Causal inference and control

Week 4:

Identification and adjustment

Søren Wengel Mogensen
Department of Automatic Control
soren.wengel_mogensen@control.lth.se

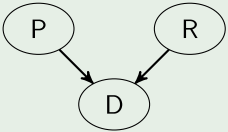
Assignment 1

Berkson's paradox

Berkson's paradox is an example of *collider bias* and is illustrated by the next example.

Example

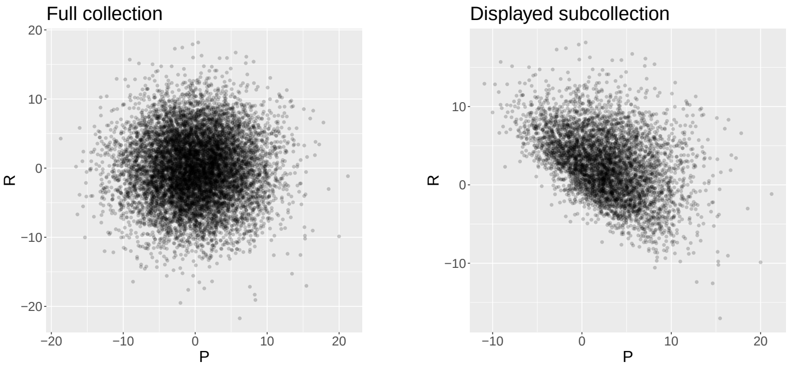
Assume stamps can be PRETTY and RARE, and that these properties are independent in the collection of some diligent collector. The collector now chooses a subset of stamps to put on display, and the collector prefers PRETTY and RARE stamps.



$$P = N_P$$
$$R = N_R$$
$$D = 1_{N_D < \exp(P+R)/(1+\exp(P+R))}$$

where $N_P, N_R \sim N(0, 25)$, and $N_D \sim \mathcal{U}[0, 1]$.

Berkson's paradox



Causal identification

Causal identification includes the task of computing interventional quantities from the observational distribution. We need to specify the quantity we want to identify, and what we assume known (which distributions are known, which parts of the graph are known, etc). We saw in Week 1 that we can compute the interventional distributions from the observational distribution when we know the graph. Today, we will discuss conditions that allow causal identification even if some variables are unobserved.

Causal identification

Causal identification includes the task of computing interventional quantities from the observational distribution. We need to specify the quantity we want to identify, and what we assume known (which distributions are known, which parts of the graph are known, etc). We saw in Week 1 that we can compute the interventional distributions from the observational distribution when we know the graph. Today, we will discuss conditions that allow causal identification even if some variables are unobserved.

We often write $p^{C,do(X=x)}(y)$ as simply $p^{do(x)}(y)$.

Adjustment

Definition (Valid adjustment set)

We consider a structural causal model on nodes $\boldsymbol{V}$. Let $X, Y \in \boldsymbol{V}$ such that $Y \notin \boldsymbol{PA}_X$. We say that $\boldsymbol{Z} \subseteq \boldsymbol{V} \setminus \{X, Y\}$ is a *valid adjustment set* for (X, Y) if

$$p^{do(x)}(y) = \sum_{z} p(y|x, \boldsymbol{z})p(\boldsymbol{z}).$$

A valid adjustment set allows a simple type of identification.

Confounding

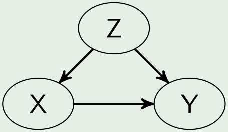
We say that the causal effect from X to Y is *confounded* if the empty set is not a valid adjustment set.

