Questioning the foundations of Causal Discovery

Ferenc Huszár
University of Cambridge

Students:

Siyuan Guo

Viktor Tóth

Szilvia Ujváry

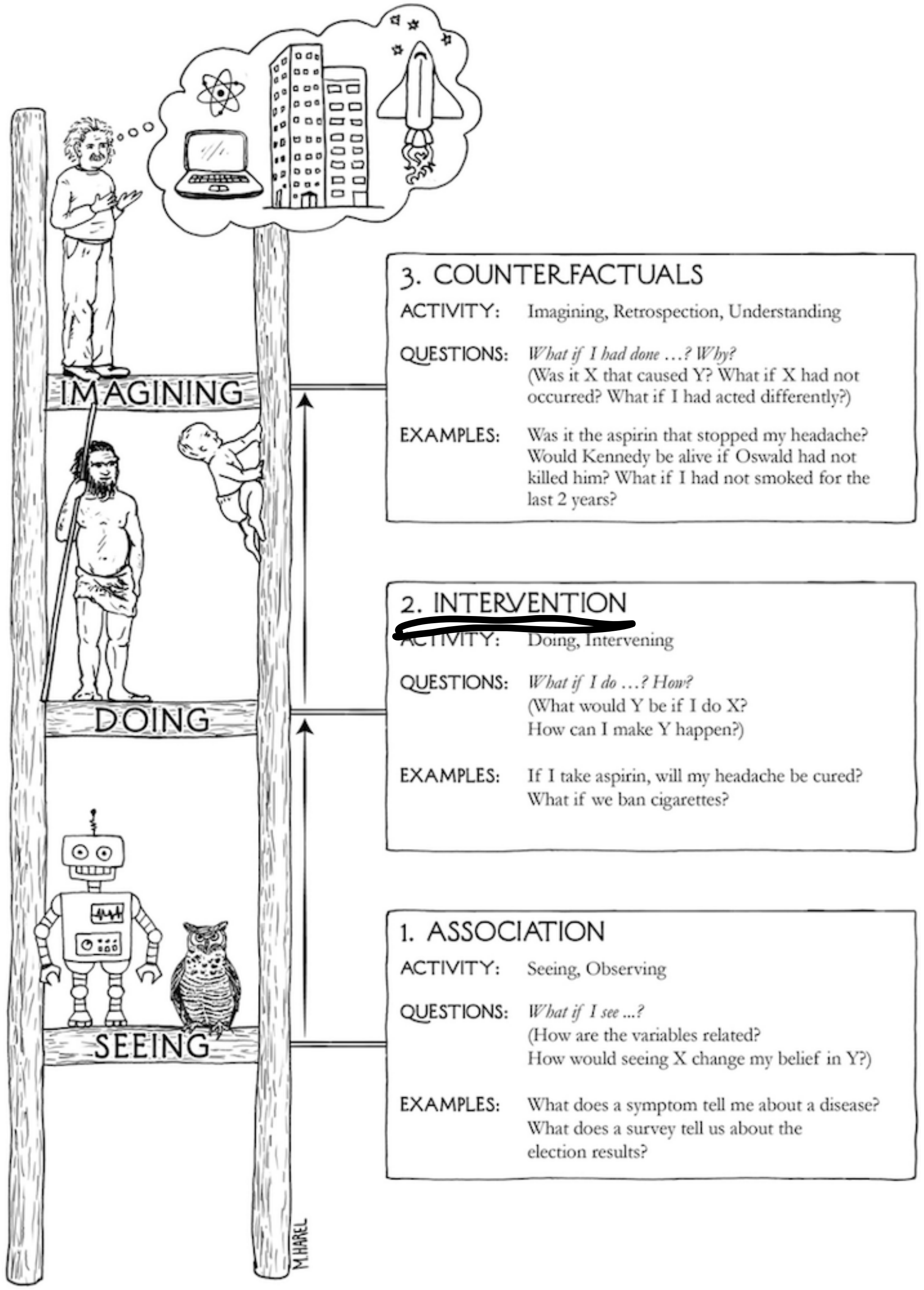
Collaborators:

Bernhard Schölkopf

Karthika Mohan

Gergely Flamich

Context: Causal Hierarchy



Counterfactuals

What would have happened if … ?

