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 $Y$

- $do(x)$: an intervention that sets variables $X$ to $x$. 

Observational data

Randomized control studies
Goal

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

- $do(x)$: an intervention that sets variables $X$ to $x$. 

Observational data

Randomized control studies
Goal

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

- $do(x)$: an intervention that sets variables $X$ to $x$.
Framework

Observational data

Learn the causal structure

Causal graph

? Estimate total causal effect
Framework

Observational data:
- no cycles, no selection bias, latent confounders.

Learn the causal structure:
- PC, GES, ARGES, FCI, RFCI, FCI+, ...

Causal graph:
- DAG, CPDAG, MAG, PAG.

Graphically find adjustment set $S$.

Estimate total causal effect:
- use $Y \sim X + S$, or adjustment formula.

- PC (Spirtes et al, 1993), GES (Chickering, 2002), ARGES (Nandy et al, 2016).
Observational data:
- no cycles, no selection bias, latent confounders.

Learn the causal structure:
- PC, GES, ARGES, FCI, RFCI, FCI+, ...

Causal graph:
- DAG, CPDAG, MAG, PAG.

Graphically find adjustment set $S$.

Estimate total causal effect:
- use $Y \sim X + S$, or adjustment formula.

- 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.

A probability density function $f$ is compatible with the causal DAG $D = (V, E)$ if:

$$f(v) = \prod_{j=1}^{p} f(x_j | \text{pa}(x_j, D))$$

and

$$f(v | \text{do}(x)) = \prod_{X_j \in V \setminus X} f(x_j | \text{pa}(x_j, D)).$$

**$S$** is an adjustment set relative to $(X, Y)$ in causal DAG $D$ if for any $f$ compatible with $D$:

$$f(y | \text{do}(x)) = \begin{cases} f(y | x) & \text{if } S = \emptyset, \\ \int_S f(y | x, s) f(s) \, ds = \mathbb{E}_S \{ f(y | x, s) \} & \text{otherwise.} \end{cases}$$

The total causal effect of $X$ on $Y$ can be defined:

$$\frac{\partial}{\partial x} \mathbb{E}(Y | \text{do}(x)).$$

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

Emilija Perković, ETH Zurich
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:
  \[
  f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|\text{pa}(x_j, \mathcal{D}))
  \]

- $S$ is an adjustment set relative to $(X, Y)$ in causal DAG $\mathcal{D}$ if for any $f$ compatible with $\mathcal{D}$:
  \[
  f(y|\text{do}(x)) = \begin{cases} 
  f(y|x) & \text{if } S = \emptyset, \\
  \int S f(y|x, s) f(s) \, ds = \mathbb{E}_S \{ f(y|x, s) \} & \text{otherwise.}
  \end{cases}
  \]

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

- In a linear setting the total causal effect of $X$ on $Y$ is then the linear regression coefficient of $X$ in the regression $Y \sim X + S$. 
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:
  \[
  f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|\text{pa}(x_j, \mathcal{D})) \quad \text{and} \quad f(\mathbf{v}|\text{do}(\mathbf{x})) = \prod_{X_j \in \mathbf{V} \setminus \mathbf{x}} f(x_j|\text{pa}(x_j, \mathcal{D})).
  \]
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:
  
  \[
  f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | \text{pa}(x_j, \mathcal{D})) \quad \text{and} \quad f(\mathbf{v} | do(\mathbf{x})) = \prod_{x_j \in \mathbf{V} \setminus \mathbf{x}} f(x_j | \text{pa}(x_j, \mathcal{D})).
  \]
- $\mathbf{S}$ 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:
  
  $$
  f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j | \text{pa}(x_j, \mathcal{D})) \quad \text{and} \quad f(\mathbf{v}| \text{do}(\mathbf{x})) = \prod_{x_j \in \mathbf{V} \setminus \mathbf{x}} f(x_j | \text{pa}(x_j, \mathcal{D})).
  $$

- $\mathbf{S}$ is an **adjustment set** relative to $(\mathbf{X}, \mathbf{Y})$ in causal DAG $\mathcal{D}$ if for any $f$ compatible with $\mathcal{D}$:

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

- The **total causal effect** of $\mathbf{X}$ on $\mathbf{Y}$ can be defined: $\frac{\partial}{\partial \mathbf{x}} E(\mathbf{Y}| \text{do}(\mathbf{x}))$. 
What is an adjustment set?

