Causal Inference Theory with Information Algebras (2/2): the Information Dependency Model

Causality in Practice, Institute Pascal, 12-16 June 2023

Benjamin Heymann, Michel De Lara, Jean-Philippe Chancelier

BH: Criteo AI Lab, Paris, France
MDL and JPC : Cermics, École des Ponts, Marne-la-Vallée, France

In case you forgot a few bits from the first part of the talk...

Refresher from the first half of the talk

- $\mathbb{H} = \Omega \times_{a \in \mathbb{A}} \mathbb{U}_a$ is the common product domain
- λ_a is $\mathcal{I}_a$-measurable:

$$\lambda_a : (\mathbb{H}, \mathcal{H}) \to (\mathbb{U}_a, \mathcal{U}_a)$$

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{I}_a$$

for all $a \in \mathbb{A}$.

How do we relate Witsenhausen's framework and causality?

Please welcome the Information Dependency Model!

Relation with SCMs[1]

An SCM formulation takes the form

- $(\lambda_a)_{a \in \mathbb{A}}$: assignments
- $P : \mathbb{A} \to 2^{\mathbb{A}}$: parental mapping

$$U_a(\omega) = \lambda_a(\omega_a, U_{P(a)}(\omega)) \quad \forall \omega \in \Omega \quad \forall a \in \mathbb{A} \ .$$

[1]Structural Causal Models

Information Dependency Model (IDM)

1. A product space $\mathbb{H} = \prod_{a \in \mathbb{A}} \Omega_a \times \mathbb{U}_a$;
2. A collection $(\mathcal{I}_a)_{a \in \mathbb{A}}$ of subalgebras of $\mathcal{H}$;

Information Dependency Model (IDM)

1. A product space $\mathbb{H} = \prod_{a \in \mathbb{A}} \Omega_a \times \mathbb{U}_a$;
   - $\mathcal{H} = \bigotimes_{b \in \mathbb{A}} \mathcal{F}_b \otimes \mathcal{U}_b$ is the product algebra of $\mathbb{H}$
2. A collection $(\mathcal{I}_a)_{a \in \mathbb{A}}$ of subalgebras of $\mathcal{H}$;
   - $\mathcal{I}_a \subset \mathcal{F}_a \otimes \bigotimes_{b \in \mathbb{A} \setminus \{a\}} \mathcal{U}_b$

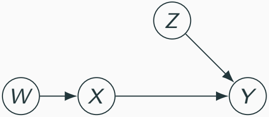
Information Dependency Model (IDM)

1. A product space $\mathbb{H} = \prod_{a \in \mathbb{A}} \Omega_a \times \mathbb{U}_a$;
   - $\mathcal{H} = \bigotimes_{b \in \mathbb{A}} \mathcal{F}_b \otimes \mathcal{U}_b$ is the product algebra of $\mathbb{H}$
2. A collection $(\mathcal{I}_a)_{a \in \mathbb{A}}$ of subalgebras of $\mathcal{H}$;
   - $\mathcal{I}_a \subset \mathcal{F}_a \otimes \bigotimes_{b \in \mathbb{A} \setminus \{a\}} \mathcal{U}_b$

The SCM is now defined by the $\mathcal{I}_a$-measurability conditions

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{I}_a \quad \forall a \in \mathbb{A}$$

How is parentality encoded?



X is a parent of Y

In SCMs, a random variable gets the arguments of its assignment function from its parents:

$$Y = \lambda_Y(\omega, X(\omega), Z(\omega)).$$

conditional precedence = Adding a pinch of flexibility

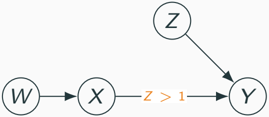
Context-Specific independence (CSI)

$$A \perp\!\!\!\perp B \text{ when } C = 1$$

- A generalization of indepence between random variables.
- Used in many applications.

$\rightarrow$ CSI for $\mathbb{FREE}$ with the Information Dependency Model (IDM)

How is parentality encoded?