Interventions

What happens if … ?

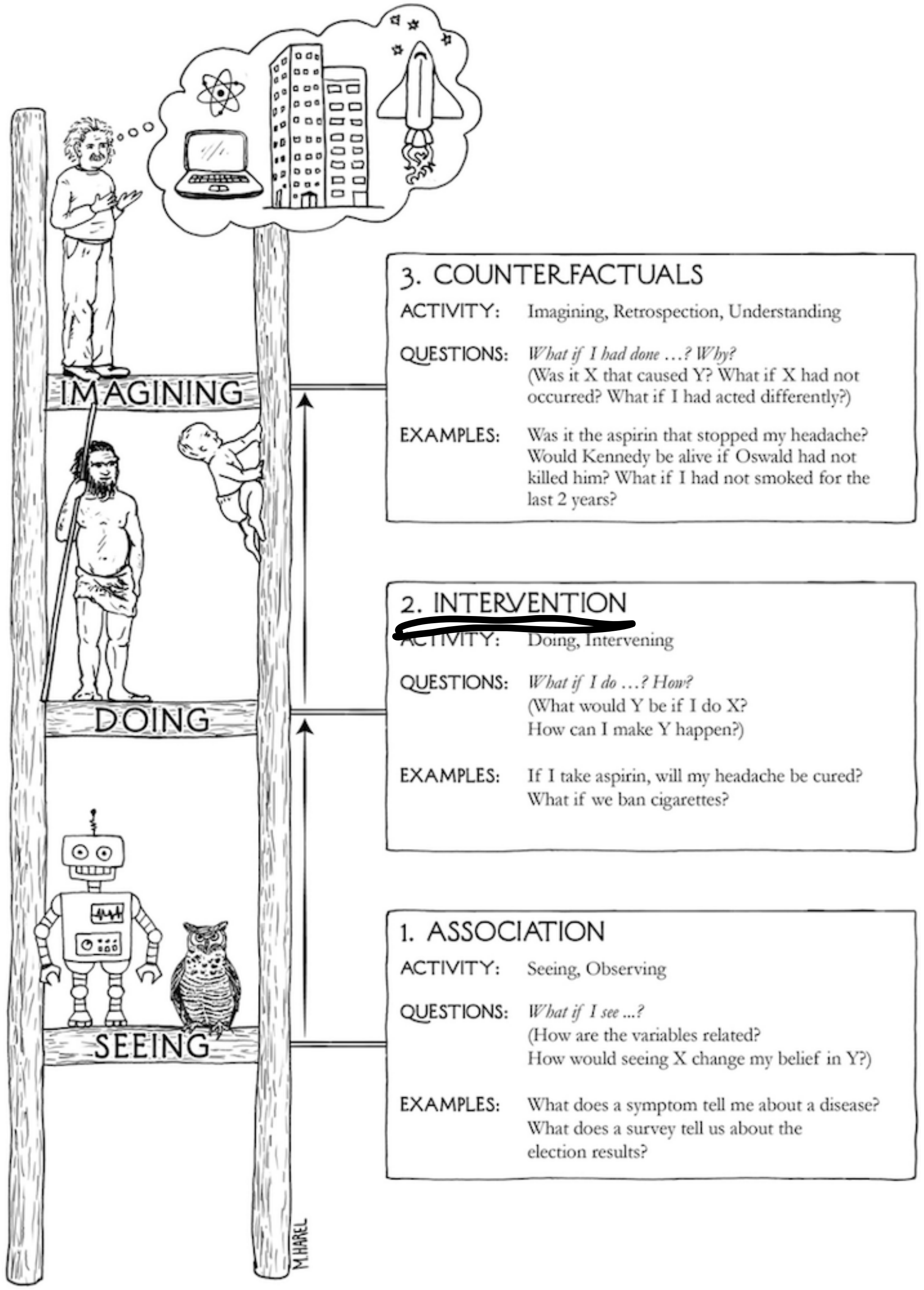
Statistical analysis

conditionals, prediction, compression

more fine grained

empirical verifiability

Context: Causal Hierarchy



Counterfactuals

What would have happened if ... ?