• (causal) DAG: (causal) Directed Acyclic Graph.

• A prob. density $f$ is compatible with the causal DAG $D = (V, E)$ if:
  
  $$
  f(v) = \prod_{j=1}^{p} f(x_j|pa(x_j, D)) \quad \text{and} \quad f(v|do(x)) = \prod_{x_j \in V \setminus X} f(x_j|pa(x_j, D))
  $$

• $S$ is an adjustment set relative to $(X, Y)$ in causal DAG $D$ if for any $f$ compatible with $D$:
  
  $$
  f(y|do(x)) = \begin{cases} 
  f(y|x) & \text{if } S = \emptyset, \\
  \int_S f(y|x, s)f(s)ds = E_S\{f(y|x, s)\} & \text{otherwise.}
  \end{cases}
  $$

• 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 $Y \sim X + S$. 
Which sets $S$ are adjustment sets?

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

Some intuition for DAGs:

1. $X \rightarrow Z \rightarrow Y$
2. $X \leftarrow Z \rightarrow Y$
3. $X \leftarrow Z \leftarrow Y$
4. $X \rightarrow Z \leftarrow Y$
Which sets $S$ are adjustment sets?

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

Some intuition for DAGs:

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

(2) $X \leftarrow Z \rightarrow Y$

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

(4) $X \rightarrow Z \leftarrow Y$

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:

\[ X \rightarrow Z_1 \rightarrow Y \]

\[ Z_2 \quad Z_3 \quad Z_4 \]

Descendants of nodes on a causal path (except of $X$) are forbidden.

(Each node is a descendant of itself)
Which sets $S$ are adjustment sets?

Some more intuition:

\[\begin{array}{c}
\overset{X}{\longleftarrow}\overset{Z_1}{\rightarrow}\overset{Y}{\rightarrow} \\
\downarrow\downarrow\downarrow \\
Z_2\quad Z_3\quad Z_4
\end{array}\]

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

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

\[
\begin{align*}
X & \rightarrow Z_1 \rightarrow Y \\
& \downarrow \quad \downarrow \quad \downarrow \\
Z_2 & \quad Z_3 \quad Z_4
\end{align*}
\]

\begin{verbatim}
> 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
\end{verbatim}
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 $D$, disjoint sets $X, Y$ and $S$ can check whether $S$ 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 $D$, disjoint sets $X,Y$ and $S$ can check whether $S$ 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:

```
    X_2
   / \
X_1   X_3
    \
     \
    X_4
   / \
X_5
```
We cannot always learn a DAG

DAG:

```
X_1 → X_2 → X_3 → X_4 → X_5
```

Unknown causal directions (CPDAG):

```
X_1 → X_2, X_4 → X_3 → X_5
```

Can be learned from observational data.
We cannot always learn a DAG

DAG:

Unobserved confounders (MAG):

Unknown causal directions (CPDAG):

\( X_1 \leftrightarrow X_3 \leftrightarrow X_5 \)
We cannot always learn a DAG

DAG:

Unobserved confounders (MAG):

Unknown causal directions (CPDAG):

Unobserved confounders and unknown causal directions (PAG):

Emilija Perković, ETH Zurich
Characterizing and constructing adjustment sets 16 / 23
We cannot always learn a DAG

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

Observational data:
- no cycles,
- no selection bias,
- latent confounders.

Learn the causal structure:
- PC, GES, ARGES, FCI, RFCI, FCI+,...

Causal graph:
- DAG, CPDAG, MAG, PAG.

Graphically find adjustment set $S$.

Estimate total causal effect:
- use $Y \sim X + S$, or adjustment formula.

