Characterizing and constructing adjustment sets

Emilija Perković, ETH Zurich

Joint work with Johannes Textor, Markus Kalisch and Marloes Maathuis

Causal effects

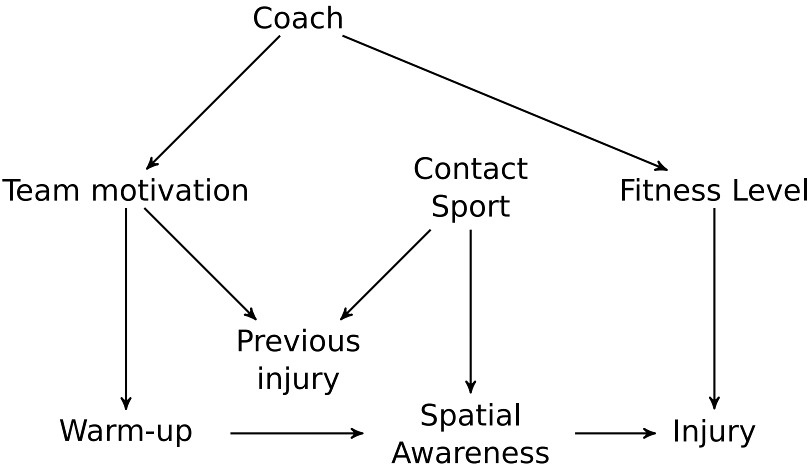


Figure: DAG $\mathcal{D}$; cf. Shrier and Platt, 2008.

Causal effects

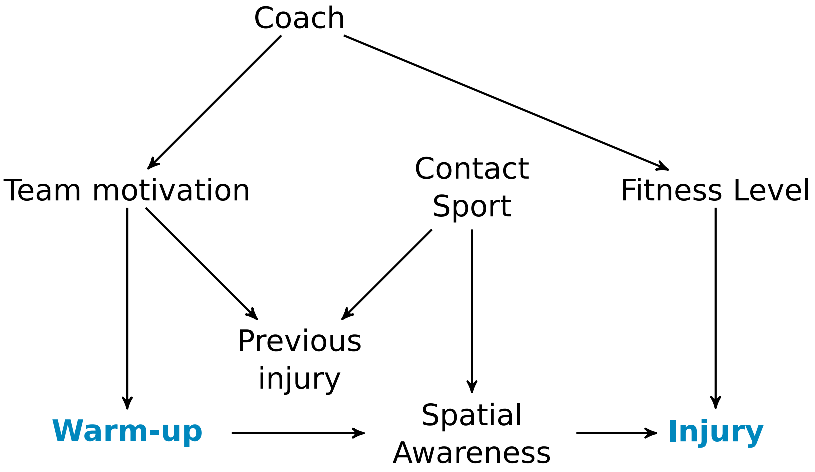


Figure: DAG $\mathcal{D}$; cf. Shrier and Platt, 2008.

Goal

- Estimate the total causal effect of **X** on **Y**

- $do(\mathbf{x})$: an intervention that sets variables **X** to **x**.

Observational data	Randomized control studies

Goal

- Estimate the total causal effect of $\mathbf{X}$ on $\mathbf{Y}$
  - the average change in $\mathbf{Y}$ due to $do(\mathbf{x})$ -

- $do(\mathbf{x})$: an intervention that sets variables $\mathbf{X}$ to $\mathbf{x}$.

Observational data

Randomized control studies

Goal

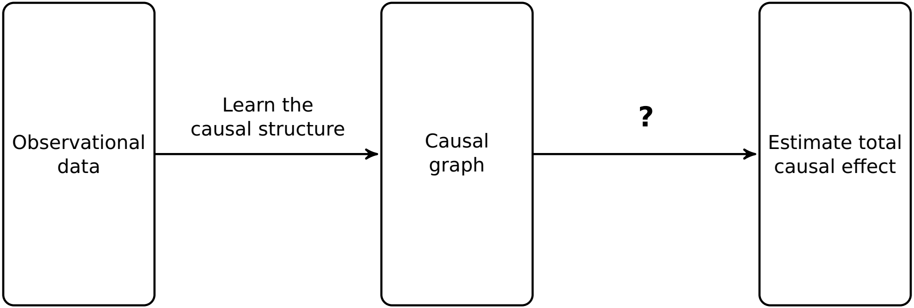
- Estimate the total causal effect of $\mathbf{X}$ on $\mathbf{Y}$
  - the average change in $\mathbf{Y}$ due to $do(\mathbf{x})$ -
  from observational data.

