UNIVERSITÄT ZU LÜBECK

Algorithmics of Causal Inference

Maciej Liśkiewicz

Institut für Theoretische Informatik, Universität zu Lübeck

KI-Kolloquium der Universität zu Lübeck, November 2019

IM FOCUS DAS LEBEN

Talk's Overview

1. Introduction: discovering of causal relationships
2. Causal inference: causal relationships from *observed data*
3. Algorithmic estimations of causal effects
4. Conclusions

Discovering of causal relationships

What causal relationships can be established? In ...

... economy?

[wikipedia]

... health?

[wikipedia]

... environment and nature?

Discovering of causal relationships

What causal relationships can be established? In ...

... economy?

[wikipedia]

... health?

[wikipedia]

... environment and nature?

What effect does have ...

... the tax increase on economic growth?

... physical activity on health?

... recycling on the environment.?

Discovering of causal relationships

What causal relationships can be established? In ...

... economy?

[wikipedia]

... health?

[wikipedia]

... environment and nature?

What are the major causes of ...

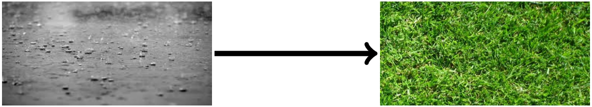
... economic crises?

... diseases?

... pollinator decline (Insektensterben)?

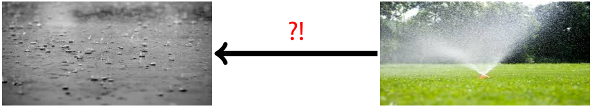
Discovering of causal relationships

What does it mean cause/effect (causal relation)?



Discovering of causal relationships

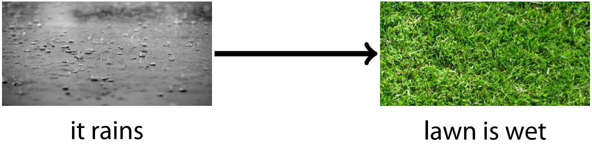
What does it mean cause/effect (causal relation)?



it rains

lawn is wet

Discovering of causal relationships
What does it mean cause/effect (causal relation)?



causality ≠ correlation

Discovering of causal relationships

Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth

Discovering of causal relationships
Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth
- *Direct experimentation*
  - ethically problematic
  - expensive
  - impossible

Discovering of causal relationships
Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth
- *Direct experimentation*
  - ethically problematic
  - expensive
  - impossible

[wikimedia.org]

Discovering of causal relationships
Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth
- *Direct experimentation*
  - ethically problematic
  - expensive
  - impossible

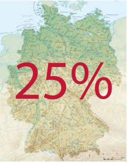

[wikimedia.org]


[wikimedia.org]

Discovering of causal relationships
Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth
- *Direct experimentation*
  - ethically problematic
  - expensive
  - impossible

[wikimedia.org]



[wikimedia.org]

Discovering of causal relationships

Complex systems

- Causal relationships in *complex systems*:
  - Does smoking cause lung cancer?
  - Predict effect of the tax increase on economic growth
- *Direct experimentation*
  - ethically problematic
  - expensive
  - impossible
- On the other hand, there are often available large amounts of *observed data* that can provide relevant information about these issues

	16	23	26	30	33	41	45	53	55	58	66	75	77	78	81	86	87	89	90	...
L77V	6,1	4	7,5	5,4	3	5,5	8,7	1,7	1,4	8,7	14	8,3	141,4	100,9	22,3	14,8	25,3	42,4	6,5	...
M81L	3,3	2,6	10,8	8,2	7,5	7,4	2,6	7,8	15,4	12,8	19,5	4,3	144,5	54,8	383,9	13,8	33,9	17,6	17,9	...
M109L	7,1	6,7	11,9	10,6	5,1	9,1	5,2	8,7	17	8,7	7,4	10,1	7,9	25	8,4	0	31,5	117	0	...
M112L	12,5	5,2	5,3	4,1	14,2	18,7	17,9	31,1	4,8	3	2,4	4	18,8	13,9	17,5	12,6	39	16,1	1,1	...
L116V	8,8	5,1	3,4	7	4,8	8,6	6,9	10,4	3,4	2,5	5,6	2,8	11,7	13,8	15,6	3,8	4,8	15,7	1,9	...
M120L	7,2	3,5	5,3	0,6	4,3	0,4	1,1	8,5	0	0,9	5,8	1,3	10,7	8,1	10,9	4,9	1,5	5,1	2,7	...
M137L	3,5	2,2	3,2	3,4	4,2	9,7	6,9	9,6	1,5	3,3	3,9	4,2	42,2	13,8	36,4	2,4	3	11,8	8,4	...
I150L	6	4,1	2,8	7,2	5,7	7,3	5,1	5,7	4,8	5,1	19,8	5,2	48,1	11,6	53,3	3,1	8,4	10,7	5	...
L159V	4,1	4,9	1,9	4,7	0,8	8,4	10,3	7,6	3,8	3,6	7,7	2,5	19,7	12,8	24,7	1,9	13,7	12,1	5,3	...
L167V	4,9	4,2	5,1	5,3	7,2	4,2	11,2	8,9	4,7	2,1	11,7	0,6	4,5	4,8	4	3,3	1,4	9	8,9	...
L174V	7,8	3,3	6,1	6,6	4,7	9,2	10,3	3,3	4,8	14,4	12,5	1,1	93	39,4	134,8	26,7	20,4	15,8	18,5	...
V186A	2,1	3,8	1,3	1,5	0,9	4,1	5,7	1,2	3,3	1,1	10,1	2	15	7,6	13,9	3,6	3,8	6	5,3	...
V187A	7,6	3,7	3,2	6,9	3	7,4	11,4	4,6	5,3	5,9	7,5	4,3	15,8	9,8	5,5	11,9	5,3	6	6,6	...
L198V	5	4,2	1,8	10,7	6,2	5,2	6,8	5	1,2	2,6	3,5	4,2	9,9	5,1	6,6	1,8	7,7	5,6	9,4	...
M216L	1,9	1,4	1,6	3,5	5,8	10	6,2	7,6	5,3	3,5	4,4	1,4	7,7	4,7	5,3	0,8	4,6	5,9	4,4	...
L225V	4	8,1	1,8	3,4	3,6	4,2	7	2,8	3,5	1,6	8,9	1,5	5,5	4,2	4,1	4,5	6,8	4,1	4	...
...																				

