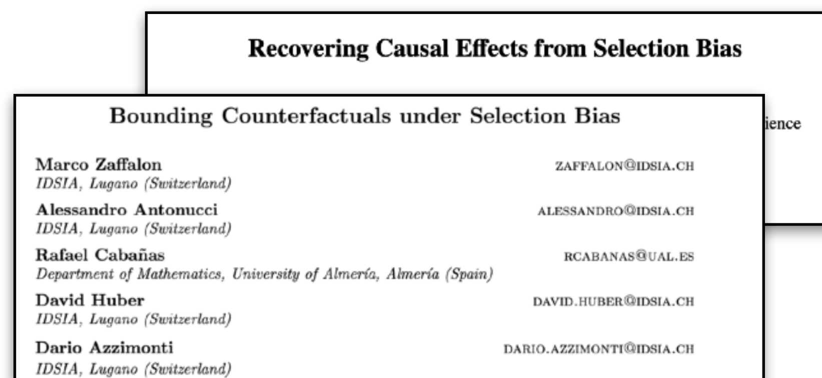
Jin Tian*
Department of Computer Science
Iowa State University
Ames, IA. 50011
jtian@iastate.edu

Causal Analysis from **Biased** Data

- Selective data acquisition
  (untreated M and treated F missing)
- A (Boolean) **selector** variable $S \equiv (X \neq Z)$
- Assume we know $n(S=0) \propto P(S=0)$
- Interventional queries with bias?
- Do calculus for selection bias
  Barenboim & Tian (AAAI, 2015)
- Unidentifiable queries?
- Our EM(CC) can be used for that!

Treat, X	Recovery Y	Gender Z	Selector S	counts
1	0	0	1	41
1	1	0	1	313
0	0	1	1	107
0	1	1	1	13
*	*	*	0	226

[Müeller et al., 2022]


Recovering Causal Effects from Selection Bias

ience

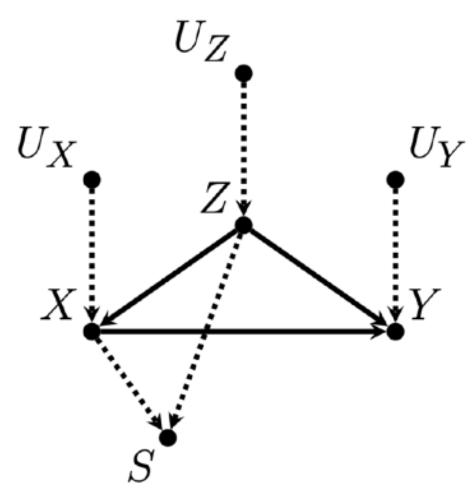
Bounding Counterfactuals under Selection Bias

Marco Zaffalon ZAFFALON@IDSIA.CH
IDSIA, Lugano (Switzerland)
Alessandro Antonucci ALESSANDRO@IDSIA.CH
IDSIA, Lugano (Switzerland)
Rafael Cabañas RCABANAS@UAL.ES
Department of Mathematics, University of Almería, Almería (Spain)
David Huber DAVID.HUBER@IDSIA.CH
IDSIA, Lugano (Switzerland)
Dario Azzimonti DARIO.AZZIMONTI@IDSIA.CH
IDSIA, Lugano (Switzerland)

Back to the Biased Data ...

- S determined by an equation, a SCM!

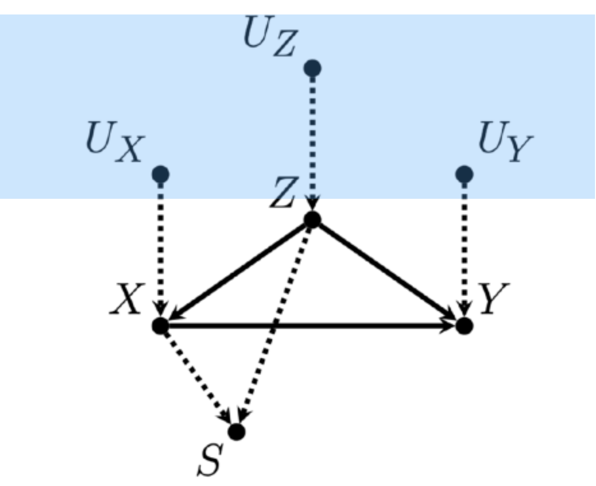
UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226



Back to the Biased Data ...