Interventions

What happens if ... ?

Statistical analysis

conditionals, prediction, compression

more fine grained

empirical verifiability

Intervention: What is the distribution of Y conditioned on Z if I force variable X to take value x?

ICM Principle: Independence of Causal Modules

$X \sim P_X$

$Y|X \sim P_{Y|X}$

$Z|Y \sim P_{Z|Y}$

Intervention: What is the distribution of Z conditioned on X if I force variable Y to take value y?

ICM Principle: Independence of Causal Modules

$$X \sim P_X \xrightarrow{\text{keep}} X \sim P_X$$

$$Y|X \sim P_{Y|X} \xrightarrow{\text{replace}} Y \sim \delta_y$$

$$Z|Y \sim P_{Z|Y} \xrightarrow{\text{keep}} Z|Y \sim P_{Z|Y}$$

Intervention: What is the distribution of **Z** conditioned on **X** if I force variable **Y** to take value **y**?

ICM Principle: Independence of Causal Modules

$$X \sim P_X \xrightarrow{\text{keep}} X \sim P_X$$

$$Y|X \sim P_{Y|X} \xrightarrow{\text{replace}} Y \sim \delta_y$$

$$Z|Y \sim P_{Z|Y} \xrightarrow{\text{keep}} Z|Y \sim P_{Z|Y}$$

$$P_{X,Z|do(Y=y)} = \int P_X(x)\, \delta_y(y)\, P_{Z|Y}(z|y)\, dy$$

Intervention: What is the distribution of Z conditioned on X if I force variable Y to take value y?

ICM Principle: Independence of Causal Modules

$$\mathcal{M}_1$$

$$X \sim P_X \xrightarrow{\text{keep}} X \sim P_X$$

$$Y|X \sim P_{Y|X} \xrightarrow{\text{replace}} Y \sim \delta_y$$

$$Z|Y \sim P_{Z|Y} \xrightarrow{\text{keep}} Z|Y \sim P_{Z|Y}$$

$$P^{\mathcal{M}_1}_{X,Z|do(Y=y)} = \int P_X(x)\,\delta_y(y)\,P_{Z|Y}(z|y)\,dy$$

$$\mathcal{M}_2$$

$$Z \sim P_Z \xrightarrow{\text{keep}} Z \sim P_Z$$

$$Y|Z \sim P_{Y|Z} \xrightarrow{\text{replace}} Y \sim \delta_y$$

$$X|Y \sim P_{X|Y} \xrightarrow{\text{keep}} X|Y \sim P_{X|Y}$$

$$P^{\mathcal{M}_2}_{X,Z|do(Y=y)} = \int P_{X|Y}(x|y)\,\delta_y(y)\,P_Z(z)\,dy$$

Intervention: What is the distribution of Z conditioned on X if I force variable Y to take value y?

ICM Principle: Independence of Causal Modules

$X \sim P_X \xrightarrow{\text{keep}} X \sim P_X$ $\quad \mathcal{M}_1 \quad P^{\mathcal{M}_1}_{X,Z \mid do(Y=y)} =$

$Y|X \sim P_Y$ $\quad (z|y)\, dy$

$Z|Y \sim P_Z$

$\mathcal{M}_1$ and $\mathcal{M}_2$ are

OBSERVATIONALLY EQUIVALENT

but support different conclusions

$Z \sim P_Z$ $\quad) =$

$Y|Z \sim P_Y$ $\quad P_Z(z)\, dy$

$X|Y \sim P_{X|Y} \xrightarrow{\text{keep}} X|Y \sim P_{X|Y}$

Intervention: What is the distribution of Z conditioned on X if I force variable Y to take value y?