X is a parent of Y
UNLESS $Z > 1$

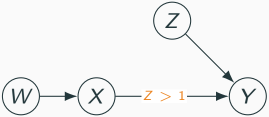
(W, H)-Conditional parentality

Definition

Let $W \subset \mathbb{A}$, $H \subset \mathbb{H}$ and $a \in \mathbb{A}$. The conditional parents set $\mathcal{E}^{W,H}a$ is the smallest subset $B \subset \mathbb{A}$ such that

$$\mathcal{I}_a \cap H \subset \mathcal{H}_{B \cup W} \cap H$$

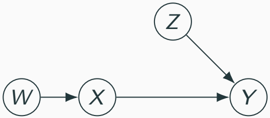
Conditional parentality: conditioning on the context



X is a parent of Y UNLESS $Z > 1$

$$\mathcal{E}^{\emptyset,\{Z>1\}} Y = \{Z\}$$

Conditional parentality: conditioning on a variable



$$\mathcal{E}^{\emptyset,\mathbb{H}}Y = \{Z, X\}$$

$$\mathcal{E}^{\{Z\},\mathbb{H}}Y = \{X\}$$

$$\mathcal{E}^{\{X\},\mathbb{H}}Y = \{Z\}$$

$$\mathcal{E}^{\{X,Z\},\mathbb{H}}Y = \emptyset$$

$\rightarrow$ an alternative way of expressing that a path is *blocked*. Very handy for algebric manipulations.

(W, H)-Conditional parentality

Definition

Let $W \subset \mathbb{A}$, $H \subset \mathbb{H}$ and $a \in \mathbb{A}$. The conditional parent set $\mathcal{E}^{W,H}a$ is the smallest subset $B \subset \mathbb{A}$ such that

$$\mathcal{I}_a \cap H \subset \mathcal{H}_{B \cup W} \cap H$$

We denote by $\bar{B}$ (or $\bar{B}^{W,H}$) the topological closure of B, which is the smallest subset of $\mathbb{A}$ that contains B and its own parents under $\mathcal{E}^{W,H}$.

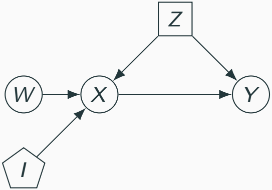
Modeling interventions

Atomic intervention

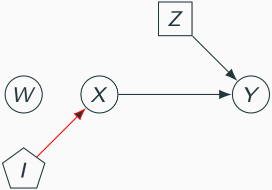
$$\mathcal{I}_X \leftarrow (\mathcal{I}_X \otimes \{I = 0\}) \cup (\{\emptyset, \mathbb{H}\} \otimes \{I = 1\})$$

Atomic intervention

$$\mathcal{I}_X \leftarrow \underbrace{(\mathcal{I}_X \otimes \{I = 0\})}_{\text{normal regime}} \cup \underbrace{(\{\emptyset, \mathbb{H}\} \otimes \{I = 1\})}_{\text{intervention}}$$



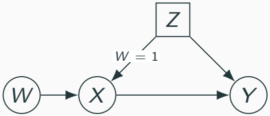
Intervention not activated, $I = 0$



Intervention activated, $I = 1$

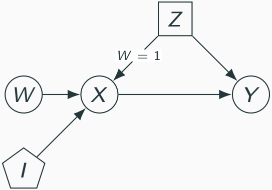
Application[2]

Can we estimate $\Pr(Y \mid \mathrm{do}(X))$ from the observational distribution?



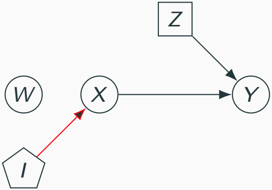
[2]Example taken from Tikka et al. 2019

Application[3] (hand waving style)

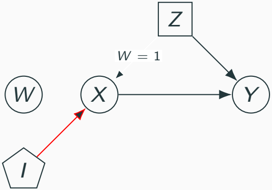


We picture the original graph, with the additional node

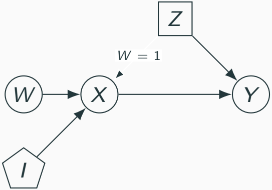
[3]We do it in the graphical world because it is possible to do so. Note however that Information Dependency Models can deal with more complex situations



we start with the situation $I = 1$



By independence, we can set $W = 1$



We now set $I = 0$

Hence $P(Y \mid \mathrm{do}(X)) = P(Y \mid X, W = 1)$.