- **CPDAG:** PC (Spirtes et al, 1993), GES (Chickering, 2002), ARGES (Nandy et al, 2016).
Overview of graphical criteria for adjustment

<table>
<thead>
<tr>
<th>Back-door (Pearl ’93)</th>
<th>DAG</th>
<th>MAG</th>
<th>CPDAG</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇒</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjustment (Shpitser et al ’10)</th>
<th>DAG</th>
<th>MAG</th>
<th>CPDAG</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjustment (Van der Zander et al ’14)</th>
<th>DAG</th>
<th>MAG</th>
<th>CPDAG</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇔</td>
<td></td>
<td>⇔</td>
<td>⇔</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generalized back-door (Maathuis &amp; Colombo ’15)</th>
<th>DAG</th>
<th>MAG</th>
<th>CPDAG</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇒</td>
<td>⇔</td>
<td>⇔</td>
<td>⇔</td>
<td>⇒</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generalized adjustment (Perkovic et al ’15)</th>
<th>DAG</th>
<th>MAG</th>
<th>CPDAG</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇔</td>
<td>⇔</td>
<td>⇔</td>
<td>⇔</td>
<td>⇔</td>
</tr>
</tbody>
</table>

⇒ - sufficient, ⇔ - necessary and sufficient

Can be learned from observational data
Generalized adjustment criterion

Theorem (Perković et al, 2015):

\( S \) is an adjustment set relative to \((X, Y)\) and \( \mathcal{G} \) if:

Amenability \( \mathcal{G} \) is **amenable** relative to \((X, Y)\).

Forbidden Set \( S \) does not contain nodes in **Forbidden** \((X, Y, \mathcal{G})\).

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 \((X, Y)\).
Further questions

(1) Does an adjustment set always exist?

(2) Can we construct adjustment sets using a fast algorithm?
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 \((X, Y)\) and \(G\) if and only if `adjustmentSets(G, X, 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 \((X, Y)\) and \(G\) if and only if `adjustmentSets(G, X, 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.
Algorithms implemented in \texttt{R} package \texttt{dagitty} on CRAN:

\begin{itemize}
  \item \texttt{adjustmentSets(G, X, Y, type="canonical")} runs in $O(|V|+|E|)$ runtime.
  \item \texttt{isAdjustmentSet(G, S, X, Y)} checks whether $S$ is an adjustment set for $(X, Y)$ and $G$ in $O(|V|+|E|)$ runtime.
  \item \texttt{adjustmentSets(G, X, Y, type="all")} lists all (or all minimal, if \texttt{type="minimal"}) adjustment sets for $(X, Y)$ and $G$ in $O(|V|^2(|V|+|E|))$ runtime per set.
\end{itemize}
Algorithms implemented in R package *dagitty* on CRAN:

- `adjustmentSets(G, X, Y, type="canonical")` runs in $O(|V|+|E|)$ runtime.
- `isAdjustmentSet(G, S, X, Y)` checks whether $S$ is an adjustment set for $(X, Y)$ and $G$ in $O(|V|+|E|)$ runtime.
- `adjustmentSets(G, X, Y, type="all")` lists all (or all minimal, if `type="minimal"`) adjustment sets for $(X, Y)$ and $G$ in $O(|V|(|V|+|E|))$ runtime per set.
Algorithms implemented in \(\texttt{R}\) package \texttt{dagitty} on CRAN:

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

- \texttt{isAdjustmentSet(\(\mathcal{G}\), \texttt{S}, \texttt{X}, \texttt{Y})}\) checks whether \texttt{S} is an adjustment set for \((\texttt{X}, \texttt{Y})\) and \(\mathcal{G}\) in \(O(|V|+|E|)\) runtime.
Implementation

Algorithms implemented in \( \texttt{R} \) package \texttt{dagitty} on CRAN:

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

- \texttt{isAdjustmentSet(} \( G, S, X, Y \) \) checks whether \( S \) is an adjustment set for \( (X, Y) \) and \( G \) in \( O(|V|+|E|) \) runtime.

- \texttt{adjustmentSets(} \( G, X, Y \), \texttt{type="all"} \) lists all (or all minimal, if \texttt{type="minimal"}) adjustment sets for \( (X, Y) \) and \( G \) 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 $(X, Y)$ if there is one in $O(|\mathbf{V}|+|\mathbf{E}|)$ runtime.

- An algorithm that finds all (minimal) adjustment sets relative to $(X, Y)$ in $O(|\mathbf{V}|(|\mathbf{V}|+|\mathbf{E}|))$ runtime per set.
Thanks!

Joint work with Marloes Maathuis, Markus Kalisch, Johannes Textor

References: