Joint causal inference on observational and experimental datasets

Sara Magliacane, Tom Claassen, Joris M. Mooij

s.magliacane@uva.nl

What If?, 10th December, 2016

Part I

Introduction

Causal inference: learning causal relations from data

Causal inference: learning causal relations from data

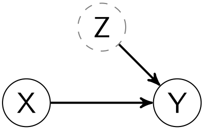
Definition

X causes Y ($X \dashrightarrow Y$) = *intervening upon* (changing) X changes Y

Definition

X causes Y ($X \dashrightarrow Y$) = *intervening upon* (changing) X changes Y

- We can represent causal relations with a causal DAG (hidden vars):



E.g. X = Smoking, Y = Cancer, Z = Genetics

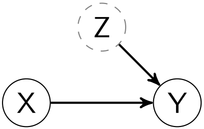
- Causal inference = structure learning of the causal DAG

Causal inference: learning causal relations from data

Definition

X causes Y ($X \dashrightarrow Y$) = *intervening upon* (changing) X changes Y

- We can represent causal relations with a causal DAG (hidden vars):



E.g. X = Smoking, Y = Cancer, Z = Genetics

- Causal inference = structure learning of the causal DAG
- Traditionally, causal relations are inferred from interventions.
- Sometimes, interventions are unethical, unfeasible or too expensive

Causal inference from observational and experimental data

Holy Grail of Causal Inference

Learn as much causal structure as possible from observations, integrating background knowledge and experimental data.

Causal inference from observational and experimental data

Holy Grail of Causal Inference

Learn as much causal structure as possible from observations, integrating background knowledge and experimental data.

Current causal inference methods:

- **Score-based**: evaluate models using a penalized likelihood score
- **Constraint-based**: use statistical independences to express constraints over possible causal models

Causal inference from observational and experimental data

Holy Grail of Causal Inference

Learn as much causal structure as possible from observations, integrating background knowledge and experimental data.

Current causal inference methods:

- **Score-based**: evaluate models using a penalized likelihood score
- **Constraint-based**: use statistical independences to express constraints over possible causal models

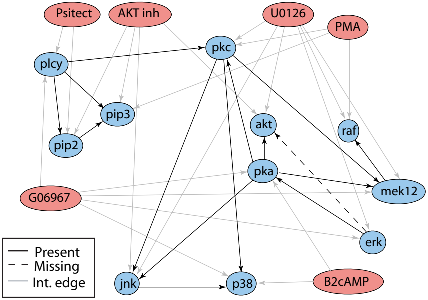
Advantage of constraint-based methods:

- can handle latent confounders naturally

Joint inference on observational and experimental data

Advantage of **score-based** methods:

- can formulate joint inference on observational and experimental data and learn the targets of interventions, e.g. [Eaton and Murphy, 2007].

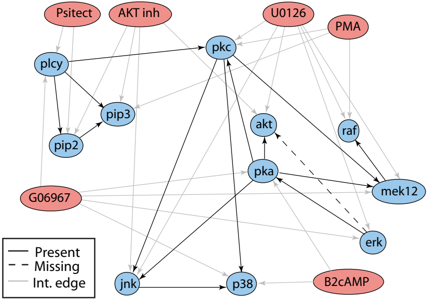


Example from [Eaton and Murphy, 2007]

Joint inference on observational and experimental data

Advantage of **score-based** methods:

- can formulate joint inference on observational and experimental data and learn the targets of interventions, e.g. [Eaton and Murphy, 2007].



Example from [Eaton and Murphy, 2007]

Goal: Can we perform joint inference using constraint-based methods?

Part II

Joint Causal Inference

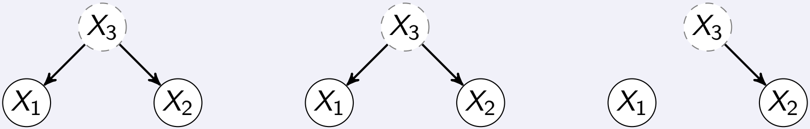
Joint Causal Inference: Assumptions

Idea: Model jointly several observational or experimental datasets $\{D_r\}_{r \in \{1 \dots n\}}$ with zero or more possibly unknown intervention targets.

Idea: Model jointly several observational or experimental datasets $\{D_r\}_{r \in \{1 \ldots n\}}$ with zero or more possibly unknown intervention targets.

We assume a **unique underlying causal DAG** across datasets defined over system variables $\{X_j\}_{j \in \mathcal{X}}$ (some of which possibly hidden).