Topological separation

Topological separation

Definition (Topological Separation)

We say that B and C are (conditionally) *topologically separated* (wrt (W, H)), and write

$$B \perp\!\!\!\perp_t C \mid (W, H),$$

if there exists $W_B, W_C \subset W$ such that

$$W_B \sqcup W_C = W \text{ and } \overline{B \cup W_B} \cap \overline{C \cup W_C} = \emptyset$$

Do-calculus in the IDM

Definition (Topological Separation)

We say that B and C are (conditionally) *topologically separated* (wrt (W, H)), and write

$$B \underset{t}{\perp\!\!\!\perp} C \mid (W, H),$$

if there exists $W_B, W_C \subset W$ such that

$$W_B \sqcup W_C = W \text{ and } \overline{B \cup W_B} \cap \overline{C \cup W_C} = \emptyset$$

Theorem (Do-calculus)

$$Y \underset{t}{\perp\!\!\!\perp} Z \mid (W, H) \Longrightarrow \Pr(U_Y \mid U_W, U_{\bar{Z}}, H) = \Pr(U_Y \mid U_W, H)$$

Illustration

Example 1

Are Y_1 and Y_2 independent when conditioned on W?

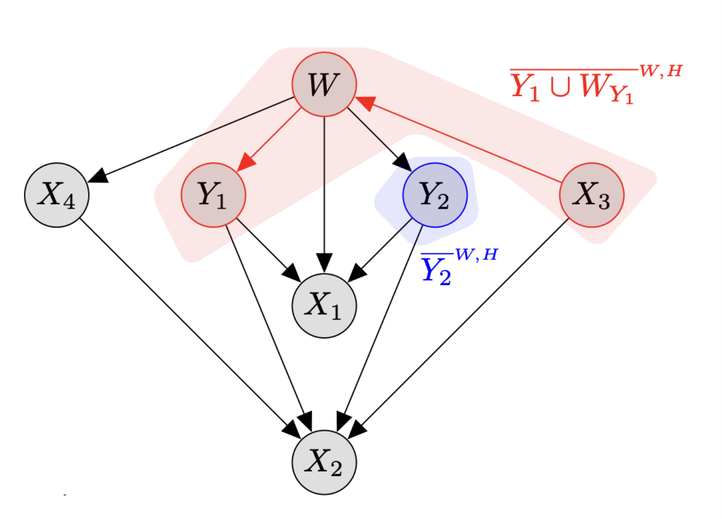


Figure 1: The split of W is a piece of information that can be usefull.

Example 2

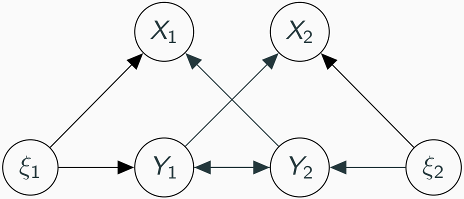


Figure 2: Are the Y's independent when conditioning on the X's ?

Example 2

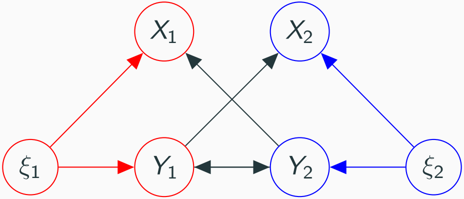


Figure 3: Let $W_{X_i} = Y_i$, for $i = 1, 2$. The closure of $X_1 \cup Y_1$ (resp. $X_2 \cup Y_2$), with the edges followed to build the closure, is in red (resp. blue).