Definition (Confounding)

Let $X, Y \in \boldsymbol{V}$ such that there is a directed path from X to Y. We say that the causal effect from X to Y is *confounded* if

$$p^{do(x)}(y) \neq p(y|x).$$

Example (Confounding)



Truncated factorization gives

$$p^{do(x)}(y) = \sum_z p(y|x,z)p(z).$$

Finding valid adjustment sets

We explain how one finds valid adjustment sets. Note that if

$$p^{do(x)}(y|x, \boldsymbol{z}) = p(y|x, \boldsymbol{z}), \qquad p^{do(x)}(\boldsymbol{z}) = p(\boldsymbol{z}) \tag{1}$$

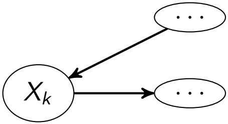
then

$$\begin{aligned} p^{do(x)}(y) &= \sum_{\mathbf{z}} p^{do(x)}(y, \boldsymbol{z}) = \sum_{\mathbf{z}} p^{do(x)}(y, x, \boldsymbol{z}) = \sum_{\mathbf{z}} p^{do(x)}(y|x, \boldsymbol{z}) p^{do(x)}(x, \boldsymbol{z}) \\ &= \sum_{\mathbf{z}} p^{do(x)}(y|x, \boldsymbol{z}) p^{do(x)}(\boldsymbol{z}) = \sum_{\mathbf{z}} p(y|x, \boldsymbol{z}) p(\boldsymbol{z}). \end{aligned}$$

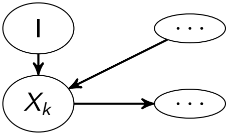
The invariance of conditionals in Equation (1) is sufficient for $\boldsymbol{Z}$ to be a valid adjustment set for (X, Y).

Finding valid adjustment sets

We consider now an SCM, C, and an intervention $do(X_k = x_k)$. We construct a new SCM, C^*, by adding a (non-degenerate) binary variable I which is a parent of X_k.



$$X_j = f_j(\boldsymbol{PA}_j, N_j)$$



$$\begin{aligned} I &= N_I, \\ X_j &= f_j(\boldsymbol{PA}_j, N_j), \quad j \neq k \\ X_k &= \begin{cases} f_k(\boldsymbol{PA}_k, N_k) & \text{if } I = 0 \\ x_k & \text{otherwise} \end{cases} \end{aligned}$$

Finding valid adjustment sets

Using that I is a source node, we have

$$
\begin{aligned}
p^{C^*}(x_1, \ldots, x_d | I = 0) &= \frac{1}{p^{C^*}(I = 0)} \left(\prod_j p^{C^*}(x_j | \boldsymbol{PA}_j^{C^*}) \right) p^{C^*}(I = 0) \\
&= p^{C^*, do(I=0)}(x_1, \ldots, x_d) = p^{C}(x_1, \ldots, x_d) \\
p^{C^*}(x_1, \ldots, x_d | I = 1) &= p^{C^*, do(I=1)}(x_1, \ldots, x_d) = p^{C, do(X_k = x_k)}(x_1, \ldots, x_d)
\end{aligned}
$$

Finding valid adjustment sets

If we now have a variable A and a set of variables $\boldsymbol{B}$ such that $A \perp\!\!\!\perp_{\mathcal{G}^*} I \mid \boldsymbol{B}$, then $p^{C^*}(a|b, I=0) = p^{C^*}(a|b, I=1)$, and therefore

$$p^{C}(a|\boldsymbol{b}) = p^{C,do(X_k=x_k)}(a|\boldsymbol{b}).$$

This means that Equation (1),

$$p^{do(x)}(y|x,\boldsymbol{z}) = p(y|x,\boldsymbol{z}), \qquad p^{do(x)}(\boldsymbol{z}) = p(\boldsymbol{z}),$$

is satisfied if

$$Y \perp\!\!\!\perp_{\mathcal{G}^*} I \mid X, \boldsymbol{Z}, \qquad \boldsymbol{Z} \perp\!\!\!\perp_{\mathcal{G}^*} I$$

Finding valid adjustment sets

Let $\boldsymbol{Z}$ denote the parent set of X. In this case,

$$Y \perp\!\!\!\perp_{\mathcal{G}^*} I \mid X, \boldsymbol{Z}, \qquad \boldsymbol{Z} \perp\!\!\!\perp_{\mathcal{G}^*} I$$

The second statement follows from the fact that every path between $\boldsymbol{Z}$ and I must contain a collider as

$$I \to X \to \ldots \to \boldsymbol{Z}$$

creates a cycle. The first statement follows from the fact that every path between I and Y must be of the form

$$I \to X \leftarrow Z \ldots Y.$$

Valid adjustment sets

Proposition (ECI, Proposition 6.41)

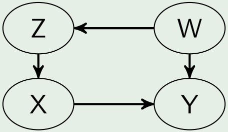
Assume that $Y \notin \boldsymbol{PA}_X$. The following sets are valid adjustment sets for (X, Y).