Example



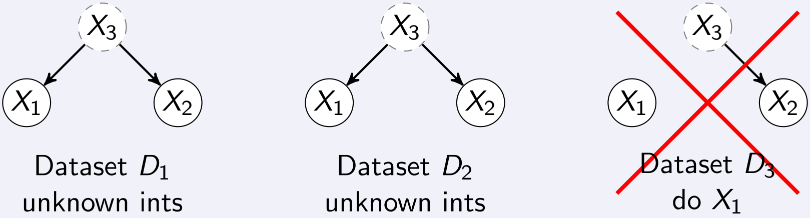
Dataset D_1 unknown ints | Dataset D_2 unknown ints | Dataset D_3 do X_1

Joint Causal Inference: Assumptions

Idea: Model jointly several observational or experimental datasets $\{D_r\}_{r\in\{1\dots n\}}$ with zero or more possibly unknown intervention targets.

We assume a **unique underlying causal DAG** across datasets defined over system variables $\{X_j\}_{j\in\mathcal{X}}$ (some of which possibly hidden).

Example



Note: cannot handle certain intervention types, e.g. perfect interventions.

Joint Causal Inference: SCM

We introduce two types of dummy variables in the data:

- a **regime variable** R, indicating which dataset D_r a data point is from
- **intervention variables** $\{I_i\}_{i \in \mathcal{I}}$, functions of R

Joint Causal Inference: SCM

We introduce two types of dummy variables in the data:

- a **regime variable** R, indicating which dataset D_r a data point is from
- **intervention variables** $\{I_i\}_{i\in\mathcal{I}}$, functions of R

We assume that we can represent the whole system as an acyclic SCM:

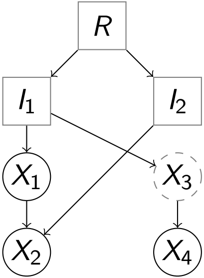
$$\begin{cases} R &= E_R, \\ I_i &= g_i(R), \quad i \in \mathcal{I}, \\ X_j &= f_j(X_{\mathrm{pa}(X_j)\cap\mathcal{X}}, I_{\mathrm{pa}(X_j)\cap\mathcal{I}}, E_j), \quad j \in \mathcal{X}, \end{cases}$$

$$P\big((E_k)_{k\in\mathcal{X}\cup\{R\}}\big) = \prod_{k\in\mathcal{X}\cup\{R\}} P(E_k).$$

We represent the SCM with a causal DAG $\mathcal{C}$ representing all datasets jointly:

R	I_1	I_2	X_1	X_2	X_4
1	20	0	0.1	0.2	0.5
1	20	0	0.13	0.21	0.49
1	20	0	...	...	...
2	20	1	...	...	...
3	30	0	...	...	...
4	30	1	...	...	...

4 datasets with 2 interventions



Causal DAG $\mathcal{C}$

We assume Causal Markov and Minimality hold in $\mathcal{C}$.

Joint Causal Inference: Faithfulness violations

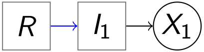
Causal Faithfulness assumption:

$$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \implies \boldsymbol{X} \perp_d \boldsymbol{Y} \mid \boldsymbol{W} \; [\mathcal{C}]$$

Joint Causal Inference: Faithfulness violations

Causal Faithfulness assumption:

$$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \implies \boldsymbol{X} \perp_d \boldsymbol{Y} \mid \boldsymbol{W} \; [\mathcal{C}]$$

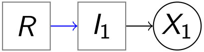


$I_1 = g_1(R) \implies X_1 \perp\!\!\!\perp I_1 \mid R$ ... but $X_1 \not\perp_d I_1 \mid R$

Joint Causal Inference: Faithfulness violations

Causal Faithfulness assumption:

$$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \implies \boldsymbol{X} \perp_d \boldsymbol{Y} \mid \boldsymbol{W} \; [\mathcal{C}]$$



$I_1 = g_1(R) \implies X_1 \perp\!\!\!\perp I_1 \mid R$ ... but $X_1 \not\perp_d I_1 \mid R$

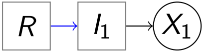
Solution: D-separation [Spirtes et al., 2000]: $X_1 \perp_D I_1 \mid R$

Problem: $\perp_D$ provably complete (for now) for *functionally determined* relations

Joint Causal Inference: Faithfulness violations

Causal Faithfulness assumption:

$$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \,|\, \boldsymbol{W} \implies \boldsymbol{X} \perp_d \boldsymbol{Y} \,|\, \boldsymbol{W} \; [\mathcal{C}]$$