Non-atomic interventions for free

Type	Strategy	$P(x \mid pa_x, u_x; \sigma_X)$	
Idle	$\emptyset$	(unaltered)	
Atomic/*do*	$do(X = x')$	$\delta(x, x')$	(4)
Conditional	$do(X = g(pa_x^*))$	$\delta(x, g(pa_x^*))$	(5)
Stochastic/Random	$P^*(X \mid pa_x^*)$	$P^*(x \mid pa_x^*)$	(6)

Figure 4: From [Correa2020]

Topological separation extends d-separation to a more general settings

Topological separation and d-separation

Theorem

Let $(\mathcal{V}, \mathcal{E})$ be a graph, that is, $\mathcal{V}$ is a set and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, and let $W \subset \mathcal{V}$ be a subset of vertices, we have the equivalence

$$b \underset{t}{\perp\!\!\!\perp} c \mid W \iff b \underset{d}{\perp\!\!\!\perp} c \mid W \qquad \big(\forall b, c \in W^{\mathsf{c}}\big)$$

Proofing toolbox: binary relations

Binary relation

We can define:

- Conditional parentality relation
- Conditional ancestry relation
- Conditional common cause relation
- Conditional "cousinhood relation"
- t-separation relation

An illustration of equational reasoning

Proof We have that

$$
\begin{aligned}
&\Delta_{W^c}\big(\Delta \cup (\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W\big)\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\big(\Delta \cup \mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\big)\Delta_{W^c} \\
&\quad = \Delta_{W^c}\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\Delta_{W^c} && \text{(by developing)} \\
&\qquad \cup \Delta_{W^c}\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\big(\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\big)\Delta_{W^c} \\
&\qquad \cup \Delta_{W^c}\big((\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W\big)\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\Delta_{W^c} \\
&\qquad \cup \Delta_{W^c}\big((\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W\big)\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\big(\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\big)\Delta_{W^c} \\
&\quad = \Delta_{W^c}\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\Delta_{W^c} \\
&\qquad \cup \Delta_{W^c}\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\Delta_{W^c} && \text{(as } \mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\mathcal{C}^W = \mathcal{C}^W \text{ by (34c))} \\
&\qquad \cup \Delta_{W^c}(\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W\mathcal{E}^{-W*}\mathcal{E}^{W*}\Delta_{W^c} && \text{(also by (34c))} \\
&\qquad \cup \Delta_{W^c}(\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\Delta_{W^c} && \text{(also by (34c) applied twice)} \\
&\quad = \Delta_{W^c}\big(\mathcal{B}^W \cup \mathcal{K}^W\big)\mathcal{C}^W\big(\mathcal{B}^{-W} \cup \mathcal{K}^W\big)\Delta_{W^c} && \text{(by (34d) and (34e))} \\
&\qquad \cup \Delta_{W^c}\big(\mathcal{B}^W \cup \mathcal{K}^W\big)(\mathcal{B}^{-W} \cup \mathcal{K}^W)\Delta_{W^c} && \text{(by (34e))} \\
&\qquad \cup \Delta_{W^c}(\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W\big(\mathcal{B}^{-W} \cup \mathcal{K}^W\big)\Delta_{W^c} && \text{(by (34d))} \\
&\qquad \cup \Delta_{W^c}(\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\Delta_{W^c} \\
&\quad = \Delta_{W^c}(\mathcal{B}^W \cup \mathcal{K}^W)\mathcal{C}^W(\mathcal{B}^{-W} \cup \mathcal{K}^W)\Delta_{W^c}\ .
\end{aligned}
$$

This ends the proof. ∎

Conclusion

Summary

- **IDM**, as a generalization of causal graphs/an alternative language to describe causal dependencies
- **Topological separation**, as an alternative definition of d-separation

Making the case for Information Dependency Model (IDM)

- Unlock **mathematical toolboxes**
- **Unifying, generalizing and versatile** framework for causality
- Elegant style of expression and proof : **equational reasoning**
  - compositionality
  - binary relations
- Potential to **bridge** causality, game theory, control and Reinforcement Learning

Some References

H. S. Witsenhausen.
On information structures, feedback and causality.
SIAM J. Control, 9(2):149–160, May 1971.

S. Tikka, A. Hyttinen, and J. Karvanen.
Identifying causal effects via context-specific independence relations.
Proceedings of the AAAI Conference on Artificial Intelligence, 2019.

J. Correa, E. Bareinboim
A Calculus for Stochastic Interventions:Causal Effect Identification and Surrogate Experiments.
In *Advances in Neural Information Processing Systems*, pages 2804–2814, 2019.

B. Heymann, M. De Lara, J. P. Chancelier.
Kuhn's equivalence theorem for games in product form,
In *Games and Economic Behavior*, Volume 135, 2022