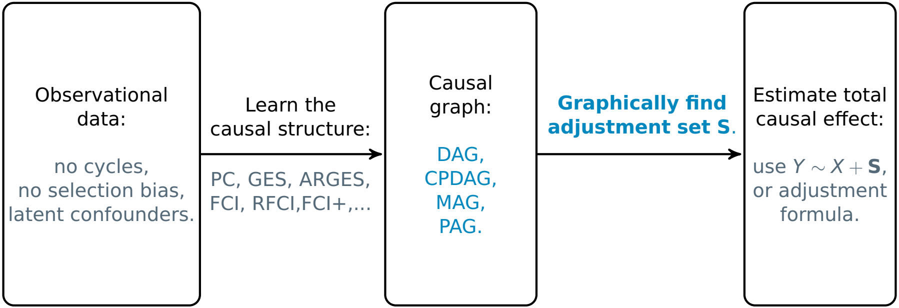
- $do(\mathbf{x})$: an intervention that sets variables $\mathbf{X}$ to $\mathbf{x}$.

Observational data

Randomized control studies

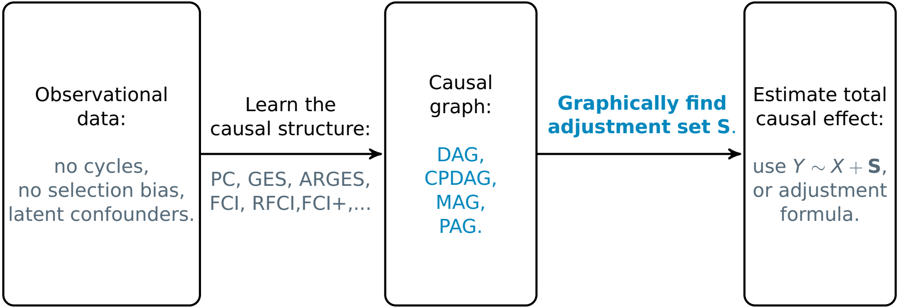
Framework


Observational data
Learn the causal structure
Causal graph
?
Estimate total causal effect



- PC (Spirtes et al, 1993), GES (Chickering, 2002), ARGES (Nandy et al, 2016).
- FCI (Spirtes et al 1993, Zhang 2008), RFCI (Colombo et al, 2012), FCI+ (Claassen et al, 2013).

Framework



- Causal effects are often estimated by adjusted regression.
- Adjustment sets depend on the causal structure, which can be represented by a graph.

What is an adjustment set?

- DAG: Directed Acyclic Graph.

What is an adjustment set?

- DAG: Directed Acyclic Graph.
- A prob. density f is compatible with the DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | pa(x_j, \mathcal{D}))$

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.
- A prob. density f is compatible with the causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | pa(x_j, \mathcal{D}))$ and $f(\mathbf{v} | do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j | pa(x_j, \mathcal{D}))$.

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.
- A prob. density f is compatible with the causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|pa(x_j, \mathcal{D}))$ and $f(\mathbf{v}|do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j|pa(x_j, \mathcal{D}))$.
- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in causal DAG $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x}, \mathbf{s})\} & \text{otherwise.} \end{cases}$$

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.
- A prob. density f is compatible with the causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | pa(x_j, \mathcal{D}))$ and $f(\mathbf{v} | do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j | pa(x_j, \mathcal{D}))$.
- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in causal DAG $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y} | do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x}, \mathbf{s})\} & \text{otherwise.} \end{cases}$$

- The total causal effect of X on Y can be defined: $\frac{\partial}{\partial x} E(Y | do(x))$.