- $\boldsymbol{Z} = \boldsymbol{PA}_X$.
- $\boldsymbol{Z} \subseteq \boldsymbol{V} \setminus \{X, Y\}$ *such that* $\boldsymbol{Z}$ *contains no descendants of* X *and blocks all back-door paths between* X *and* Y.
- *no member of* $\boldsymbol{Z}$ *is a descendant of any* $W \in \boldsymbol{V} \setminus \{X\}$ *which lies on a directed path from* X *to* Y, *and* $\boldsymbol{Z}$ *blocks all nondirected paths between* X *and* Y.

We say that W is a *descendant* of X if there exists a directed path $X \to \ldots \to W$. We say that a path between X and Y is a *back-door path* if $X \leftarrow \ldots Y$. We say that $\boldsymbol{Z} \subseteq \boldsymbol{X} \setminus \{X, Y\}$ *blocks* a path between X and Y if the path is not *d*-connecting given $\boldsymbol{Z}$.

Parent set of X

Example

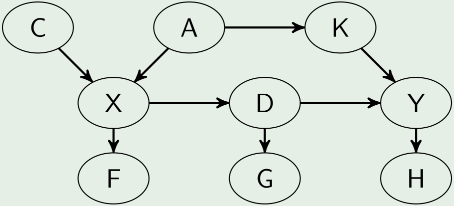


Z is the only parent of X, and $\{Z\}$ is therefore a valid adjustment set for (X, Y). This means that

$$p^{do(x)}(y) = \sum_z p(y|x, z)p(z)$$

Back-door example

Example



We see that $X \leftarrow A \rightarrow K \rightarrow Y$ is a back-door path so we need $A \in \boldsymbol{Z}$ or $K \in \boldsymbol{Z}$ to apply the back-door criterion. D, F, G, and H are not allowed in $\boldsymbol{Z}$ when using the back-door criterion, and C may be included or excluded.

Linear structural causal model

We say that an SCM is a *linear Gaussian SCM* if for each $j = 1, \ldots, d$,

$$X_j = f_j(\boldsymbol{PA}_j, N_j) = \sum_{X_i} a_{ji} X_i + N_j$$

where the sum is over $\boldsymbol{PA}_j$ and N_j is Gaussian.

Adjustment in linear structural causal model

Let $\boldsymbol{Z}$ be a valid adjustment set for (X, Y). In a zero-mean linear Gaussian SCM, we have

$$E(Y|X = x, \boldsymbol{Z} = \boldsymbol{z}) = ax + \boldsymbol{b}^t \boldsymbol{z}$$

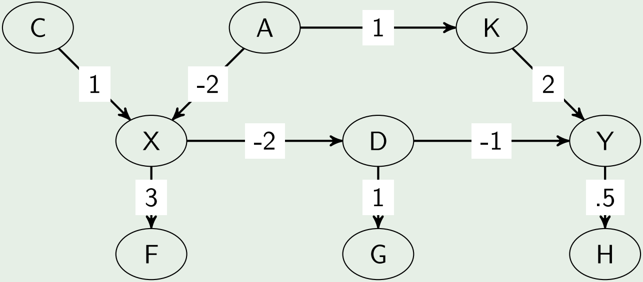
We also have (ECI Problem 6.63)

$$\frac{\partial}{\partial x} E^{do(x)}(Y) = a$$

for a constant a. If there is a single directed path from X to Y, then a equals the product of the corresponding coefficients.

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

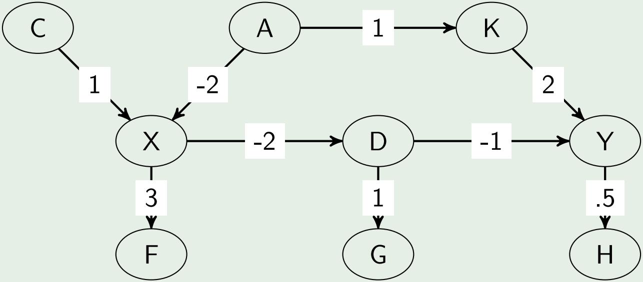
Example



- $\boldsymbol{Z} = \boldsymbol{PA}_X$.
- $\boldsymbol{Z} \subseteq \boldsymbol{V} \setminus \{X, Y\}$ such that $\boldsymbol{Z}$ contains no descendants of X and blocks all back-door paths between X and Y.
- no member of $\boldsymbol{Z}$ is a descendant of any $W \in \boldsymbol{V} \setminus \{X\}$ which lies on a directed path from X to Y, and $\boldsymbol{Z}$ blocks all nondirected paths between X and Y.

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

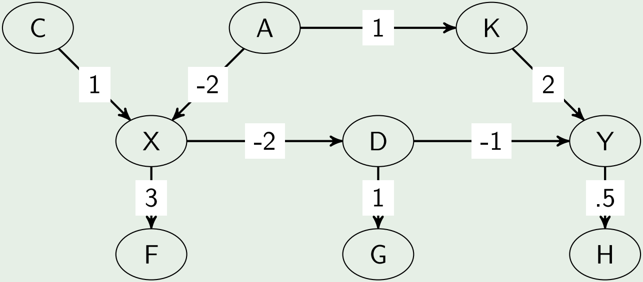
Example




```
lm(D~X)$coefficients
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

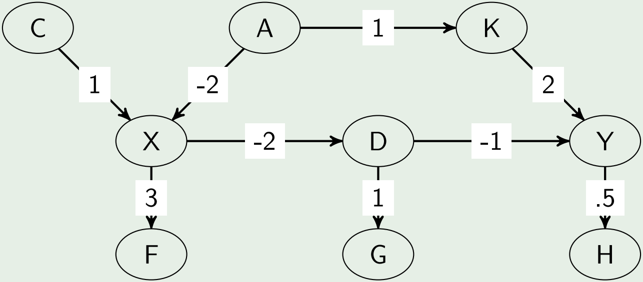
Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

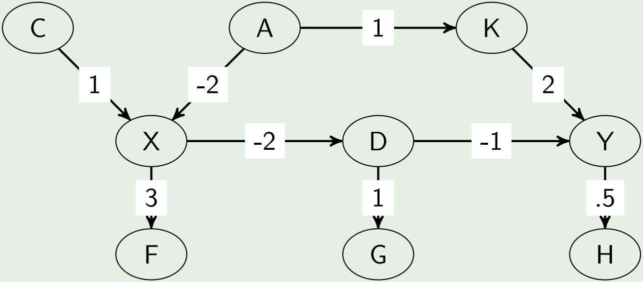
Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

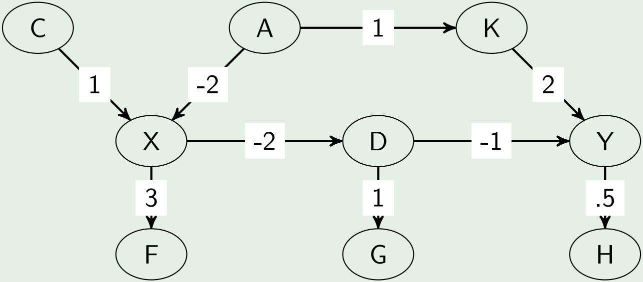
Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

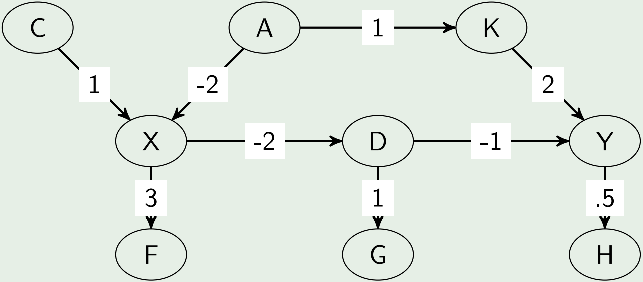
Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

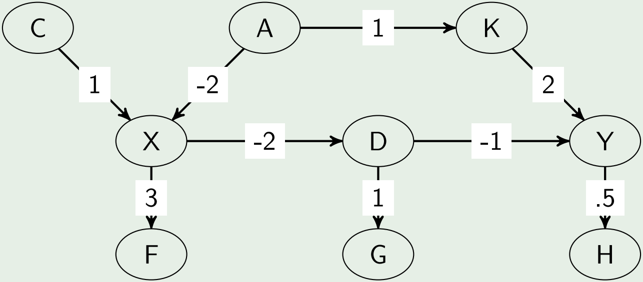
Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
# (Intercept)---------X
# 0.01428974 2.07038809
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

Example

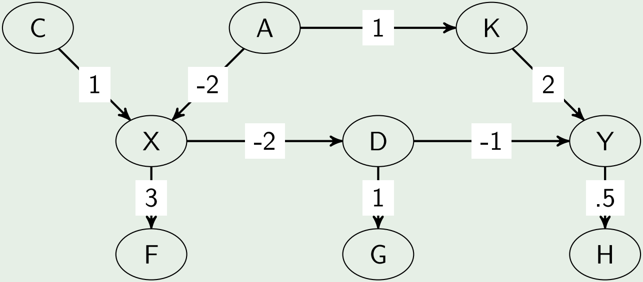



