Causality with Information Algebras

Causality Discussion Group, October 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

σ-algebra

- A *set algebra* over a set $\mathbb{D}$ is a subset $\mathcal{D} \subset 2^{\mathbb{D}}$, containing $\mathbb{D}$, and which is **stable** under complement, finite union and intersection
- A *σ-algebra* over a set $\mathbb{D}$ is a subset $\mathcal{D} \subset 2^{\mathbb{D}}$, containing $\mathbb{D}$, and which is stable under complement and **countable** union and intersection

Agents, action spaces and Nature space

- Let A be a (finite or infinite) set, whose elements are called agents (or decision-makers)
- With each agent $a \in A$ is associated a measurable space

$$\Big(\underbrace{\mathbb{U}_a}_{\substack{\text{set of}\\ \text{actions}\\ \text{for agent } a}} , \underbrace{\mathcal{U}_a}_{\substack{\sigma\text{-algebra}\\ \subset 2^{\mathbb{U}_a}}} \Big)$$

- With Nature is associated a measurable space

$$(\Omega, \mathcal{F})$$

(at this stage of the presentation, we do not need to equip $(\Omega, \mathcal{F})$ with a probability distribution, as we only focus on **information**)

The configuration space is a product space

Configuration space

The configuration space is the product space

$$\mathbb{H} = \prod_{a \in A} \mathbb{U}_a \times \Omega$$

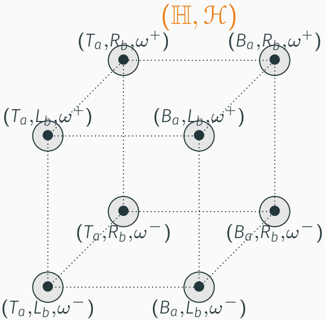
equipped with the product σ-algebra, called configuration σ-algebra

$$\mathcal{H} = \bigotimes_{a \in A} \mathcal{U}_a \otimes \mathcal{F}$$

so that $(\mathbb{H}, \mathcal{H})$ is a measurable space

Example of configuration space

$\mathbb{U}_a = \{T_a, B_a\}$, $\mathbb{U}_b = \{R_b, L_b\}$, $\Omega = \{\omega^+, \omega^-\}$
$\mathcal{U}_a = 2^{\mathbb{U}_a}$, $\mathcal{U}_b = 2^{\mathbb{U}_b}$, $\mathcal{F} = 2^{\Omega}$



- product configuration space

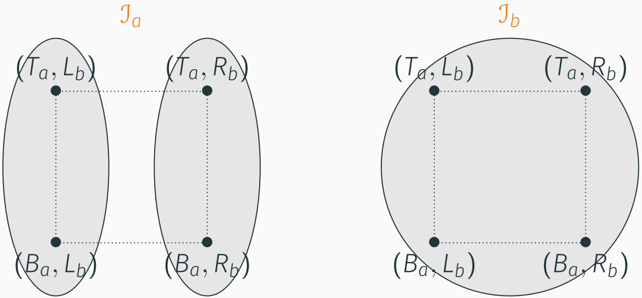
$$\mathbb{H} = \prod_{a \in A} \mathbb{U}_a \times \Omega$$

- product configuration σ-algebra

$$\mathcal{H} = \bigotimes_{a \in A} \mathcal{U}_a \otimes \mathcal{F}$$

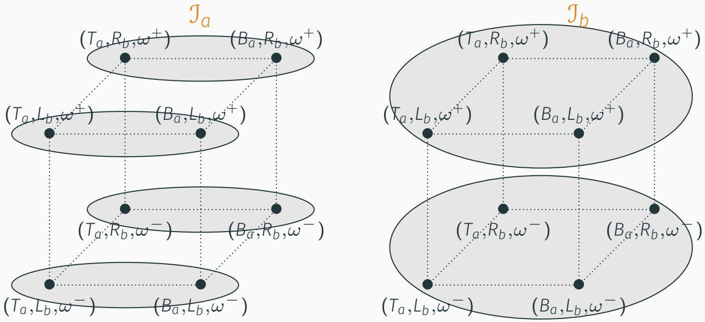
  represented by
  the partition of its atoms

”Alice and Bob” information partitions



- $\mathcal{I}_b = \{\emptyset, \{T_a, B_a\}\} \otimes \{\emptyset, \{R_b, L_b\}\}$
  Bob knows nothing
- $\mathcal{I}_a = \{\emptyset, \{T_a, B_a\}\} \otimes \{\emptyset, \{R_b\}, \{L_b\}, \{R_b, L_b\}\}$
  Alice knows what Bob does
  (as she can distinguish between Bob’s actions $\{R_b\}$ and $\{L_b\}$)