[Aoto et al., Scientific reports 6 (2016)]

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

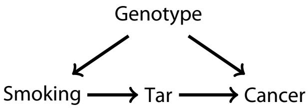
	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85

1 *[Pearl, Causality, 2009]*

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85

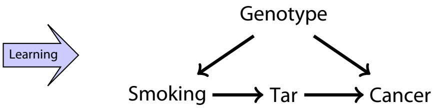


1 *[Pearl, Causality, 2009]*

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85

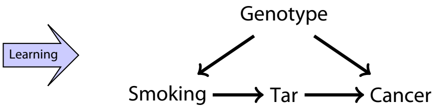


1 *[Pearl, Causality, 2009]*

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85



- *Causal Inference* (sub-field of AI)

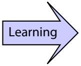
  - Mathematical modeling of direct experimentation
  - Estimation of causal effects

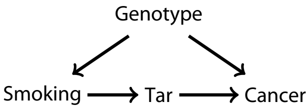
1 *[Pearl, Causality, 2009]*

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85





- *Causal Inference* (sub-field of AI)
  - Mathematical modeling of direct experimentation
  - Estimation of causal effects

© 2011 ACM

Judea Pearl © 2011 ACM

In 2011 Judea Pearl won the ACM A.M. Turing Award, recognized as the Nobel Prize of Computing, for his groundbreaking work in the field of Bayesian networks, which greatly advanced both Artificial Intelligence and Causality.

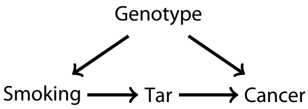
[1] *[Pearl, Causality, 2009]*

Discovering of causal relationships from *observed data*

- Does smoking cause lung cancer?
- Relevant factors: Genotype, Tar in the lungs [1]

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85

Learning



- *Causal Inference* (sub-field of AI)
  - Mathematical modeling of direct experimentation
  - Estimation of causal effects
- *Focus of our research*
  - Algorithmics of causal inference
  - Effective methods for identification and estimation of causal effects

© 2011 ACM

Judea Pearl © 2011 ACM

In 2011 Judea Pearl won the ACM A.M. Turing Award, recognized as the Nobel Prize of Computing, for his groundbreaking work in the field of Bayesian networks, which greatly advanced both Artificial Intelligence and Causality.

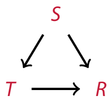
[1] [Pearl, Causality, 2009]

Structural causal models and the do-calculus

Central components: *causal structure (diagram)* and the *do-operator* introduced by Pearl

- A *causal diagram* used as a model of causal relationships is represented (typically) by a directed acyclic graph (DAG) $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ whose:
  - vertices $\mathbf{V} = \{X_1, \ldots, X_n\}$ represent random variables of interest and
  - edges $X_i \to X_j$ express direct causal effects of one variable on another.

  Example: Kidney Stones



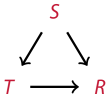
S – size of stone (small, large), T – treatment (A or B), R – recovery (0, 1).

Structural causal models and the do-calculus

Central components: *causal structure (diagram)* and the *do-operator* introduced by Pearl

- A *causal diagram* used as a model of causal relationships is represented (typically) by a directed acyclic graph (DAG) $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ whose:
  - vertices $\mathbf{V} = \{X_1, \ldots, X_n\}$ represent random variables of interest and
  - edges $X_i \to X_j$ express direct causal effects of one variable on another.

  Example: Kidney Stones



S – size of stone (small, large), T – treatment (A or B), R – recovery (0, 1).

- The *do-operator* allows mathematical modeling of interventions and to predict causal effects from observational data.

  Example: What is the expected recovery if all get treatment A?

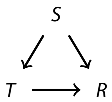
Structural causal models and the do-calculus
Graphical concepts

Let P be a j.p.d. of $\mathbf{V} = \{X_1, \ldots, X_n\}$ and $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG.

- P and $\mathcal{G}$ are *consistent* (or $\mathcal{G}$ represents P) if P admits the factorization:

$$P(\mathbf{v}) = \prod_{j=1}^{n} P(x_j | pa_j),$$

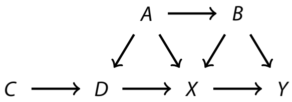
  where pa_j denotes a particular realization of the parent variables of X_j in $\mathcal{G}$.
- Notation: $P(y)$ means $P(Y = y)$.
- For example, the graph



  induces the factorization: $P(s, t, p) = P(t|s)P(r|s, t)P(s)$.

Structural causal models and the do-calculus
Graphical concepts

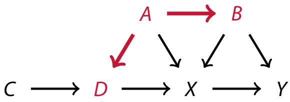
- Causal Model: DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$



Structural causal models and the do-calculus

Graphical concepts

- Causal Model: DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$

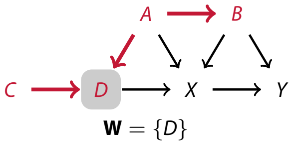


- *D* and *B* are *d-connected* if there is a path π betwen *D* and *B* which does not contain a collider $\rightarrow X \leftarrow$.