```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
# (Intercept)---------X
# 0.01428974 2.07038809
```

```
lm(Y~X+K+G)$coefficients
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

Example

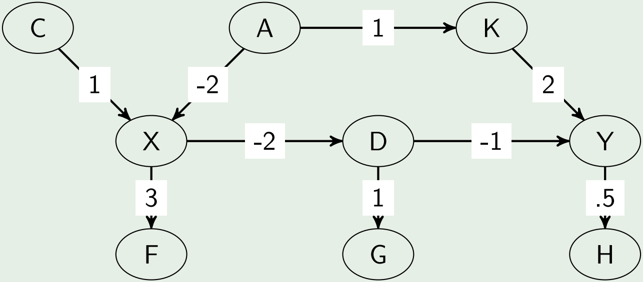



```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
# (Intercept)---------X
# 0.01428974 2.07038809
```

```
lm(Y~X+K+G)$coefficients
#(Intercept)----------X
#-0.0393745   1.1512565
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

Example

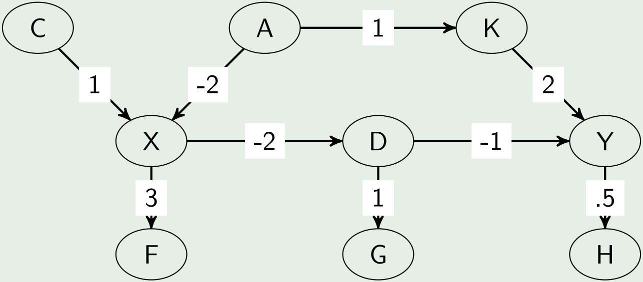



```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
# (Intercept)---------X
# 0.01428974 2.07038809
```

```
lm(Y~X+K+G)$coefficients
#(Intercept)----------X
#-0.0393745   1.1512565
lm(Y~X+F+C+K)$coefficients
```

Adjustment example, ECI Figure 6.5 and Code Snippet 6.43

Example




```
lm(D~X)$coefficients
# (Intercept)----------X
#-0.01990805 -2.01308637
lm(Y~X)$coefficients
# (Intercept)---------X
# 0.09724282 1.27941073
lm(Y~X+K)$coefficients
# (Intercept)---------X
# 0.01428974 2.07038809
```