- S determined by an equation, a SCM!
- CN approach? No, $S = 1$ induces relations between $P(U)$'s in the CN

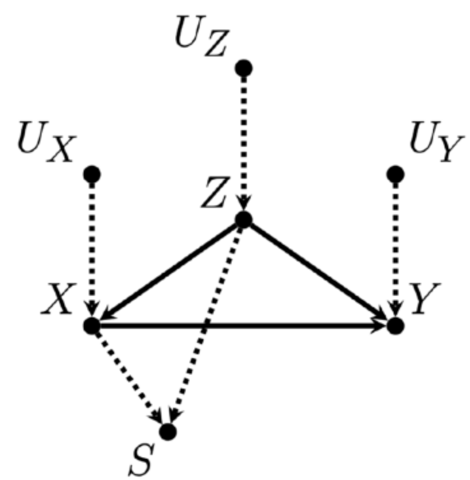
UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226



Back to the Biased Data ...

- S determined by an equation, a SCM!
- CN approach? No, $S = 1$ induces relations between $P(U)$'s in the CN
- EM? Maybe, but "non-rectangular" missingness, might kill unimodality ...
- Convergence to max preserved? (hence inner points of $[\underline{P}, \overline{P}]$)

UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226

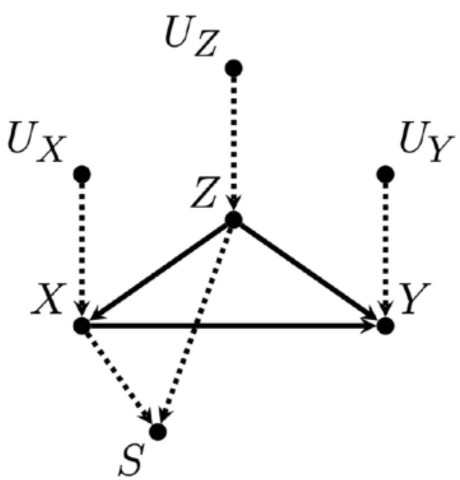


Back to the Biased Data ...

- S determined by an equation, a SCM!
- CN approach? No, $S = 1$ induces relations between $P(U)$'s in the CN
- EM? Maybe, but "non-rectangular" missingness, might kill unimodality ...
- Convergence to max preserved? (hence inner points of $[\underline{P}, \overline{P}]$) Yes!

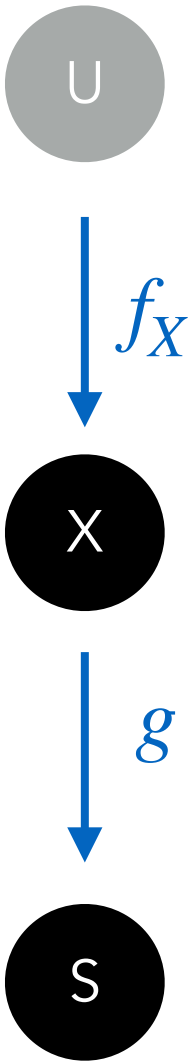
UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226

Theorem 4 *As a function of* $\{P(U)\}_{U \in \boldsymbol{U}}$*, the log-likelihood in Eq. (7) has no local maxima and a global maximum equal to the value* LL^* *in Eq. (6). Such a maximum is achieved if and only if the* M*-compatibility constraints in Eqs. (8) and (9) are satisfied.*