What is an adjustment set?

- (causal) DAG: (causal) Directed Acyclic Graph.
- A prob. density f is compatible with the causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if: $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|pa(x_j, \mathcal{D}))$ and $f(\mathbf{v}|do(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{X}} f(x_j|pa(x_j, \mathcal{D}))$.
- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in causal DAG $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x}, \mathbf{s})\} & \text{otherwise.} \end{cases}$$

- The total causal effect of X on Y can be defined: $\frac{\partial}{\partial x} E(Y|do(x))$.

- In a linear setting the total causal effect of X on Y is then the linear regression coefficient of X in the regression $Y \sim X + \mathbf{S}$.

Which sets **S** are adjustment sets?

Adjusting for **too many** or **too few** variables leads to **bias**.

Some intuition for DAGs:

(1) $X \longrightarrow Z \longrightarrow Y$

(2) $X \longleftarrow Z \longrightarrow Y$

(3) $X \longleftarrow Z \longleftarrow Y$

(4) $X \longrightarrow Z \longleftarrow Y$

Which sets **S** are adjustment sets?

Adjusting for **too many** or **too few** variables leads to **bias**.

Some intuition for DAGs:

(1) $X \longrightarrow Z \longrightarrow Y$

(2) $X \longleftarrow Z \longrightarrow Y$

(3) $X \longleftarrow Z \longleftarrow Y$

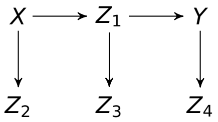
(4) $X \longrightarrow Z \longleftarrow Y$

Answer: (1) $\emptyset$; (2) $\{Z\}$; (3) $\{Z\}$; (4) $\emptyset$

Do not disturb causal paths,
block all non-causal paths.

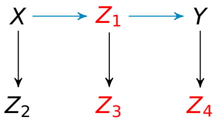
Which sets **S** are adjustment sets?

Some more intuition:



Which sets **S** are adjustment sets?

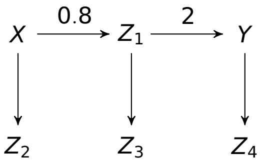
Some more intuition:



Answer: $\emptyset$ or $\{Z_2\}$

Descendants of nodes on a causal path (except of X) are forbidden.
(Each node is a descendant of itself)

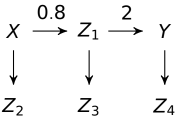
R example




```
n  <- 100000
eps <- matrix(rnorm(6*n,0,1), ncol=6)
X  <- eps[,1]
Z1 <- 0.8*X + eps[,2]
Y  <- 2*Z1 + eps[,3]
Z2 <- X + eps[,4]
Z3 <- Z1 + eps[,5]
Z4 <- Y + eps[,6]
```

The total effect of X on Y is $0.8 \cdot 2 = 1.6$.

R example




```
> lm(Y~X)$coeff[2]
1.598732
> lm(Y~X+Z1)$coeff[2]
0.003260167
> lm(Y~X+Z2)$coeff[2]
1.596347
> lm(Y~X+Z3)$coeff[2]
0.7985709
> lm(Y~X+Z4)$coeff[2]
0.2730853
> lm(Y~X+Z2+Z3)$coeff[2]
0.7996412
```

Existing results for DAGs

- Back-door criterion (Pearl, 1993)
  - **sufficient**
- Adjustment criterion (Shpitser et al, 2010)
  - **necessary and sufficient**

Existing results for DAGs

- Back-door criterion (Pearl, 1993)
  - **sufficient**
- Adjustment criterion (Shpitser et al, 2010)
  - **necessary and sufficient**

Given a DAG $\mathcal{D}$, disjoint sets $\mathbf{X}$,$\mathbf{Y}$ and $\mathbf{S}$ can **check** whether $\mathbf{S}$ satisfies back-door/adjustment criterion.