Structural causal models and the do-calculus

Graphical concepts

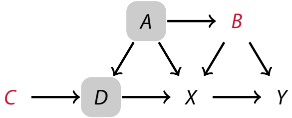
- Causal Model: DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$



- D and B are *d-connected* if there is a path π betwen D and B which does not contain a collider $\to X \leftarrow$.
- C and B are *d-connected by a set* $\mathbf{W}$ if there is π between them on which
  - every non-collider is not in $\mathbf{W}$ and
  - every collider is an ancestor of $\mathbf{W}$.

Structural causal models and the do-calculus
Graphical concepts

- Causal Model: DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$



$\mathbf{W} = \{A, D\}$

- *D* and *B* are *d-connected* if there is a path π betwen *D* and *B* which does not contain a collider $\rightarrow X \leftarrow$.
- *C* and *B* are *d-connected by a set* $\mathbf{W}$ if there is π between them on which
  - every non-collider is not in $\mathbf{W}$ and
  - every collider is an ancestor of $\mathbf{W}$.
- $\mathbf{W}$ *d*-separates *C* and *B* if they are not *d*-connected by $\mathbf{W}$.

Structural causal models and the do-calculus

Graphical concepts

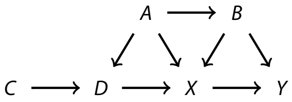
Theorem (*d*-separation vs. conditional independence (Verma, Pearl))

For any three disjoint subsets of nodes $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *in a DAG* $\mathcal{G}$ *and for all probability functions P, we have:*

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{\mathcal{G}} \quad \Longrightarrow \quad (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_P$$

whenever $\mathcal{G}$ *and P are consistent.*

For example, for any P over $\mathbf{V} = \{A, B, C, D, X, Y\}$ consistent with:



we have, e.g.: $(C \perp\!\!\!\perp B)_P$ and $(C \perp\!\!\!\perp X \mid \{D, A\})_P$

Structural causal models and the do-calculus

The do-calculus

Let P be a probability distribution and $\mathcal{G}$ a P-consistent DAG.

- The operator

$$do(\mathbf{X} = \mathbf{x})$$

($do(\mathbf{x})$, for short) models an *external intervention* that fixes the cause variables $\mathbf{X}$ to the values $\mathbf{x}$. This corresponds to removing all edges entering $\mathbf{X}$ and fixing $\mathbf{X}$ to $\mathbf{x}$.

- The *causal effect* of $\mathbf{X}$ on outcome variables $\mathbf{Y}$, denoted as

$$P(\mathbf{y}|do(\mathbf{x}))$$

is defined as the probability distribution of variables $\mathbf{Y}$ after the intervention.

Structural causal models and the do-calculus

The do-calculus

Let P be a probability distribution and $\mathcal{G}$ a P-consistent DAG.

- The operator
$$do(\mathbf{X} = \mathbf{x})$$
($do(\mathbf{x})$, for short) models an *external intervention* that fixes the cause variables $\mathbf{X}$ to the values $\mathbf{x}$. This corresponds to removing all edges entering $\mathbf{X}$ and fixing $\mathbf{X}$ to $\mathbf{x}$.

- The *causal effect* of $\mathbf{X}$ on outcome variables $\mathbf{Y}$, denoted as
$$P(\mathbf{y}|do(\mathbf{x}))$$
is defined as the probability distribution of variables $\mathbf{Y}$ after the intervention.

- *Observation vs. Intervention* R: it is raining ($R = 1$) or not ($R = 0$);
W: the lawn is wet ($W = 1$) or not ($W = 0$)

$$R \longrightarrow W$$

Structural causal models and the do-calculus

The do-calculus

Let P be a probability distribution and $\mathcal{G}$ a P-consistent DAG.

- The operator

$$do(\mathbf{X} = \mathbf{x})$$

  ($do(\mathbf{x})$, for short) models an *external intervention* that fixes the cause variables $\mathbf{X}$ to the values $\mathbf{x}$. This corresponds to removing all edges entering $\mathbf{X}$ and fixing $\mathbf{X}$ to $\mathbf{x}$.

- The *causal effect* of $\mathbf{X}$ on outcome variables $\mathbf{Y}$, denoted as

$$P(\mathbf{y}|do(\mathbf{x}))$$

  is defined as the probability distribution of variables $\mathbf{Y}$ after the intervention.

- *Observation vs. Intervention* R: it is raining ($R = 1$) or not ($R = 0$);
  W: the lawn is wet ($W = 1$) or not ($W = 0$)

$$R \longrightarrow W$$

  - Suppose: $P(R = 1) = 0.01$, and $P(W = 1 \mid R = 1) = 1$, $P(W = 1 \mid R = 0) = 0.001$.
  - Let's suppose we *observe* that the lawn is wet. Using Bayes Theorem, we get:

$$P(R = 1 \mid W = 1) = 0.91$$

Structural causal models and the do-calculus

The do-calculus

Let P be a probability distribution and $\mathcal{G}$ a P-consistent DAG.

- The operator

$$do(\mathbf{X} = \mathbf{x})$$

  ($do(\mathbf{x})$, for short) models an *external intervention* that fixes the cause variables $\mathbf{X}$ to the values $\mathbf{x}$. This corresponds to removing all edges entering $\mathbf{X}$ and fixing $\mathbf{X}$ to $\mathbf{x}$.

- The *causal effect* of $\mathbf{X}$ on outcome variables $\mathbf{Y}$, denoted as

$$P(\mathbf{y}|do(\mathbf{x}))$$

  is defined as the probability distribution of variables $\mathbf{Y}$ after the intervention.

- *Observation vs. Intervention* R: it is raining ($R = 1$) or not ($R = 0$);
  W: the lawn is wet ($W = 1$) or not ($W = 0$)

$$R \qquad 1 = W$$

  - Suppose: $P(R = 1) = 0.01$, and $P(W = 1 \mid R = 1) = 1$, $P(W = 1 \mid R = 0) = 0.001$.
  - Let's suppose we *observe* that the lawn is wet. Using Bayes Theorem, we get:

$$P(R = 1 \mid W = 1) = 0.91$$

  - If we *intervene* to make the lawn wet, we get:

$$P(R = 1 \mid do(W = 1)) = 0.01$$

Structural causal models and the do-calculus

The do-calculus

Let P be a probability distribution and $\mathcal{G}$ a P-consistent DAG.