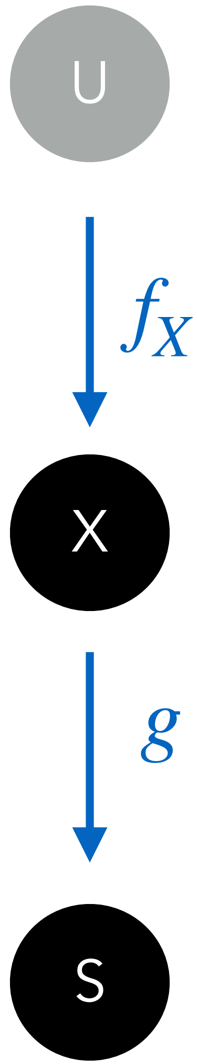
Sketch of the proof

UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226



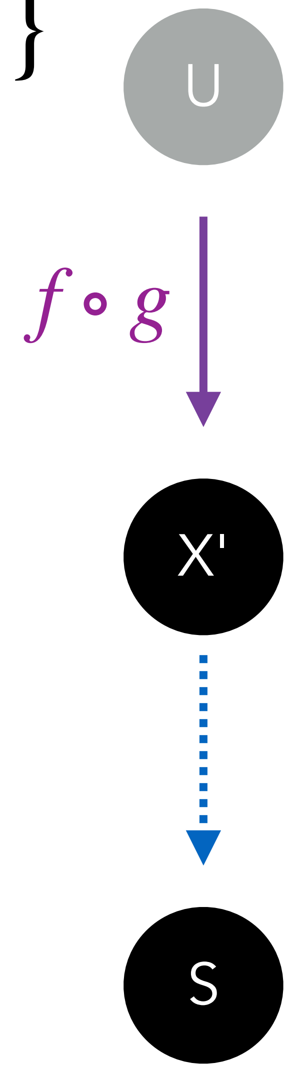
Sketch of the proof

UX	UY	UZ	X	Y	Z	S	n
*	*	*	1	0	0	1	41
*	*	*	1	1	0	1	313
*	*	*	0	0	1	1	107
*	*	*	0	1	1	1	13
*	*	*	*	*	*	0	226



$$\Omega_{X'} := \Omega_X \cup \{ * \}$$

UX	UY	UZ	X'	S	n
*	*	*	{100}	1	41
*	*	*	{110}	1	313
*	*	*	{001}	1	107
*	*	*	{011}	1	13
*	*	*	{*}	0	226

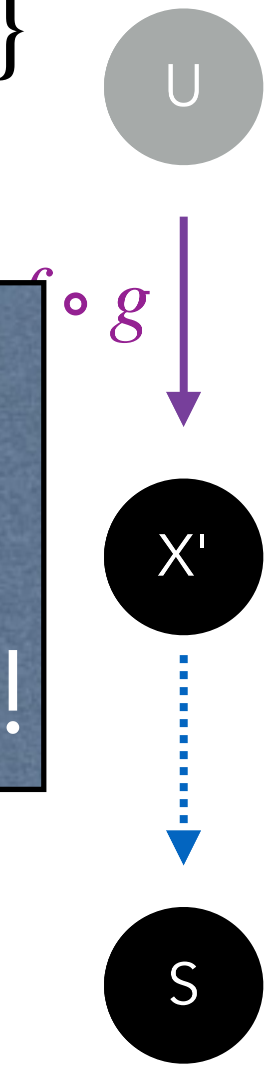


Sketch of the proof

$$\Omega_{X'} := \Omega_X \cup \{ * \}$$

UX	UY	UZ	X	Y	Z	S	n
*							
*							
*							
*							
*	*	*	*	*	*	0	226

UX	UY	UZ	X'	S	n
*	*	*	{*}	0	226

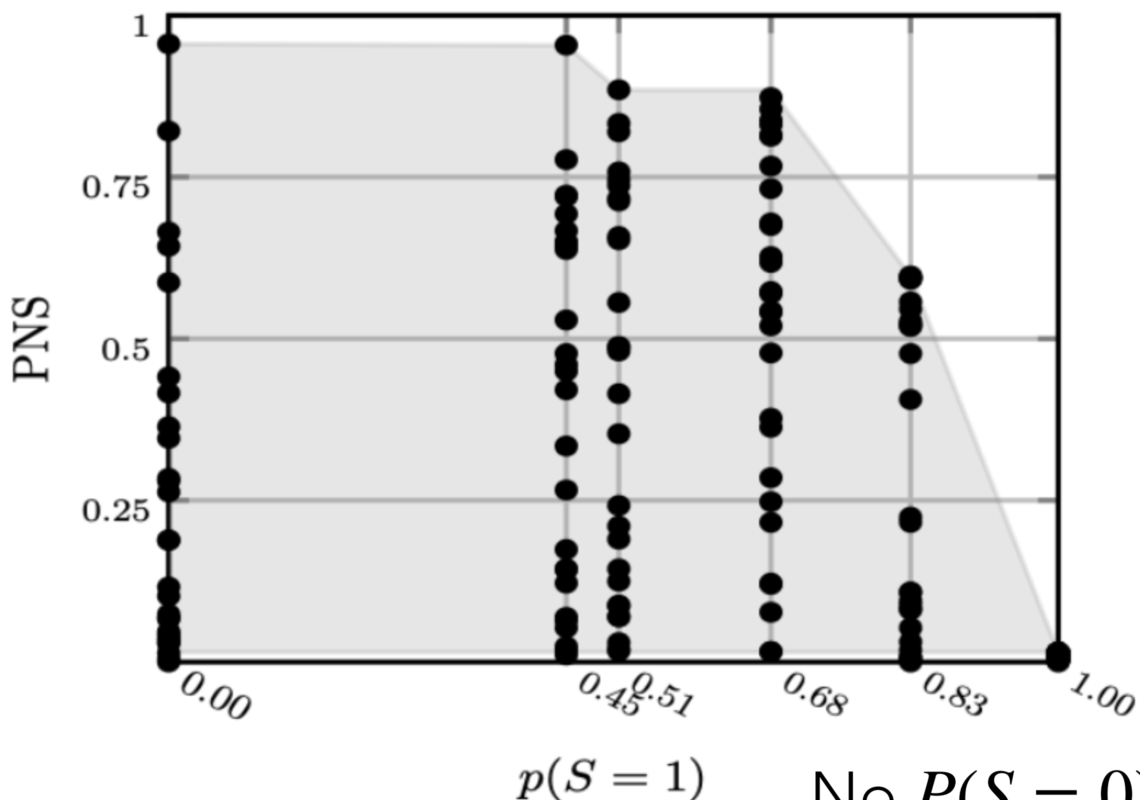


Both SCMs have the same lik
The new one has rectangular
missingness and EM converges!

Counterfactual Bounds for Biased Data

Probability of Necessity and Sufficiency

$$\text{PNS} := P(Y_{X=0} = 0, Y_{X=1} = 1)$$



Treatment X	Recovery Y	Gender Z	counts
0	0	0	2
1	0	0	41
0	1	0	114
1	1	0	313
0	0	1	107
1	0	1	109
0	1	1	13
1	1	1	1

[Müeller et al., 2022]

Real Bounds (Gray)
EM Points (Black)

No $P(S=0)$? limit or expert bound

Current Work: Hybrid Data

Learning to Bound Counterfactual Inference from Observational, Biased and Randomised Data

Marco Zaffalon[a], Alessandro Antonucci[a,*], Rafael Cabañas[b], David Huber[a]

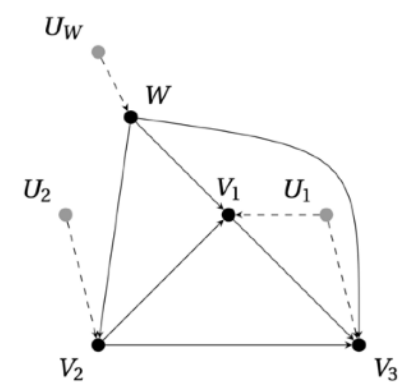
[a] *IDSIA, Lugano (Switzerland)*
[b] *Department of Mathematics, University of Almería, Almería (Spain)*

Study	Treatment	Gender	Survival	Counts
interventional	do(drug)	female	survived	489
	do(drug)	female	dead	511
	do(drug)	male	survived	490
	do(drug)	male	dead	510
	do(no drug)	female	survived	210
	do(no drug)	female	dead	790
	do(no drug)	male	survived	210
	do(no drug)	male	dead	790
observational	drug	female	survived	378
	drug	female	dead	1022
	drug	male	survived	980
	drug	male	dead	420
	no drug	female	survived	420
	no drug	female	dead	180
	no drug	male	survived	420
	no drug	male	dead	180

Table 1: Data from interventional and observational studies on the potential effects of a drug on patients affected by a deadly disease.