$I_1 = g_1(R) \implies X_1 \perp\!\!\!\perp I_1 \,|\, R$ ... but $X_1 \not\perp_d I_1 \,|\, R$

Solution: D-separation [Spirtes et al., 2000]: $X_1 \perp_D I_1 \,|\, R$

Problem: $\perp_D$ provably complete (for now) for *functionally determined* relations

Solution: restrict deterministic relations to only $\forall i \in \mathcal{I}: \; I_i = g_i(R)$

D-Faithfulness: $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \,|\, \boldsymbol{W} \implies \boldsymbol{X} \perp_D \boldsymbol{Y} \,|\, \boldsymbol{W} \; [\mathcal{C}]$.

Joint Causal Inference

Joint Causal Inference (JCI) = Given all the assumptions, reconstruct the causal DAG $\mathcal{C}$ from independence test results.

Joint Causal Inference (JCI) = Given all the assumptions, reconstruct the causal DAG $\mathcal{C}$ from independence test results.

$$X_1 \not\perp\!\!\!\perp I_1$$

$$X_1 \perp\!\!\!\perp I_1 \mid R \qquad \Longrightarrow$$

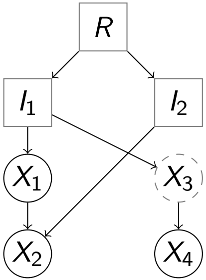
$$X_4 \perp\!\!\!\perp X_2 \mid I_1$$

$$\ldots$$

Joint Causal Inference (JCI) = Given all the assumptions, reconstruct the causal DAG $\mathcal{C}$ from independence test results.

$$X_1 \not\perp\!\!\!\perp I_1$$
$$X_1 \perp\!\!\!\perp I_1 \mid R$$
$$X_4 \perp\!\!\!\perp X_2 \mid I_1$$
$$\ldots$$

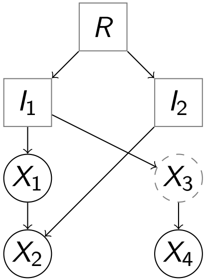
$$\Longrightarrow$$



Joint Causal Inference (JCI) = Given all the assumptions, reconstruct the causal DAG $\mathcal{C}$ from independence test results.

$$X_1 \not\perp\!\!\!\perp I_1$$
$$X_1 \perp\!\!\!\perp I_1 \mid R$$
$$X_4 \perp\!\!\!\perp X_2 \mid I_1$$
$$\dots$$

$$\Longrightarrow$$



Problem: Current constraint-based methods cannot work with JCI, because of faithfulness violations.

Part III

Extending constraint-based methods for JCI

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations $\rightarrow$ Partial inputs

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations $\rightarrow$ Partial inputs

A simple strategy for dealing with functionally determined relations:

1. Rephrase constraint-based method in terms of d-separations

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations $\rightarrow$ Partial inputs

A simple strategy for dealing with functionally determined relations:

1. Rephrase constraint-based method in terms of d-separations
2. For each independence test result derive sound d-separations:

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations → Partial inputs

A simple strategy for dealing with functionally determined relations:

1. Rephrase constraint-based method in terms of d-separations
2. For each independence test result derive sound d-separations:
   - $X \not\perp\!\!\!\perp Y | \boldsymbol{W} \implies X \not\perp_d Y | \boldsymbol{W}$

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations → Partial inputs

A simple strategy for dealing with functionally determined relations:

1. Rephrase constraint-based method in terms of d-separations
2. For each independence test result derive sound d-separations:
   - $X \not\perp\!\!\!\perp Y | \boldsymbol{W} \implies X \not\perp_d Y | \boldsymbol{W}$
   - $X \notin \textsc{Det}(\boldsymbol{W})$ and $Y \notin \textsc{Det}(\boldsymbol{W})$ and $X \perp\!\!\!\perp Y \,|\, \boldsymbol{W} \implies X \perp_d Y \,|\, \textsc{Det}(\boldsymbol{W})$

where $\textsc{Det}(\boldsymbol{W})$ = variables determined by (a subset of) $\boldsymbol{W}$.