Identifiability of Correct Causal Model

$$p(x, y, z) = \begin{cases} p(x)\; p(y|x)\; p(z|y,x) \\ p(x)\; p(y|x)\; p(z|y) \\ p(x|y)\; p(y|z)\; p(z) \\ p(x|y,z)\, p(y|z)\, p(z) \\ p(x|y)\; p(z|y)\, p(y) \\ \vdots \end{cases}$$

Identifiability of Correct Causal Model

what the model predicts

$$p(x,y,z) = \left\{ \begin{array}{l} p(x)\,p(y|x)\,p(z|y,x) \xrightarrow{\text{ICM}} \\ p(x)\,p(y|x)\,p(z|y) \longrightarrow \\ p(x|y)\,p(y|z)\,p(z) \longrightarrow \\ p(x|y,z)\,p(y|z)\,p(z) \longrightarrow \\ p(x|y)\,p(z|y)\,p(y) \longrightarrow \\ \vdots \end{array} \right.$$

Identifiability of Correct Causal Model

what the model predicts

ICM

$$p(x,y,z) = \begin{cases} \cancel{p(x)\,p(y|x)\,p(z|y,x)} \\ p(x)\,p(y|x)\,p(z|y) \\ p(x|y)\,p(y|z)\,p(z) \\ \cancel{p(x|y,z)\,p(y|z)\,p(z)} \\ p(x|y)\,p(z|y)\,p(y) \\ \vdots \end{cases}$$

if only there was a way of ruling some of these out!...

Identifiability of Correct Causal Model

what the model predicts

ICM

$$p(x,y,z) = \begin{cases} \cancel{p(x)\,p(y|x)\,p(z|y,x)} \\ p(x)\,p(y|x)\,p(z|y) \\ p(x|y)\,p(y|z)\,p(z) \\ \cancel{p(x|y,z)\,p(y|z)\,p(z)} \\ p(x|y)\,p(z|y)\,p(y) \\ \vdots \end{cases}$$