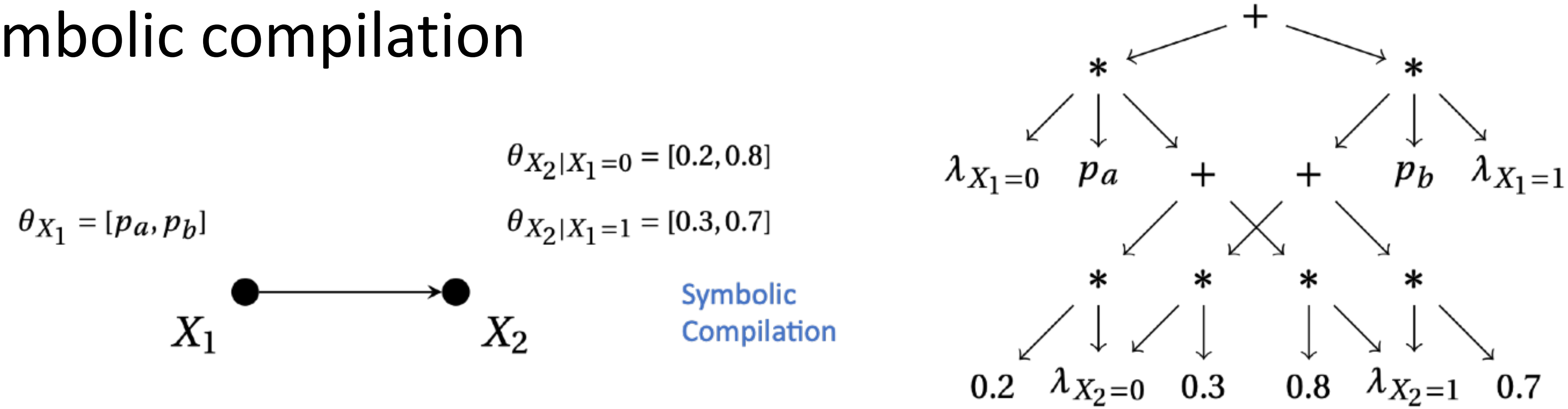
Treatment	Gender	Survival	W	Counts
drug	female	survived	drug	489
drug	female	dead	drug	511
drug	male	survived	drug	490
drug	male	dead	drug	510
no drug	female	survived	no drug	210
no drug	female	dead	no drug	790
no drug	male	survived	no drug	210
no drug	male	dead	no drug	790
drug	female	survived	$w_\emptyset$	378
drug	female	dead	$w_\emptyset$	1022
drug	male	survived	$w_\emptyset$	980
drug	male	dead	$w_\emptyset$	420
no drug	female	survived	$w_\emptyset$	420
no drug	female	dead	$w_\emptyset$	180
no drug	male	survived	$w_\emptyset$	420
no drug	male	dead	$w_\emptyset$	180

Table 2: A merged version of the two datasets in Table 1 with the index variable W.



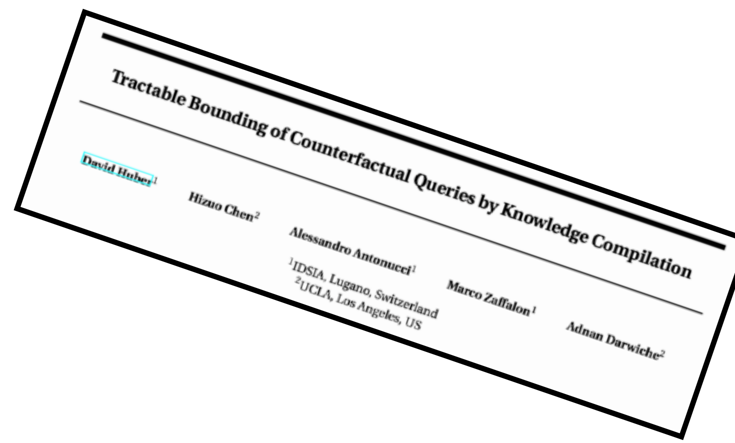
Current Work: Symbolic Knowledge Compilation (TPM 2023)

- Joint work with Adnan Darwiche and Hizuo Chen
- Our EM requires many (BN) queries
- Equations remain constant
- Compile BN once, use many times
- Symbolic compilation