A simple strategy for dealing with faithfulness violations

Idea: Faithfulness violations $\rightarrow$ Partial inputs

A simple strategy for dealing with functionally determined relations:

1. Rephrase constraint-based method in terms of d-separations
2. For each independence test result derive sound d-separations:
   - $X \not\perp\!\!\!\perp Y | \boldsymbol{W} \implies X \not\perp_d Y | \boldsymbol{W}$
   - $X \notin \mathrm{Det}(\boldsymbol{W})$ and $Y \notin \mathrm{Det}(\boldsymbol{W})$ and $X \perp\!\!\!\perp Y \mid \boldsymbol{W} \implies$
     $X \perp_d Y \mid \mathrm{Det}(\boldsymbol{W})$

where $\mathrm{Det}(\boldsymbol{W})$ = variables determined by (a subset of) $\boldsymbol{W}$.

Conjecture: sound also for a larger class of deterministic relations.

Idea: Faithfulness violations → Partial inputs

Ancestral Causal Inference with Determinism (ACID)

Idea: Faithfulness violations → Partial inputs

Traditional constraint-based methods (e.g., PC, FCI, ...):

- cannot handle partial inputs
- cannot exploit the rich background knowledge in JCI

Ancestral Causal Inference with Determinism (ACID)

Idea: Faithfulness violations → Partial inputs

Traditional constraint-based methods (e.g., PC, FCI, ...):

- cannot handle partial inputs
- cannot exploit the rich background knowledge in JCI

Solution: Logic-based methods, e.g., ACI [Magliacane et al., 2016]

Ancestral Causal Inference with Determinism (ACID)

Idea: Faithfulness violations → Partial inputs

Traditional constraint-based methods (e.g., PC, FCI, ...):

- cannot handle partial inputs
- cannot exploit the rich background knowledge in JCI

Solution: Logic-based methods, e.g., ACI [Magliacane et al., 2016]

⟹ we implement the strategy in an extension of ACI called:

Rephrasing ACI rules in terms of d-separations

ACI rules:

Example

For X, Y, $\boldsymbol{W}$ disjoint (sets of) variables:

$$(X \perp\!\!\!\perp Y \mid \boldsymbol{W}) \wedge (X \not\dashrightarrow \boldsymbol{W}) \implies X \not\dashrightarrow Y$$

Rephrasing ACI rules in terms of d-separations

ACI rules:

Example

For X, Y, $\boldsymbol{W}$ disjoint (sets of) variables:

$$(X \perp\!\!\!\perp Y \mid \boldsymbol{W}) \land (X \not\dashrightarrow \boldsymbol{W}) \implies X \not\dashrightarrow Y$$

ACID rules:

Example

For X, Y, $\boldsymbol{W}$ disjoint (sets of) variables:

$$(X \perp_d Y \mid \boldsymbol{W}) \land (X \not\dashrightarrow \boldsymbol{W}) \implies X \not\dashrightarrow Y$$

ACID-JCI = ACID rules + sound d-separations from strategy

+ JCI background knowledge

ACID-JCI = ACID rules + sound d-separations from strategy

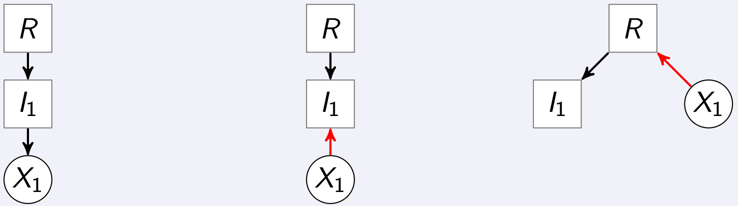
+ JCI background knowledge

For example:

$$\forall i \in \mathcal{I}, \forall j \in \mathcal{X} : (X_j \not\dashrightarrow R) \wedge (X_j \not\dashrightarrow I_i)$$

"Standard variables cannot cause dummy variables"

Example



ACID-JCI

ACID-JCI = ACID rules + sound d-separations from strategy

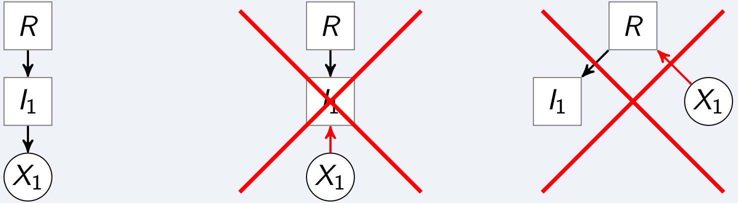
+ JCI background knowledge

For example:

$$\forall i \in \mathcal{I}, \forall j \in \mathcal{X} : (X_j \not\dashrightarrow R) \wedge (X_j \not\dashrightarrow I_i)$$