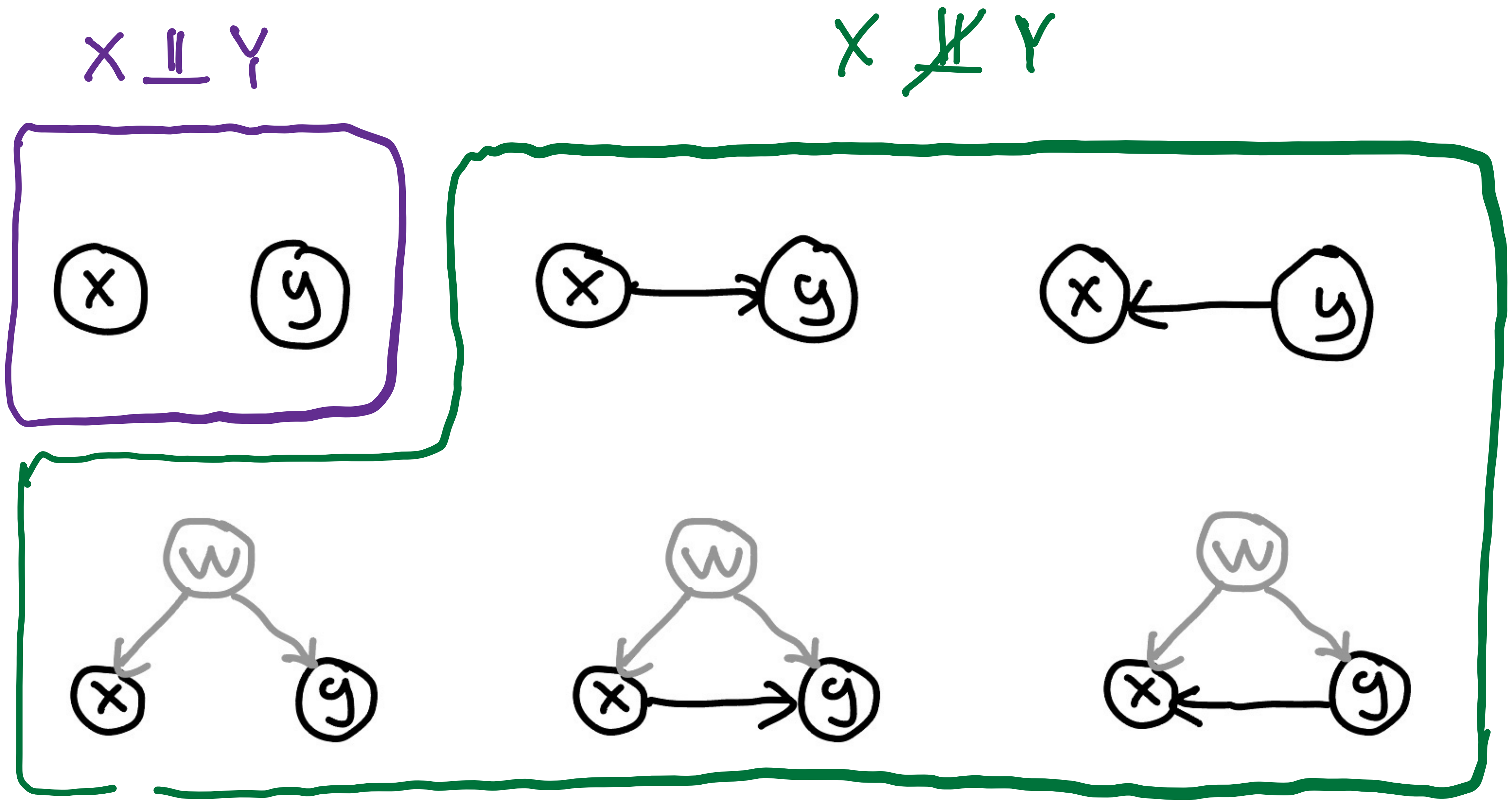
conditional independence structure

$X \perp\!\!\!\perp Z \mid Y$

if only there was a way of ruling some of these out!...

There is!

Example : two-variable causal models



The conditional independence structure is not rich enough!

Causal de Finetti: On the Identification of Invariant Causal Structure in Exchangeable Data

Siyuan Guo[1,2,*] **Viktor Tóth**[1,*] **Bernhard Schölkopf**[2] **Ferenc Huszár**[1]
[1]University of Cambridge [2]Max Planck Institute for Intelligent Systems
{syg26,fh277}@cam.ac.uk toth.viktor7400@gmail.com
bs@tuebingen.mpg.de

Abstract

Constraint-based causal discovery methods leverage conditional independence tests to infer causal relationships in a wide variety of applications. Just as the majority of machine learning methods, existing work focuses on studying *independent and identically distributed* data. However, it is known that even with infinite i.i.d. data, constraint-based methods can only identify causal structures up to broad Markov

Causal inference assumes i.i.d. data

Exchangeable processes are equally justified but richer

Can we exploit this richness?

Yes, we can!

Exchangeability

Exchangeability

$\forall N > 0$, $\forall$ permutation π

$$p(x_1, y_1, x_2, y_2, \dots, x_N, y_N) = p(x_{\pi_1}, y_{\pi_1}, x_{\pi_2}, y_{\pi_2}, \dots, x_{\pi_N}, y_{\pi_N})$$

$$\begin{array}{cc} \boxed{X_1} & \boxed{Y_1} \\ \boxed{X_2} & Y_2 \\ X_3 & Y_3 \\ \vdots & \\ X_n & Y_n \end{array}$$

i.i.d. case

$X_1 \perp\!\!\!\perp Y_1$?

automatically hold

$$\left\{ \begin{array}{l} X_i \perp\!\!\!\perp Y_j \\ X_i \perp\!\!\!\perp X_j \mid Y_i \\ \vdots \end{array} \right.$$

1 bit

exch case

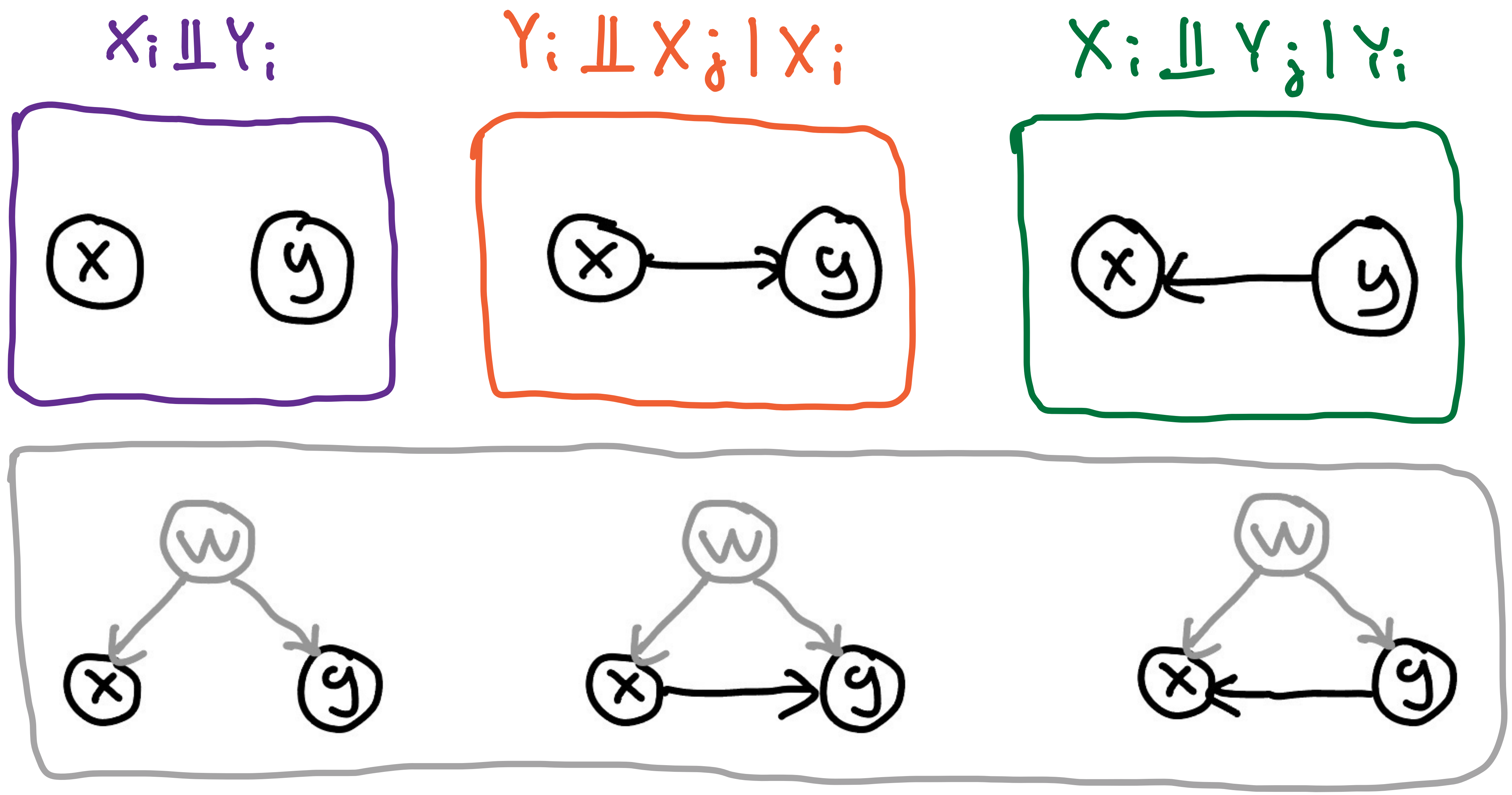
$X_1 \perp\!\!\!\perp Y_1$?

$X_1 \perp\!\!\!\perp X_2$?

$Y_1 \perp\!\!\!\perp X_2 \mid X_1$?

2 bits of info

Causal identification with CdeF



Causal de Finetti

Causal de Finetti theorem (bi-variate case)

Let $(X_i, Y_i)_{i \in \mathbb{N}}$ be an exchangeable pair

If $Y_1 \perp\!\!\!\perp X_2 \mid X_1$ holds, then $\exists\ \Theta, \Psi$ such that

$$P(x_1, y_1, x_2, y_2, \dots, x_N, y_N) = \int\!\!\int \prod_{n=1}^{N} p(x_n \mid \Theta)\, p(y_n \mid x_n, \Psi)\, d\pi(\theta)\, d\pi(\psi)$$

<u>Interpretation</u>: if $X_1 \perp\!\!\!\perp X_2 \mid X_1$ holds, there exist statistically independent mechanisms Θ and Ψ

A statistical ICM principle

Causal de Finetti: On the Identification of Invariant Causal Structure in Exchangeable Data