”Alice, Bob and a coin tossing” information partitions



$$\mathcal{I}_b = \overbrace{\{\emptyset, \{T_a, B_a\}\}}^{\text{Bob does not know what Alice does}} \otimes \{\emptyset, \mathbb{U}_b\} \otimes \overbrace{\{\emptyset, \{\omega^+\}, \{\omega^-\}, \{\omega^+, \omega^-\}\}}^{\text{Bob knows Nature's move}}$$

$$\mathcal{I}_a = \{\emptyset, \mathbb{U}_a\} \otimes \underbrace{\{\emptyset, \{R_b\}, \{L_b\}, \{R_b, L_b\}\}}_{\text{Alice knows what Bob does}} \otimes \underbrace{\{\emptyset, \{\omega^+\}, \{\omega^-\}, \{\omega^+, \omega^-\}\}}_{\text{Alice knows Nature's move}}$$

Strategies

- $\mathbb{H} = \prod_{a \in A} \mathbb{U}_a \times \Omega$ is the common product domain
- The agent strategy 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 $a \in \mathbb{A}$

For instance, $\lambda_a^{-1}(\mathcal{U}_a) \subset \underbrace{\{\emptyset, \mathbb{H}\}}_{\text{no information}} \iff \lambda_a$ is constant on $\mathbb{H}$

Relation with SCMs[1]

An SCM formulation takes the form

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

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

[1]Structural Causal Models

Information Dependency Model (IDM)

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

Information Dependency Model (IDM)

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

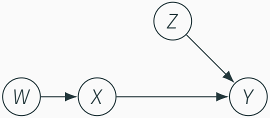
Information Dependency Model (IDM)

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

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

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

How is parentality encoded?



X is a parent of Y

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

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

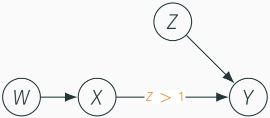
Context-specific independence (CSI)

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

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

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

Conditional Parentality



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

(W, H)-Conditional parentality

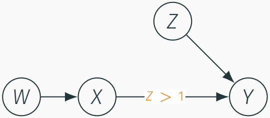
- $W \subset A$: conditioning agents (variables)
- $H \subset \mathbb{H}$: localization

Definition

For a given agent $a \in A$, the conditional parents set $\mathcal{P}^{W,H}a$ is the smallest subset $B \subset A$ such that

$$\mathcal{I}_a \cap H \subset \left(\bigotimes_{b \in B \cup W} \mathcal{U}_b \right) \otimes \mathcal{F}_a \cap H$$

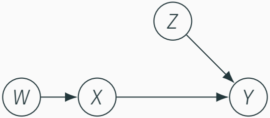
Conditional parentality: conditioning on the context



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

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

Conditional parentality: conditioning on a variable



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

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

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

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

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

- $W \subset A$: conditioning agents (variables)
- $H \subset \mathbb{H}$: localization

Definition

For a given agent $a \in A$, the conditional parents set $\mathcal{P}^{W,H}a$ is the smallest subset $B \subset A$ such that

$$\mathcal{I}_a \cap H \subset \left(\bigotimes_{b \in B \cup W} \mathcal{U}_b \right) \otimes \mathcal{F}_a \cap H$$

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

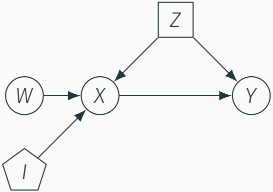
Modeling interventions

Atomic intervention

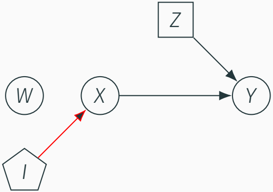
$$\mathcal{I}_X \leftarrow (\mathcal{I}_X \otimes \{I = 0\}) \cup (\{\emptyset, \mathbb{H}\} \otimes \{I = 1\})$$

Atomic intervention

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



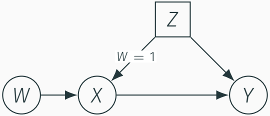
Intervention not activated, $I = 0$



Intervention activated, $I = 1$

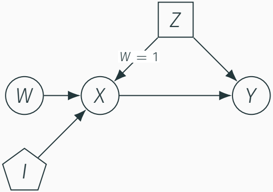
Application[2]

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



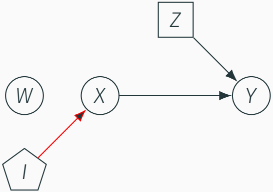
[2]Example taken from Tikka et al. 2019

Application[3] (hand waving style)