"Standard variables cannot cause dummy variables"

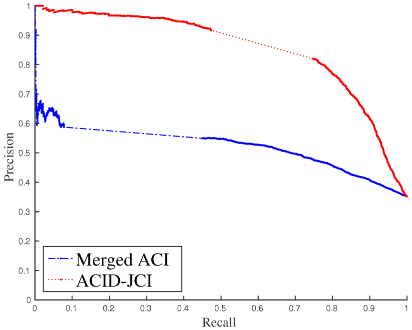
Example



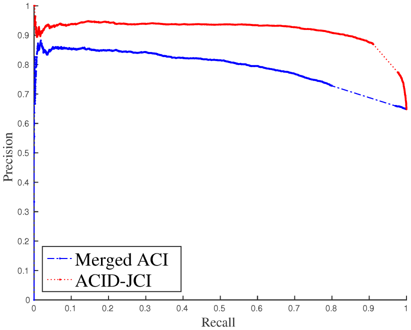
Part IV

Evaluation

Simulated data accuracy: example Precision Recall curve



Ancestral ("causes") relations



Non-ancestral ("not causes")

- Precision Recall curves of 2000 randomly generated causal graphs for 4 system variables and 3 interventions
- ACID-JCI substantially improves accuracy w.r.t. merging learnt structures (merged ACI)

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods
- A simple strategy for dealing with faithfulness violations due to functionally determined relations

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods
- A simple strategy for dealing with faithfulness violations due to functionally determined relations
- An implementation, ACID-JCI, that substantially improves the accuracy w.r.t. state-of-the-art

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods
- A simple strategy for dealing with faithfulness violations due to functionally determined relations
- An implementation, ACID-JCI, that substantially improves the accuracy w.r.t. state-of-the-art

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods
- A simple strategy for dealing with faithfulness violations due to functionally determined relations
- An implementation, ACID-JCI, that substantially improves the accuracy w.r.t. state-of-the-art

Future work:

- Improve scalability (now max 7 variables in $\mathcal{C}$).

Working paper: `https://arxiv.org/abs/1611.10351`

Conclusion

- Joint Causal Inference, a powerful formulation of causal discovery over multiple datasets for constraint-based methods
- A simple strategy for dealing with faithfulness violations due to functionally determined relations
- An implementation, ACID-JCI, that substantially improves the accuracy w.r.t. state-of-the-art

Future work:

- Improve scalability (now max 7 variables in $\mathcal{C}$).

Working paper: https://arxiv.org/abs/1611.10351

Collaborators:

Tom Claassen 

 Joris M. Mooij

References I

Claassen, T. and Heskes, T. (2011).
A logical characterization of constraint-based causal discovery.
In *UAI 2011*, pages 135–144.

Eaton, D. and Murphy, K. (2007).
Exact Bayesian structure learning from uncertain interventions.
In *Proceedings of the 10th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 107–114.

Magliacane, S., Claassen, T., and Mooij, J. M. (2016).
Ancestral causal inference.
In *NIPS*.

Mooij, J. M. and Heskes, T. (2013).
Cyclic causal discovery from continuous equilibrium data.
In Nicholson, A. and Smyth, P., editors, *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pages 431–439. AUAI Press.

Spirtes, P., Glymour, C., and Scheines, R. (2000).
Causation, Prediction, and Search.
MIT press, 2nd edition.

Ancestral Causal Inference (ACI)

- Weighted list of inputs: $I = \{(i_j, w_j)\}$:
  - E.g. $I = \{\ (Y \perp\!\!\!\perp Z \,|\, X, 0.2), \ (Y \not\perp\!\!\!\perp X, 0.1)\}\ \}$
  - Any consistent weighting scheme, e.g. frequentist, Bayesian
- For any possible ancestral structure C, we define the loss function:

$$\mathcal{L}oss(C, I) := \sum_{(i_j, w_j) \in I:\ i_j \text{ is not satisfied in } C} w_j$$

- Here: "i_j is not satisfied in C" = defined by ancestral reasoning rules
- For each possible causal relation $X \dashrightarrow Y$ provide score:

$$Conf(X \dashrightarrow Y) = \min_{C \in \mathcal{C}} \mathcal{L}oss(C,\ I + (X \not\dashrightarrow Y, \infty))$$

$$- \min_{C \in \mathcal{C}} \mathcal{L}oss(C,\ I + (X \dashrightarrow Y,\ \infty))$$