Siyuan Guo[12*] **Viktor Tóth**[1*] **Bernhard Schölkopf**[2] **Ferenc Huszár**[1]
[1]University of Cambridge [2]Max Planck Institute for Intelligent Systems
{syg26,fh277}@cam.ac.uk toth.viktor7400@gmail.com
bs@tuebingen.mpg.de

Abstract

Constraint-based causal discovery methods leverage conditional independence tests to infer causal relationships in a wide variety of applications. Just as the majority of machine learning methods, existing work focuses on studying *independent and identically distributed* data. However, it is known that even with infinite i.i.d. data, constraint-based methods can only identify causal structures up to broad Markov

Causal inference assumes i.i.d. data

Exchangeable processes are equally justified but richer

Can we exploit this richness?

Yes, we can!

Causal de Finetti: On the Identification of Invariant Causal Structure in Exchangeable Data

Siyuan Guo[1,2*] **Viktor Tóth**[1*] **Bernhard Schölkopf**[2] **Ferenc Huszár**[1]
[1]University of Cambridge [2]Max Planck Institute for Intelligent Systems
{syg26,fh277}@cam.ac.uk toth.viktor7400@gmail.com
bs@tuebingen.mpg.de

Abstract

Constraint-based causal discovery methods leverage conditional independence tests to infer causal relationships in a wide variety of applications. Just as the majority of machine learning methods, existing work focuses on studying *independent and identically distributed* data. However, it is known that even with infinite i.i.d. data, constraint-based methods can only identify causal structures up to broad Markov

Causal inference assumes i.i.d. data

Exchangeable processes are equally justified but richer

Can we exploit this richness?

Yes, we can!

...but also: No, we cannot.

New frontier: algorithmic info theory

Causal de Finetti	Algorithmic C de F
exchangeable data generating process	? ? ?
conditional independences $Y_i \perp\!\!\!\perp X_j \mid X_i$	? ? ?
statistical independence of causal modules	? ? ?

New frontier: algorithmic info theory

Causal de Finetti	Algorithmic C de F
exchangeable data generating process	? ? ?
conditional independences $Y_i \perp\!\!\!\perp X_j \mid X_i$	? ? ?
statistical independence of causal modules	algorithmic independence of causal modules Janzing & Schölkopf

New frontier: algorithmic info theory

Causal de Finetti	Algorithmic C de F
exchangeable data generating process	algorithmic exchangeability?
conditional independences $Y_i \perp\!\!\!\perp X_j \mid X_i$	? ? ?
statistical independence of causal modules	algorithmic independence of causal modules Janzing & Schölkopf

Algorithmic exchangeability

Algorithmic exchangeability

Let $x_1, x_2, \dots$ be an infinite sequence of objects, and $\langle\cdot\rangle$ a uniquely decodable source code.

The sequence is algorithmically exchangeable, iff

$\exists\, C>0: \forall$ computable permutation π, $\forall\, N>0$

$$\left| K[\langle x_1\rangle\langle x_2\rangle \cdots \langle x_N\rangle] - K[\langle x_{\pi_1}\rangle\langle x_{\pi_2}\rangle \cdots \langle x_{\pi_N}\rangle] \right| < C$$

Note: important that only computable permutations are allowed. Otherwise:

$$1\,3\,2\,2\,3\,1\,1\,3\,2\,1\,3 \xrightarrow{\pi} 1\,1\,1\,1\;\;2\,2\,2\;\;3\,3\,3\,3$$

Algorithmic de Finetti conjecture

If $x_1, \dots, x_N, \dots$ infinite sequence of objects is algo exchangeable with uniquely decodable finite code $\langle \cdot \rangle$

$\exists$ probability distribution P and constant $C > 0$ s.t.

$$\forall N \quad K[\langle x_1 \rangle \langle x_2 \rangle \cdots \langle x_N \rangle] > -\sum_{n=1}^{N} \log P(x_i) + C$$

Interpretation: algorithmic exchangeability justifies statistical modeling.

THANKS!

To you and also...

Students:

Siyuan Guo

Viktor Tóth

Szilvia Ujváry

Collaborators:

Bernhard Schölkopf

Karthika Mohan

Gergely Flamich