Existing results for DAGs

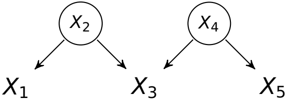
- Back-door criterion (Pearl, 1993)
  - **sufficient**
- Adjustment criterion (Shpitser et al, 2010)
  - **necessary and sufficient**

Given a DAG $\mathcal{D}$, disjoint sets $\mathbf{X}$,$\mathbf{Y}$ and $\mathbf{S}$ can **check** whether $\mathbf{S}$ satisfies back-door/adjustment criterion.

- The adjustment criterion will give you **all** adjustment sets.
- The back-door criterion will give you **some** adjustment sets.

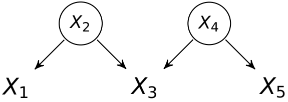
We cannot always learn a DAG

DAG:

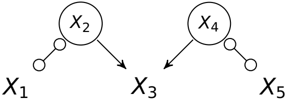


We cannot always learn a DAG

DAG:

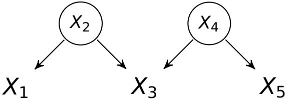


Unknown causal directions (CPDAG):

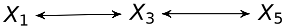


We cannot always learn a DAG

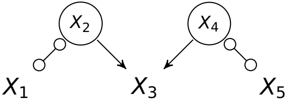
DAG:



Unobserved confounders (MAG):

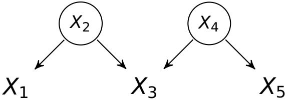


Unknown causal directions (CPDAG):

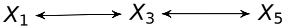


We cannot always learn a DAG

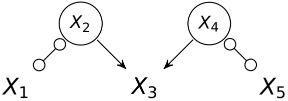
DAG:



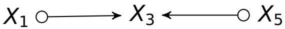
Unobserved confounders (MAG):



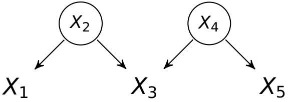
Unknown causal directions (CPDAG):



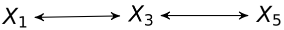
Unobserved confounders and unknown causal directions (PAG):



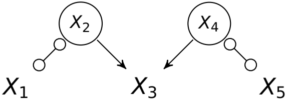
DAG:



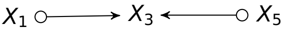
Unobserved confounders (MAG):



Unknown causal directions (CPDAG):

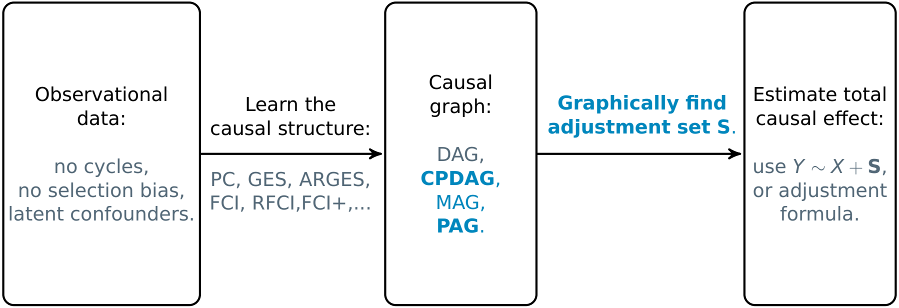


Unobserved confounders and unknown causal directions (PAG):



Can be learned from observational data.

Focus on CPDAGs and PAGs



- **CPDAG**: PC (Spirtes et al, 1993), GES (Chickering, 2002), ARGES (Nandy et al, 2016).
- **PAG**: FCI (Spirtes et al 1993, Zhang 2008), RFCI (Colombo et al, 2012), FCI+ (Claassen et al, 2013).

Overview of graphical criteria for adjustment

Can be learned from observational data (CPDAG, PAG)

	DAG	MAG	CPDAG	PAG
Back-door (Pearl '93)	$\Rightarrow$			
Adjustment (Shpitser et al '10)	$\Leftrightarrow$			
Adjustment (Van der Zander et al '14)	$\Leftrightarrow$	$\Leftrightarrow$		
Generalized back-door (Maathuis & Colombo '15)	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$
Generalized adjustment (Perkovic et al '15)	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$

$\Rightarrow$ - sufficient, $\Leftrightarrow$ - necessary and sufficient

Generalized adjustment criterion

Theorem (Perković et al, 2015):
S is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if:

Amenability	$\mathcal{G}$ is **amenable** relative to $(\mathbf{X}, \mathbf{Y})$.
Forbidden Set	**S** does not contain nodes in **Forbidden**$(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.
Blocking	**S** blocks all **proper non-causal definite status** paths from **X** to **Y** in $\mathcal{G}$.