- The operator

$$do(\mathbf{X} = \mathbf{x})$$

  ($do(\mathbf{x})$, for short) models an *external intervention* that fixes the cause variables **X** to the values **x**. This corresponds to removing all edges entering **X** and fixing **X** to **x**.

- The *causal effect* of **X** on outcome variables **Y**, denoted as

$$P(\mathbf{y}|do(\mathbf{x}))$$

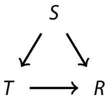
  is defined as the probability distribution of variables **Y** after the intervention.

- The causal effect of **X** on **Y** is *identifiable* if $P(\mathbf{y}|do(\mathbf{x}))$ can be expressed using only standard *pre-intervention* probabilities which involve *observed* variables.

Causal Inference

Intervention Example: Kidney Stones

S – size of stone (small, large), T – treatment (A or B), R – recovery (0,1)



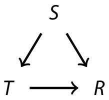
over data: $P(S = s), P(T = t|S = s), P(R = r|T = t, S = s)$.

	Treatment A	Treatment B
Small Stones $\left(\frac{357}{700} = 0.51\right)$	$\frac{81}{87} = 0.93$	$\frac{234}{270} = 0.87$
Large Stones $\left(\frac{343}{700} = 0.49\right)$	$\frac{192}{263} = 0.73$	$\frac{55}{80} = 0.69$
	$\frac{273}{350} = 0.78$	$\frac{289}{350} = 0.83$
	$\frac{562}{700} = 0.80$	

Charig et al.: *Comparison of treatment of renal calculi by open surgery (...),* British Medical Journal, 1986

Causal Inference
Intervention Example: Kidney Stones

S – size of stone (small, large), T – treatment (A or B), R – recovery (0,1)



over data: $P(S = s), P(T = t|S = s), P(R = r|T = t, S = s)$.

	Treatment A	Treatment B
Small Stones $\left(\frac{357}{700} = 0.51\right)$	$\frac{81}{87} = 0.93$	$\frac{234}{270} = 0.87$
Large Stones $\left(\frac{343}{700} = 0.49\right)$	$\frac{192}{263} = 0.73$	$\frac{55}{80} = 0.69$
	$\frac{273}{350} = 0.78$	$\frac{289}{350} = 0.83$
	$\frac{562}{700} = 0.80$	

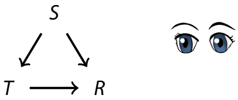
Charig et al.: *Comparison of treatment of renal calculi by open surgery (...)*, British Medical Journal, 1986

- What is the probability of recovery if all get treatment A (resp. B)?
  Solution: Make treatment independent of size.

Causal Inference

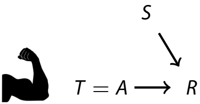
Intervention Example: Kidney Stones

S – size of stone (small, large), T – treatment (A or B), R – recovery (0,1)



over data: $P(S = s), P(T = t|S = s), P(R = r|T = t, S = s)$.

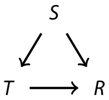
Make the intervention: *“do $T = A$”*, meaning *“use treatment A”*; We delete the edge $S \to T$ and fix T to the value A:



Causal Inference

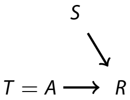
Intervention Example: Kidney Stones

S – size of stone (small, large), T – treatment (A or B), R – recovery (0,1)



over data: $P(S = s), P(T = t|S = s), P(R = r|T = t, S = s)$.

Make the intervention: *"do $T = A$"*, meaning *"use treatment A"*; We delete the edge $S \rightarrow T$ and fix T to the value A:



The causal effect $P(R = 1 \mid do(T = A))$ of the treatment A on recovery is identifiable and can be expressed by a formula:

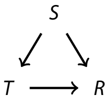
$$P(R = 1 \mid do(T = A)) \;=\; \sum_s P(R = 1|s, T = A)P(s) \;=\; 0.832$$

using only pre-intervention probabilities.

Causal Inference

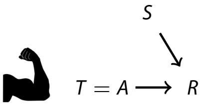
Intervention Example: Kidney Stones

S – size of stone (small, large), T – treatment (A or B), R – recovery (0,1)



over data: $P(S = s)$, $P(T = t|S = s)$, $P(R = r|T = t, S = s)$.

Make the intervention: *"do $T = A$"*, meaning *"use treatment A"*; We delete the edge $S \rightarrow T$ and fix T to the value A:



The causal effect $P(R = 1 \mid do(T = A))$ of the treatment A on recovery is identifiable and can be expressed by a formula:

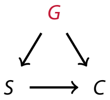
$$P(R = 1 \mid do(T = A)) \;=\; \sum_s P(R = 1|s, T = A)P(s) \;=\; 0.832$$

using only pre-intervention probabilities. Analogously we can compute

$$P(R = 1 \mid do(T = B)) \;=\; \sum_s P(R = 1|s, T = B)P(s) \;=\; 0.782 \; < 0.832\,.$$

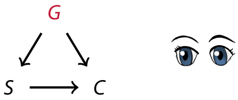
Causal Inference
Intervention Example: Smoking

S – smoking, *C* – lung cancer, *G* – carcinogenic genotype (unobserved)

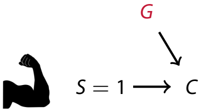


Intervention Example: Smoking

S – smoking, C – lung cancer, G – carcinogenic genotype (unobserved)

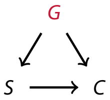


Make the intervention: *"do $S = 1$"*, meaning *"smoker"*; We delete $G \to S$ and fix S to 1:

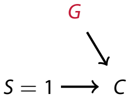


Causal Inference
Intervention Example: Smoking

S – smoking, C – lung cancer, G – carcinogenic genotype (unobserved)



Make the intervention: *"do* $S = 1$*"*, meaning *"smoker"*; We delete $G \to S$ and fix S to 1:

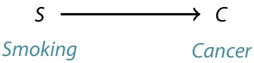


In this case the causal effect $P(C = 1 \mid do(S = 1))$ of the smoking on the lung cancer is non-identifiable!