Current Work: Symbolic Knowledge Compilation (TPM 2023)

- Joint work with Adnan Darwiche and Hizuo Chen
- Our EM requires many (BN) queries
- Equations remain constant
- Compile BN once, use many times
- Symbolic compilation

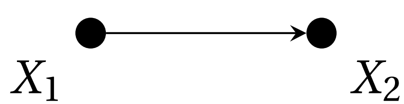

Tractable Bounding of Counterfactual Queries by Knowledge Compilation

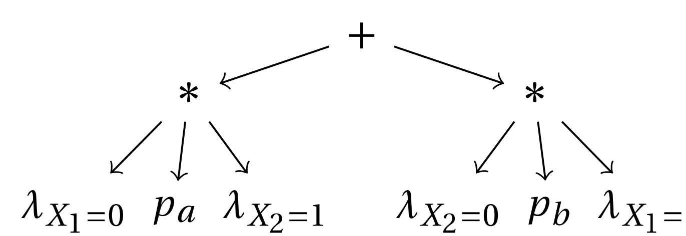
David Huber[1] Hizuo Chen[2] Alessandro Antonucci[1] Marco Zaffalon[1] Adnan Darwiche[2]

[1]IDSIA, Lugano, Switzerland
[2]UCLA, Los Angeles, US

$\theta_{X_2|X_1=0} = [0.0, 1.0]$

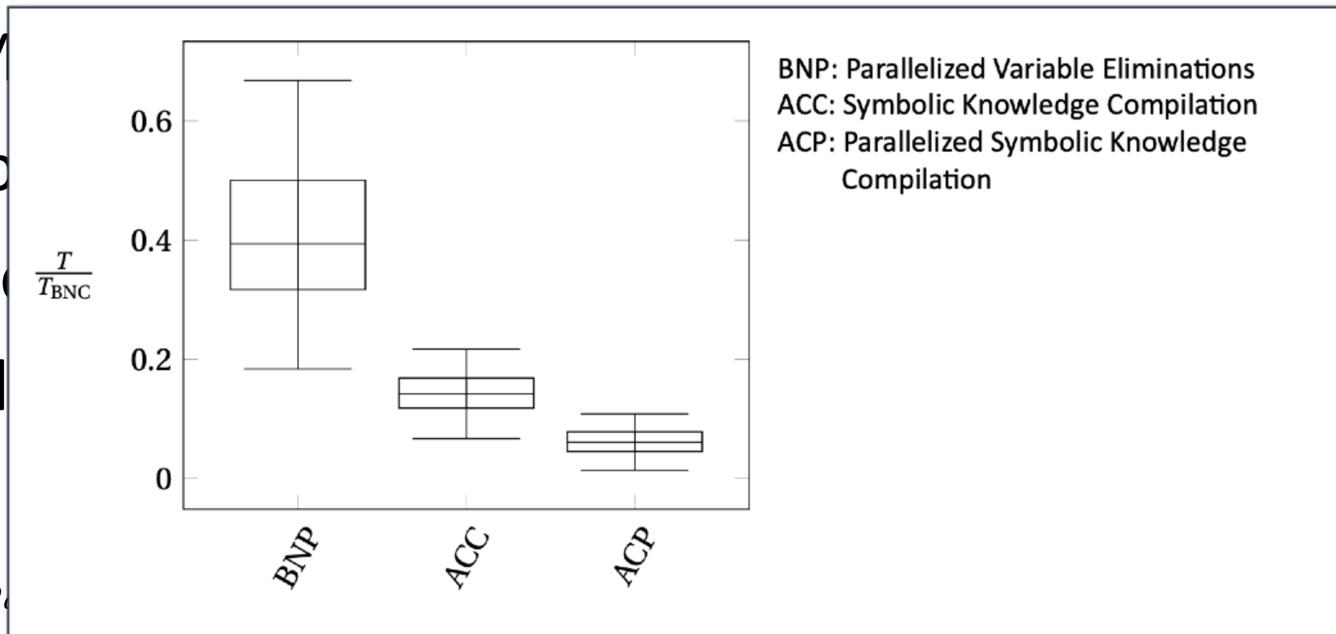
$\theta_{X_1} = [p_a, p_b]$ $\theta_{X_2|X_1=1} = [1.0, 0.0]$