```
lm(Y~X+K+G)$coefficients
#(Intercept)----------X
#-0.0393745   1.1512565
lm(Y~X+F+C+K)$coefficients
# (Intercept)---------X
# 0.01687018 1.90495456
```

Adjustment example

Imagine we have observational data on a cohort of children with variables

- neonatal vaccination status (X),
- disease status at three months (Y),
- treatment status,
- socioeconomic indicators,
- hospitalization history,
- maternal age at birth.

What seems like a reasonable adjustment set?

Do-calculus

The *do-calculus* is a set of rules that connect interventional and observational distributions. Say we have a causal graph $\mathcal{G}$ and disjoint node sets X, Y, Z, W.

1. Insertion/deletion of observations Let $\tilde{\mathcal{G}}$ be the graph obtained by removing all edges into X from $\mathcal{G}$ ($\to X$). If Y and Z are d-separated by $\{X, W\}$ in $\tilde{\mathcal{G}}$, then

$$p^{do(x)}(y|z, w) = p^{do(x)}(y|w)$$

Do-calculus

The *do-calculus* is a set of rules that connect interventional and observational distributions. Say we have a causal graph $\mathcal{G}$ and disjoint node sets X, Y, Z, W.

1. Insertion/deletion of observations Let $\tilde{\mathcal{G}}$ be the graph obtained by removing all edges into X from $\mathcal{G}$ ($\rightarrow X$). If Y and Z are d-separated by $\{X, W\}$ in $\tilde{\mathcal{G}}$, then

$$p^{do(x)}(y|z, w) = p^{do(x)}(y|w)$$

This follows from the global Markov property!

Do-calculus

2. Action/observation exchange Let $\tilde{\mathcal{G}}$ be the graph such that edges into X ($\to X$) and edges out of Z ($Z \to$) have been removed. If Y and Z are d-separated given $\{X, W\}$, then

$$p^{do(X=x,Z=z)}(y|w) = p^{do(X=x)}(y|z,w).$$

Do-calculus

3. Insertion/deletion of actions Let $\tilde{\mathcal{G}}$ denote the graph obtained from $\mathcal{G}$ by removing all edges into X. Let $Z(W)$ denote the subset of Z that are not ancestors of any node in W in $\tilde{G}$. Let $\bar{\mathcal{G}}$ denote the graph obtained from $\mathcal{G}$ by removing all edges into X and $Z(W)$. If Y and Z are d-separated given $\{X, W\}$ in $\bar{\mathcal{G}}$, then

$$p^{do(X=x,Z=z)}(y|w) = p^{do(X=x)}(y|w).$$

Do-calculus

Do-calculus is *complete* in the sense that all identifiable interventional distributions can be computed by repeatedly applying the three rules of do-calculus [Huang and Valtorta, 2006, Shpitser and Pearl, 2006].

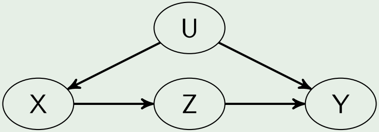
Front-door adjustment

Example

In the below example, there is no valid adjustment set for estimating the causal effect from X to Y when U is unobserved. However,

$$p^{do(x)}(y) = \sum_z p(z|x) \sum_{\tilde{x}} p(y|\tilde{x}, z) p(\tilde{x}).$$

One can show this using do-calculus (ECI Problem 6.65).



Markov equivalence of DAGs