Structural causal models and the do-calculus

The do-calculus

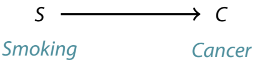
Example: Smoking



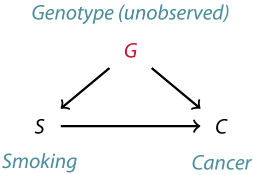
$$P(c \mid do(s)) = P(c \mid s)$$

Structural causal models and the do-calculus
The do-calculus

Example: Smoking



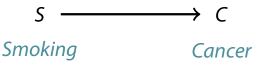
$$P(c \mid do(s)) = P(c \mid s)$$



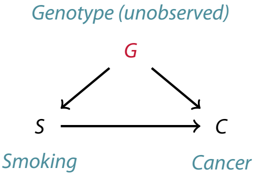
$$P(c \mid do(s)) = \text{noncomputable}$$

Structural causal models and the do-calculus
The do-calculus

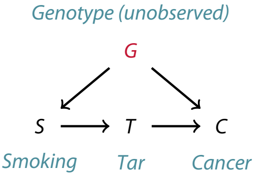
Example: Smoking



$P(c \mid do(s)) = P(c \mid s)$



$P(c \mid do(s)) =$ noncomputable

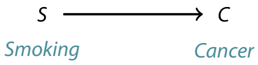


$P(c \mid do(s)) =$ computable

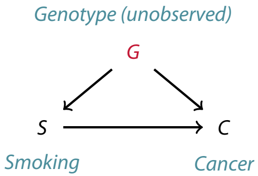
Structural causal models and the do-calculus

The do-calculus

Example: Smoking

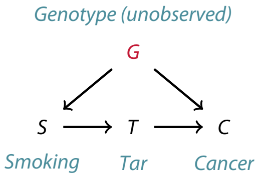


$$P(c \mid do(s)) = P(c \mid s)$$



$$P(c \mid do(s)) = \text{noncomputable}$$

	% of Population	% of Cancer cases
Nonsmokers, No tar	47.5	10
Smokers, No tar	2.5	90
Nonsmokers, Tar	2.5	5
Smokers, Tar	47.5	85



$$P(c \mid do(s)) = \text{computable}$$

$$P(c \mid do(s)) = \sum_{s'} \sum_t P(c|t, s')P(s')P(t|s)$$

Graphical models and algorithmic approaches

Problem 1. Learning Causal Structure:

$$\text{observational data} \quad \mapsto \quad \text{causal structure } \mathcal{G}$$

- A difficulty: the data-generating probability distribution(s) might be represented by different structures.
- A challenge: high dimensionality of data.

Problem 2. Inferring Interventional Distribution:

$$\mathcal{G} \text{ and data} \quad \mapsto \quad \text{predictions on the effect of interventions}$$

- Given a $\mathcal{G}$ *decide* in which situations causal effects can be identified and
- if this is possible: haw can one estimate the strength of the causal effect.

Graphical models and algorithmic approaches

Problem 1. Learning Causal Structure:

$$\text{observational data} \quad \mapsto \quad \text{causal structure } \mathcal{G}$$

- A difficulty: the data-generating probability distribution(s) might be represented by different structures.
- A challenge: high dimensionality of data.

Problem 2. Inferring Interventional Distribution:

$$\mathcal{G} \text{ and data} \quad \mapsto \quad \text{predictions on the effect of interventions}$$

- Given a $\mathcal{G}$ *decide* in which situations causal effects can be identified and
- if this is possible: haw can one estimate the strength of the causal effect.

- A crucial benefit of the graphical language is that it allows an *algorithmic* approaches.
- However, it remains still a big challenge to bridge the gap between graphical modeling and algorithmic effectiveness.

Graphical models and algorithmic approaches

Problem 1. Learning Causal Structure:

$$\text{observational data} \quad \mapsto \quad \text{causal structure } \mathcal{G}$$

- A difficulty: the data-generating probability distribution(s) might be represented by different structures.
- A challenge: high dimensionality of data.

Problem 2. Inferring Interventional Distribution:

$$\mathcal{G} \text{ and data} \quad \mapsto \quad \text{predictions on the effect of interventions}$$

- Given a $\mathcal{G}$ *decide* in which situations causal effects can be identified and
- if this is possible: haw can one estimate the strength of the causal effect.

- A crucial benefit of the graphical language is that it allows an *algorithmic* approaches.
- However, it remains still a big challenge to bridge the gap between graphical modeling and algorithmic effectiveness.
- Our research:
  - algorithmics of causal inference,
  - to provide effective and practically implementable methods,
  - data is high-dimensional.

Algorithmic estimations of causal effects: our research

- *do-calculus* by Pearl: sound and complete calculus for identification.
- *IDC* algorithm (Shpitser and Pearl): based on *do-calculus* solves the identification problem in polynomial time.
- Two drawbacks:
  - polynomial time of high degree
  - computes complex expressions, even if simple ones exists.

Algorithmic estimations of causal effects: our research

- *do-calculus* by Pearl: sound and complete calculus for identification.
- *IDC* algorithm (Shpitser and Pearl): based on *do-calculus* solves the identification problem in polynomial time.
- Two drawbacks:
  - polynomial time of high degree
  - computes complex expressions, even if simple ones exists.
- In practice the vast majority of causal effect estimations are computed either by
  - *covariate adjustment*
  - *instrumental variable*

  method.