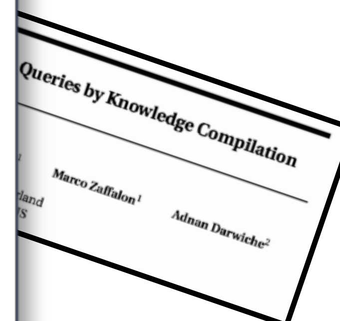


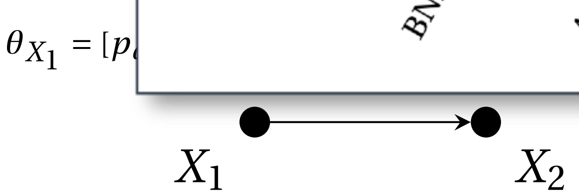


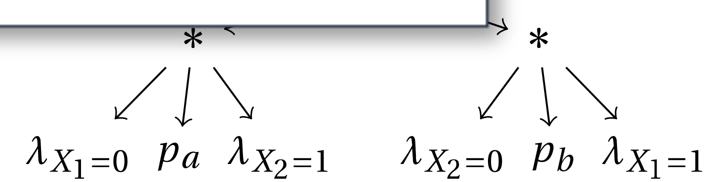
Current Work: Symbolic Knowledge Compilation (TPM 2023)

- Joint work with Adnan Darwiche and Hizuo Chen
- Our EM
- Equatio
- Compil
- Symbol



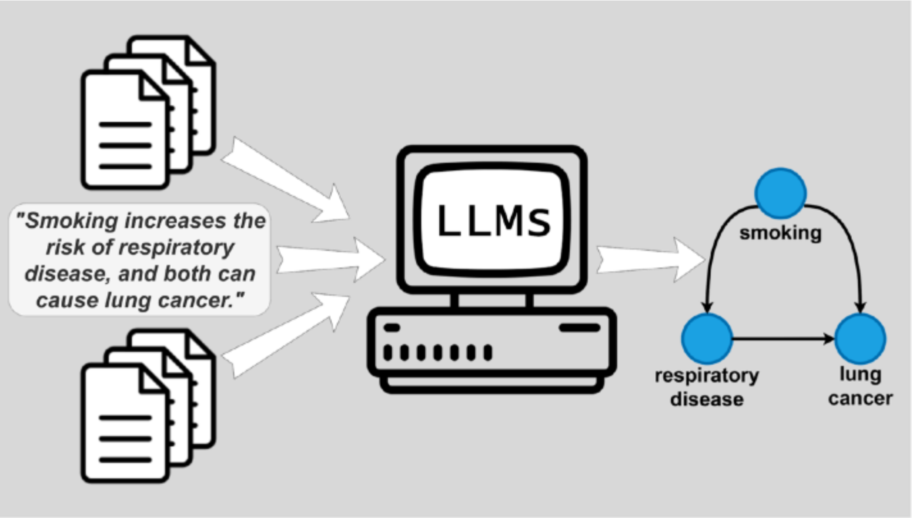






Current Work: Causal Graphs by LLMs (under preparation)

- GPT parsing causal statements in natural language



- Link with IPs? Multiple causal graphs might be returned!
- Many recent papers on bounding counterfactual wrt **ignorance** about the causal structure (credal structures?)

Conclusions

- Causality theories have an intimate connection with credal models
- Past research about CNs might offer new tools for causal analysis
- CNs offer formalism for a deeper SCMs understanding
- Lot of work to be done, causal ML/RL just at the beginning!
- Our current directions are:
  - Canonical models
  - Continuous Variables
  - XAI (Counterfactual Explanations)
  - "Credal EM" propagating credal initialisations?

Conclusions

- Causality theories have an intimate connection w... l models
- Past research about CNs might offer new ... alysis
- CNs offer formalism for a deeper S...g
- Lot of works has to be done ...d reinforcement) learning is just at the b...
- Our current dire...
  - Canoni...
  - H...
  - Co... & Parallelisation
  - Cont...uous Variables
  - Learning Causal Graphs (GPT)

slides