For any DAG/CPDAG/MAG/PAG $\mathcal{G}$ and node sets **X**,**Y** and **S**, we can **check** whether **S** is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$.

Further questions

(1) Does an adjustment set always exist?

(2) Can we construct adjustment sets using a fast algorithm?

Further questions

R package dagitty on CRAN.

(1) Does an adjustment set always exist? **No.**

Theorem (Perković et al, 2016)

There exists an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if and only if adjustmentSets($\mathcal{G}, \mathbf{X}, \mathbf{Y}$,type="canonical") returns a set.

(2) Can we construct adjustment sets using a fast algorithm?

Further questions

R package dagitty on CRAN.

(1) Does an adjustment set always exist? **No.**

Theorem (Perković et al, 2016)
There exists an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ if and only if `adjustmentSets(`$\mathcal{G}, \mathbf{X}, \mathbf{Y}$`,type="canonical")` returns a set.

(2) Can we construct adjustment sets using a fast algorithm?

Yes.
Van der Zander et al, 2014 for DAGs and MAGs.
Perković et al, 2016 for CPDAGs and PAGs.

Implementation

Algorithms implemented in `R` package `dagitty` on CRAN:

Implementation

Algorithms implemented in `R` package `dagitty` on CRAN:

- `adjustmentSets(`$\mathcal{G}$, $\mathbf{X}$, $\mathbf{Y}$`,type="canonical")` runs in $O(|\mathbf{V}|+|\mathbf{E}|)$ runtime.

Implementation

Algorithms implemented in R package `dagitty` on CRAN:

- `adjustmentSets(`$\mathcal{G}, \mathbf{X}, \mathbf{Y}$`,type="canonical")` runs in O(|**V**|+|**E**|) runtime.
- `isAdjustmentSet(`$\mathcal{G}, \mathbf{S}, \mathbf{X}, \mathbf{Y}$`)` checks whether **S** is an adjustment set for $(\mathbf{X}, \mathbf{Y})$ and $\mathcal{G}$ in O(|**V**|+|**E**|) runtime.

Implementation

Algorithms implemented in R package dagitty on CRAN:

- `adjustmentSets(`$\mathcal{G}$, **X**, **Y**`,type="canonical")` runs in O(|**V**|+|**E**|) runtime.
- `isAdjustmentSet(`$\mathcal{G}$, **S**, **X**, **Y**`)` checks whether **S** is an adjustment set for (**X**, **Y**) and $\mathcal{G}$ in O(|**V**|+|**E**|) runtime.
- `adjustmentSets(`$\mathcal{G}$, **X**, **Y**`,type="all")` lists all (or all minimal, if `type="minimal"`) adjustment sets for (**X**, **Y**) and $\mathcal{G}$ in O(|**V**|(|**V**|+|**E**|)) runtime per set.

Our contribution

For CPDAGs and PAGs $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ - the output of most causal structure learning algorithms - we developed:

- A necessary and sufficient graphical criterion for finding adjustment sets.
- An algorithm that finds an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ if there is one in $O(|\mathbf{V}|+|\mathbf{E}|)$ runtime.
- An algorithm that finds all (minimal) adjustment sets relative to $(\mathbf{X}, \mathbf{Y})$ in $O(|\mathbf{V}|(|\mathbf{V}|+|\mathbf{E}|))$ runtime per set.

Thanks!

Joint work with
Marloes Maathuis, Markus Kalisch, Johannes Textor

References:

- Perković, Textor, Kalisch and Maathuis (2015). A complete generalized adjustment criterion. *UAI 2015*.
- Perković, Textor, Kalisch and Maathuis (2016). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *arXiv:1606.06903*.