Algorithmic estimations of causal effects: our research

- *do-calculus* by Pearl: sound and complete calculus for identification.
- *IDC* algorithm (Shpitser and Pearl): based on *do-calculus* solves the identification problem in polynomial time.
- Two drawbacks:
  - polynomial time of high degree
  - computes complex expressions, even if simple ones exists.
- In practice the vast majority of causal effect estimations are computed either by
  - *covariate adjustment* $\Leftarrow$
  - *instrumental variable*

  method.

Identification of causal effects: adjustment in DAGs

- Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and observed variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ is called *adjustment set* for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ if for every P consistent with $\mathcal{G}$ we have

$$P(\mathbf{y} \mid do(\mathbf{x})) = \begin{cases} P(\mathbf{y} \mid \mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z}) & \text{otherwise.} \end{cases}$$

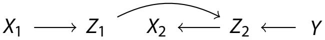
Identification of causal effects: adjustment in DAGs

- Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and observed variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ is called *adjustment set* for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ if for every P consistent with $\mathcal{G}$ we have

$$P(\mathbf{y} \mid do(\mathbf{x})) = \begin{cases} P(\mathbf{y} \mid \mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z}) & \text{otherwise.} \end{cases}$$

- The famous *back-door* criterion by Pearl: $\mathbf{Z}$ satisfies the back-door criterion if
  - no element in $\mathbf{Z}$ is a descendant of $\mathbf{X}$ and
  - $\mathbf{Z}$ *d*-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}_{\underline{\mathbf{X}}}$.

  is a simple, easily implementable rule. But it is not complete:



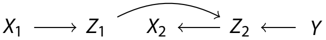
Identification of causal effects: adjustment in DAGs

- Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and observed variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ is called *adjustment set* for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ if for every P consistent with $\mathcal{G}$ we have

$$P(\mathbf{y} \mid do(\mathbf{x})) = \begin{cases} P(\mathbf{y} \mid \mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z}) & \text{otherwise.} \end{cases}$$

- The famous *back-door* criterion by Pearl: $\mathbf{Z}$ satisfies the back-door criterion if
  - no element in $\mathbf{Z}$ is a descendant of $\mathbf{X}$ and
  - $\mathbf{Z}$ *d*-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}_{\underline{\mathbf{X}}}$.

  is a simple, easily implementable rule. But it is not complete:



- Complete adjustment criterion (AC) (Shpitser, VanderWeele, Robins): $\mathbf{Z}$ satisfies AC if
  (*a*) no element in $\mathbf{Z}$ is a descendant in $\mathcal{G}_{\overline{\mathbf{X}}}$ of any $W \in \mathbf{V} \setminus \mathbf{X}$ which lies on a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ and
  (*b*) all proper non-causal paths in $\mathcal{G}$ from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by $\mathbf{Z}$.

Identification of causal effects: adjustment in DAGs

- Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and observed variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ is called *adjustment set* for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ if for every P consistent with $\mathcal{G}$ we have

$$P(\mathbf{y} \mid do(\mathbf{x})) = \begin{cases} P(\mathbf{y} \mid \mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z}) & \text{otherwise.} \end{cases}$$

- The famous *back-door* criterion by Pearl: $\mathbf{Z}$ satisfies the back-door criterion if
  - no element in $\mathbf{Z}$ is a descendant of $\mathbf{X}$ and
  - $\mathbf{Z}$ *d*-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}_{\underline{\mathbf{X}}}$.

  is a simple, easily implementable rule. But it is not complete:

$$X_1 \longrightarrow Z_1 \quad X_2 \longleftarrow Z_2 \longleftarrow Y \qquad (Z_1 \to Z_2)$$

- Complete adjustment criterion (AC) (Shpitser, VanderWeele, Robins): $\mathbf{Z}$ satisfies AC if
  (a) no element in $\mathbf{Z}$ is a descendant in $\mathcal{G}_{\overline{\mathbf{X}}}$ of any $W \in \mathbf{V} \setminus \mathbf{X}$ which lies on a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ and
  (b) all proper non-causal paths in $\mathcal{G}$ from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by $\mathbf{Z}$.

  The drawback: it does not yield a practical algorithm for adjustment set construction.

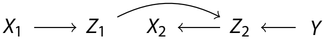
Identification of causal effects: adjustment in DAGs

- Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and observed variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z}$ is called *adjustment set* for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ if for every P consistent with $\mathcal{G}$ we have

$$P(\mathbf{y} \mid do(\mathbf{x})) = \begin{cases} P(\mathbf{y} \mid \mathbf{x}) & \text{if } \mathbf{Z} = \emptyset, \\ \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z}) & \text{otherwise.} \end{cases}$$

- The famous *back-door* criterion by Pearl: $\mathbf{Z}$ satisfies the back-door criterion if
  - no element in $\mathbf{Z}$ is a descendant of $\mathbf{X}$ and
  - $\mathbf{Z}$ *d*-separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}_{\underline{\mathbf{X}}}$.

  is a simple, easily implementable rule. But it is not complete:

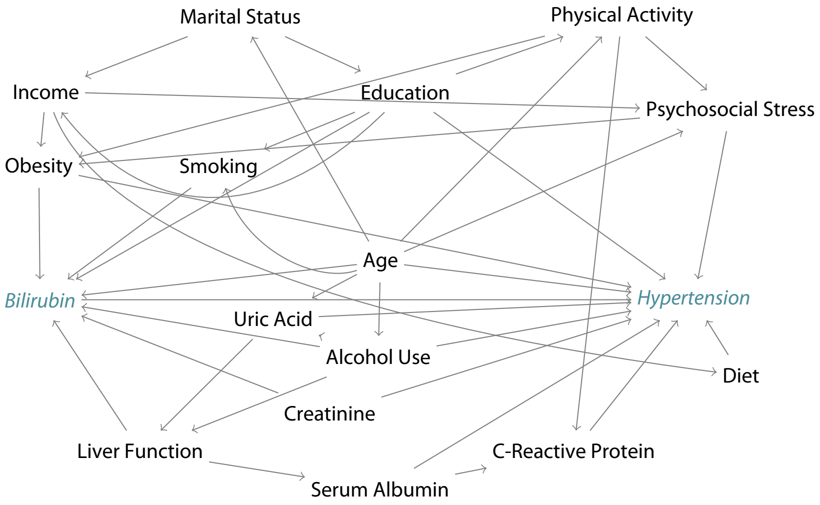


- Complete adjustment criterion (AC) (Shpitser, VanderWeele, Robins): $\mathbf{Z}$ satisfies AC if
  (*a*) no element in $\mathbf{Z}$ is a descendant in $\mathcal{G}_{\overline{\mathbf{X}}}$ of any $W \in \mathbf{V} \setminus \mathbf{X}$ which lies on a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ and
  (*b*) all proper non-causal paths in $\mathcal{G}$ from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by $\mathbf{Z}$.

  The drawback: it does not yield a practical algorithm for adjustment set construction.
- We need effective algorithms!

Identification of causal effects: adjustment in DAGs

We need effective algorithms



Identification of causal effects: adjustment in DAGs

Our results

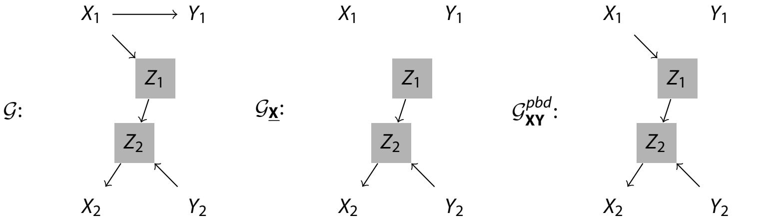
1. New constructive, sound and complete criterion for adjustment in DAGs.
2. The criterion reduces non-causal paths to ordinary *d*-separation.
3. Use algorithms for *d*-separation (with constraints).

Identification of causal effects: adjustment in DAGs

Our results

1. New constructive, sound and complete criterion for adjustment in DAGs.
2. The criterion reduces non-causal paths to ordinary *d*-separation.
3. Use algorithms for *d*-separation (with constraints).

- Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$.
- The *proper back-door graph* is obtained from $\mathcal{G}$ by removing the first edge of every proper causal path from $\mathbf{X}$ to $\mathbf{Y}$. Let $PCP(\mathbf{X}, \mathbf{Y}) = (De_{\overline{\mathbf{X}}}(\mathbf{X}) \setminus \mathbf{X}) \cap An_{\underline{\mathbf{X}}}(\mathbf{Y})$.
- CBC: $\mathbf{Z} \subseteq \mathbf{V} \setminus Dpcp(\mathbf{X}, \mathbf{Y})$ and $\mathbf{Z}$ *d*-separates $\mathbf{X}$ and $\mathbf{Y}$ in the proper back-door graph $\mathcal{G}^{pbd}_{\mathbf{XY}}$.



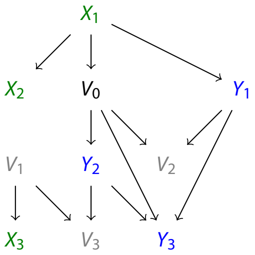
Pearl's back-door criterion vs. our CBC

Identification of causal effects: adjustment in DAGs

			Runtime
Verification: For given **X**, **Y**, **Z** and constraint **I** decide if . . .			
TestAdj	**Z** is an adjustment for (**X**, **Y**)		$\mathcal{O}(n+m)$
TestMinAdj	**Z** $\supseteq$ **I** is an adjustment for (**X**, **Y**) and **Z** is . . .		
		I-minimal	$\mathcal{O}(n^2)$
		strongly-minimal	$\mathcal{O}(n^2)$
Construction: For given **X**, **Y** and constraints **I**, **R**, output an . . .			
FindAdj	adjustment **Z** for (**X**, **Y**) with **I** $\subseteq$ **Z** $\subseteq$ **R**		$\mathcal{O}(n+m)$
FindMinAdj	adjustment **Z** for (**X**, **Y**) with **I** $\subseteq$ **Z** $\subseteq$ **R** which is . . .		
		I-minimal	$\mathcal{O}(n^2)$
		strongly-minimal	NP-hard
FindMinCostAdj	adjustment **Z** for (**X**, **Y**) with **I** $\subseteq$ **Z** $\subseteq$ **R** which is . . .		
		I-minimum	$\mathcal{O}(n^3)$
		strong-minimum	$\mathcal{O}(n^3)$
Enumeration: For given **X**, **Y**, **I**, **R** enumerate all . . .			Delay
ListAdj	adjustments **Z** for (**X**, **Y**) with **I** $\subseteq$ **Z** $\subseteq$ **R**		$\mathcal{O}(n(n+m))$
ListMinAdj	**I**-minimal adjsutments **Z** with **I** $\subseteq$ **Z** $\subseteq$ **R**		$\mathcal{O}(n^3)$

Identification of causal effects: adjustment in DAGs

- Due to the high efficiency of our algorithms we were able, for the first time, to *quantitatively analyze* in how many cases the adjustment can permit identification.
- To determine the *general identifiability* of causal effects, we use the IDC algorithm.
- Two major drawbacks of the IDC algorithm:
  - its time complexity is much larger than the runtime of our methods
  - it can return complicated identification formulas, e.g.