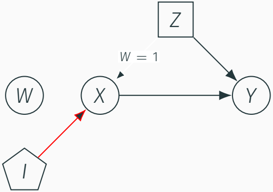


We picture the original graph, with the additional node

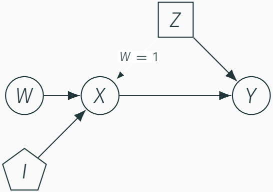
[3]We do it in the graphical world because it is possible to do so. Note however that Information Dependency Models can deal with more complex situations



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 $C \subset A$ are (conditionally) topologically separated (w.r.t. (W, H)), and write

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

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

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

Main result

Theorem

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

Illustration

Example 1

Are Y_1 and Y_2 independent when conditioned on W?

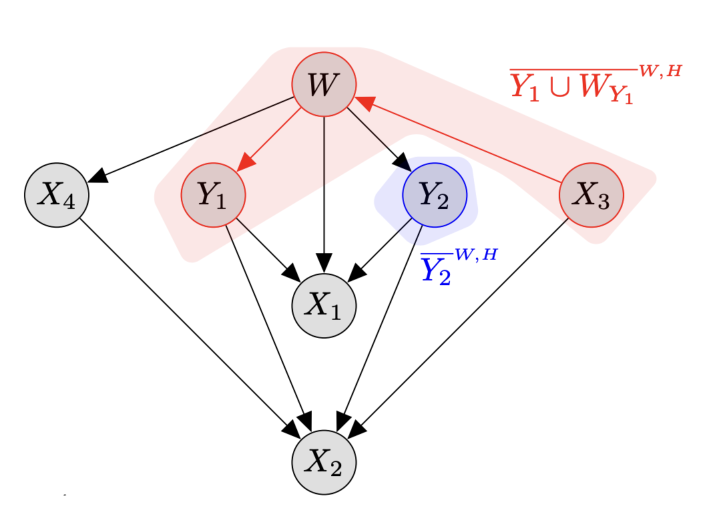


Figure 1: The split of W is a piece of information that can be useful

Example 2

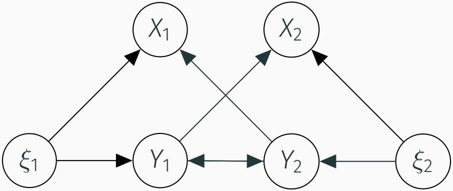


Figure 2: Are the Y's independent when conditioning on the X's ?

Example 2

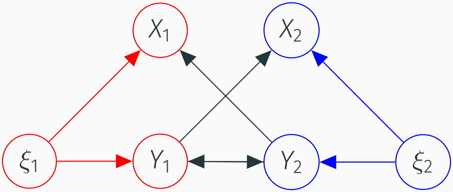


Figure 3: Let $W_{X_i} = Y_i$, for $i = 1, 2$. The closure of $X_1 \cup Y_1$ (resp. $X_2 \cup Y_2$), with the edges followed to build the closure, is in red (resp. blue).

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 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

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

Proofing toolbox: binary relations

Binary relation

We 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. ■

An illustration of equational reasoning

paper_csbr.v - GNU Emacs at wo

Latex File Edit Options Buffers Tools Coq Proof-General YASnippet Holes Outline Hide/Show Help

State Context Goal Retract Undo Next Use Goto Qed Home

```
  move ⇒ x' y'.
  by split ⇒ [[z [H1 /Singl_iff ←]]| H1];[|∃ x';split;[|apply In_singleton]].
Qed.

(* Closure intersect as a relation *)
Definition Closure_intersect :=
  λ (x y:A) ⇒ Clos_(x | E,W) ∩ Clos_(y | E,W) ≠ '∅.

Lemma Clos_Intersect : ∀ (x y:A),
    Clos_(x | E,W) ∩ Clos_(y | E,W) ≠ '∅ ↔
    (let R:= Emw.* * Ew.* in R x y).
Proof.
  move ⇒ w1' w2'; split.
  - rewrite -notempty_exists.
    move ⇒ [_ [z [w1 [H1 /Singl_iff ←]] [w2 [H2 /Singl_iff ←]]]].
    by (∃ z; split;[rewrite Emw_1 |]).
  - rewrite -notempty_exists.
    move ⇒ [z [H1 H2]]; rewrite Emw_1 in H1.
    by ∃ z;split;rewrite !Clos_Ew.
Qed.
```

Conclusion

Summary

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

Making the case for Algebraic Causality

- 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