We say that two DAGs, $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$, are *Markov equivalent* if they agree on all *d*-separations. That is, if for all $i, j \in V$ and $C \subseteq V \setminus \{i, j\}$ we have

$$i \perp\!\!\!\perp_{\mathcal{G}_1} j \mid C \Leftrightarrow i \perp\!\!\!\perp_{\mathcal{G}_2} j \mid C.$$

In the Week 1 exercises, we proved EIC Lemma 6.25.

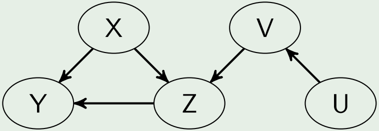
Theorem (Markov equivalence of DAGs, EIC Lemma 6.25)

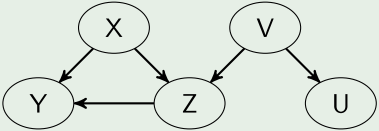
DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if they have the same skeleton and the same set of unshielded colliders.

The *skeleton* is the undirected graph obtained by replacing every directed edge in the DAG with an undirected edge. Three nodes (i, j, and k) are an *unshielded collider*, or *v-structure*, if $i \to k \leftarrow j$ and there is no edge between i and j.

Markov equivalence of DAGs

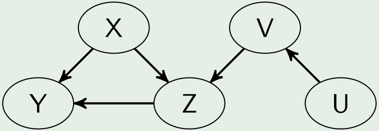
Example

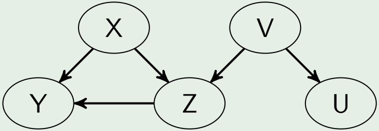




Markov equivalence of DAGs

Example

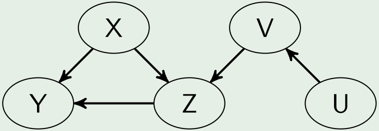


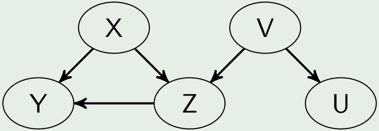


Let us argue that these constitute an equivalence class:

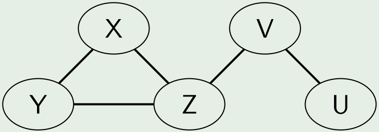
Markov equivalence of DAGs

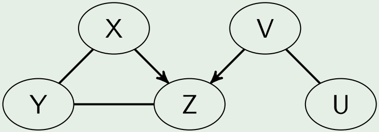
Example

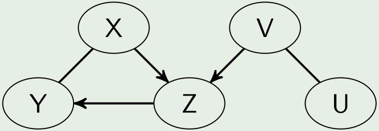


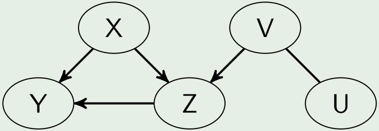


Let us argue that these constitute an equivalence class:









Markov equivalence of DAGs

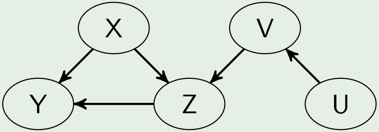
Markov equivalence defines an equivalence relation on the set of DAGs with node set V. Let $\mathcal{G} = (V, E)$ be a DAG. We say that

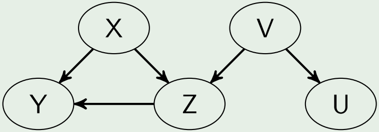
$$\{\tilde{\mathcal{G}} = (V, \tilde{E}) : \tilde{\mathcal{G}} \text{ and } \mathcal{G} \text{ are Markov equivalent}\}$$

is the *Markov equivalence class* of $\mathcal{G}$. It is useful to have a graphical representation of an entire Markov equivalence class. For this purpose, we define the *completed partially directed acyclic graph* (CPDAG) on nodes V by including $i \to j$ in the CPDAG if $i \to j$ in every graph in the Markov equivalence class and $i - j$ if $i \to j$ in some graph in the equivalence class and $i \leftarrow j$ in another.

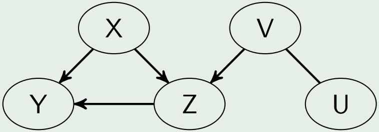
Markov equivalence of DAGs

Example





Equivalence class from before, and its CPDAG.



Markov equivalence of DAGs

Lemma (Meek [1995])

Let $\mathcal{C}$ be a CPDAG. If $i \to k - j$ in $\mathcal{C}$, then $i \to j$ in $\mathcal{C}$.

References I

Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 2006.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995.

Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 2006.