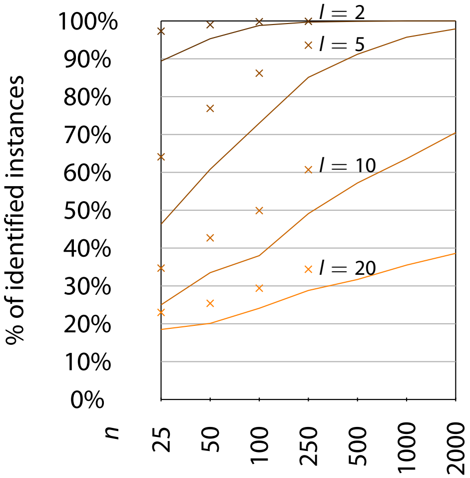


The Instance is identified by using the empty set and by the formula

$$\sum_{v_0}[P(y_1|x_1)P(v_0|x_1)P(y_3|x_1,y_1,v_0,y_2)P(y_2|x_1,v_0)]$$

found by the ID-algorithm.

Identification of causal effects: adjustment in DAGs



Percent of identifiable instances by adjustment or plain formulas (depicted as curves), resp. by do-calculus (crosses).

Identification of causal effects: adjustment in DAGs
Extensions of our techniques

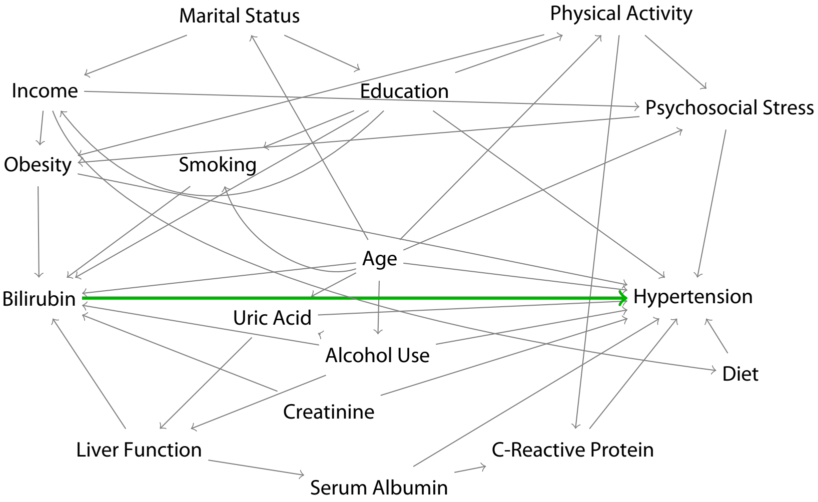
Graph class	**sound and complete crit.**	**sound and complete constructive crit.**
MAGs	van der Zander, Liśkiewicz, Textor ’14	van der Zander, Liśkiewicz, Textor ’14
CPDAGs	Perković et al. ’15	van der Zander and Liśkiewicz ’16
PAGs	Perković et al. ’15	Perković et al. ’16
CGs	van der Zander and Liśkiewicz ’16	van der Zander and Liśkiewicz ’16

Scientific software *DAGitty* for analyzing causal models

- We have implement our algorithmic solutions and developed further the package DAGitty (`www.dagitty.net`).
- Can process and analyse complex causal structures.
- Is currently widely used, especially in
  - epidemiology,
  - psychology, and
  - economical science.
- Currently the page is visited by ca. 250 users per Day / ca. 8000 users Monthly.

Scientific software *DAGitty* for analyzing causal models

Serum bilirubin and the risk of hypertension analysis with *dagitty*



[Wang et al., Int J Epidemiol 2015]

Conclusions

- Causal inference from *observed data*: causal structure + do-operator
- Our research: algorithmic estimations of causal effects
  - *covariate adjustment* (nonparametric causal model)
  - *instrumental variable* (linear systems)
- We provide easily implementable algorithms
- Scientific software *DAGitty*

Selected Current Papers

1. B. van der Zander, M. Liśkiewicz: *Finding Minimal d-separators in Linear Time and Applications*. Proc. of the 35th Conf. on Uncertainty in Artificial Intelligence (UAI'19), AUAI Press, 2019.
2. B. van der Zander, M. Liśkiewicz, J. Textor: *Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework*, Artificial Intelligence 270 (2019): 1-40.
3. J. Textor, B. van der Zander, M.S. Gilthorpe, M. Liśkiewicz, G.T.H. Ellison: *Robust causal inference using Directed Acyclic Graphs: the R package 'dagitty'*. International Journal of Epidemiology, 6(45):1887-1894, 2017.
4. B. van der Zander, M. Liśkiewicz: *On Searching for Generalized Instrumental Variables*. Proc. of the The 19th Int. Conf. on Artificial Intelligence and Statistics (AISTATS'16), pp. 1214-1222, JMLR Proceedings, 2016.
5. B. van der Zander, M. Liśkiewicz: *Separators and Adjustment Sets in Markov Equivalent DAGs*. Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI'16), pp. 3315-3321, AAAI Press, 2016.
6. B. van der Zander, J. Textor, M. Liśkiewicz: *Efficiently Finding Conditional Instruments for Causal Inference*. Proc. of the 24th Int. Joint Conf. on Artificial Intelligence (IJCAI'15), pp. 3243-3249. AAAI Press / International Joint Conferences on Artificial Intelligence, 2015.
7. J. Textor, A. Idelberger, M. Liśkiewicz: *On the Faithful DAGs of a Dependency Graph*. Proc of the 31st Conf. on Uncertainty in Artificial Intelligence (UAI'15), pp. 882-891. AUAI Press, 2015.
8. B. van der Zander, J. Textor, M. Liśkiewicz: *Constructing Separators and Adjustment Sets in Ancestral Graphs*. Proc. of the 30th Conf. on Uncertainty in Artificial Intelligence (UAI'14), pp. 907-916. AUAI Press, 2014 (**IBM Best Student Paper Award for B.Z.**)

Research supported by the Deutsche Forschungsgemeinschaft (DFG) grant LI 634/4-1 and -2.