# Introduction to causal graphical models

Charles Assaad, Emilie Devijver, Eric Gaussier

charles.assaad@ens-lyon.fr

# Table of content

# Table of content

# (conditional) Independence

Conditional independence of random variables For a distribution $P$, $X$ and $Y$ are independent conditioned on $Z$, noted $X \perp\!\!\!\perp_P Y \mid Z$, *iff*:

$$P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)$$
$$\text{or } P(X \mid Y, Z) = P(X \mid Z) \text{ if } P(Y, Z) > 0$$

# (conditional) Independence

Conditional independence of random variables For a distribution $P$, $X$ and $Y$ are independent conditioned on $Z$, noted $X \perp\!\!\!\perp_P Y \mid Z$, *iff*:

$$P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)$$
$$\text{or } P(X \mid Y, Z) = P(X \mid Z) \text{ if } P(Y, Z) > 0$$

**Properties**

Symmetry: $X \perp\!\!\!\perp_P Y \mid Z \implies Y \perp\!\!\!\perp_P X \mid Z$

Decomposition: $X \perp\!\!\!\perp_P Y, W \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

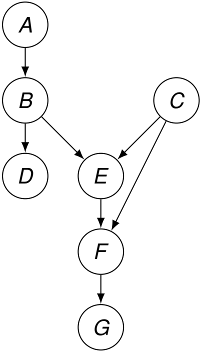
Weak union: $X \perp\!\!\!\perp_P Y, W \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z, W$

Contraction: $X \perp\!\!\!\perp_P Y \mid Z \;\&\; X \perp\!\!\!\perp_P W \mid Z, Y \implies X \perp\!\!\!\perp_P Y, W \mid Z$

Intersection: $X \perp\!\!\!\perp_P W \mid Z, Y \;\&\; X \perp\!\!\!\perp_P Y \mid Z, W \implies X \perp\!\!\!\perp_P Y, W \mid Z$

# Basic graph concepts

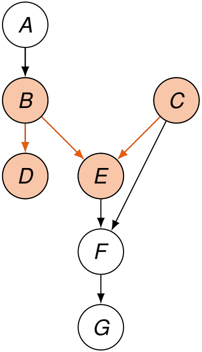
Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

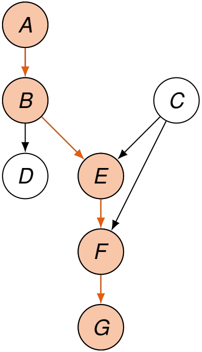
Path: $D \leftarrow B \rightarrow E \leftarrow C$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

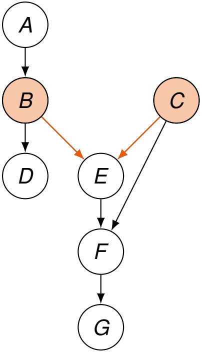
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

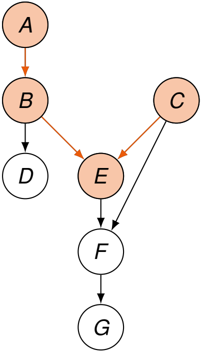
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

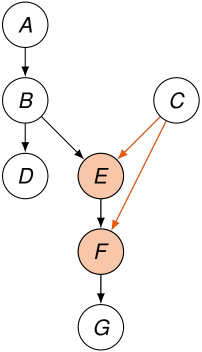
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

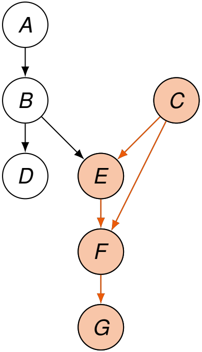
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Children: $Ch(C) = \{E, F\}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

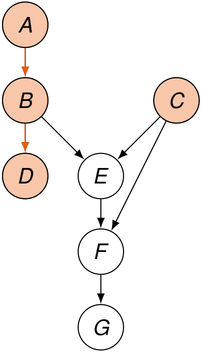
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Children: $Ch(C) = \{E, F\}$
Descendants: $De(C) = \{C, E, F, G\}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

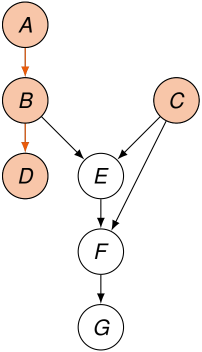
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Children: $Ch(C) = \{E, F\}$
Descendants: $De(C) = \{C, E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

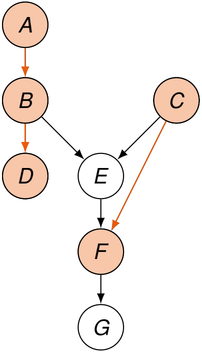
Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Children: $Ch(C) = \{E, F\}$
Descendants: $De(C) = \{C, E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$
Ancestral sets: a subset of nodes $\mathcal{S}$ is ancestral (or upward-closed) if $\forall S \in \mathcal{S}, An(S) \subseteq \mathcal{S}$



# Basic graph concepts

Consider the following graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

Path: $D \leftarrow B \rightarrow E \leftarrow C$
Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancestors: $An(E) = \{A, B, C, E\}$
Children: $Ch(C) = \{E, F\}$
Descendants: $De(C) = \{C, E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$
Ancestral sets: a subset of nodes $\mathcal{S}$ is ancestral (or upward-closed) if $\forall S \in \mathcal{S}, An(S) \subseteq \mathcal{S}$
Induced subgraph $\mathcal{G}[\mathcal{S}]$: $\mathcal{G}[\{B, C, D, F\}]$



# Table of content

# Bayesian networks and compatibility

Compatibility We say that a distribution $P(\mathcal{V})$ is compatible with (or Markov relative to) a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $P(\mathcal{V}) = \prod_{X \in \mathcal{V}} P(X \mid Pa(X))$.

# Bayesian networks and compatibility

Compatibility We say that a distribution $P(\mathcal{V})$ is compatible with (or Markov relative to) a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $P(\mathcal{V}) = \prod_{X \in \mathcal{V}} P(X \mid Pa(X))$.

Bayesian network A DAG (directed acyclic graph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a Bayesian network *iff* there exists a joint distribution $P(\mathcal{V})$ that is compatible with $\mathcal{G}$.

# Bayesian networks and compatibility

Compatibility We say that a distribution $P(\mathcal{V})$ is compatible with (or Markov relative to) a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if
$P(\mathcal{V}) = \prod_{X \in \mathcal{V}} P(X \mid Pa(X))$.

Bayesian network A DAG (directed acyclic graph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a Bayesian network *iff* there exists a joint distribution $P(\mathcal{V})$ that is compatible with $\mathcal{G}$.

Decomposing with respect to ancestral sets If $P$ is compatible with $\mathcal{G}$ and $\mathcal{S} \subseteq \mathcal{V}$ is an ancestral set, then
$P(\mathcal{S})$ is compatible with $\mathcal{G}[\mathcal{S}]$ (*i.e.*, $P(\mathcal{S}) = \prod_{S \in \mathcal{S}} P(S \mid Pa(S))$)
and $P(\mathcal{V} \backslash \mathcal{S} \mid \mathcal{S})$ is compatible with $\mathcal{G}[\mathcal{V} \backslash \mathcal{S}]$
(proof on board)

# Testing compatibility

Proposition (Ordered Markov condition) $P$ is compatible with $\mathcal{G}$ *iff* in any topological ordering $X_1, \cdots, X_n$ of $\mathcal{V}$, we have that

$$X_i \perp\!\!\!\perp X_1, \cdots, X_{i-1} \mid Pa(X_i) \qquad \text{for } i = 1, \cdots, n$$

(proof on board)

# Testing compatibility

Proposition (Ordered Markov condition) $P$ is compatible with $\mathcal{G}$ *iff* in any topological ordering $X_1, \cdots, X_n$ of $\mathcal{V}$, we have that

$$X_i \perp\!\!\!\perp X_1, \cdots, X_{i-1} \mid Pa(X_i) \qquad \text{for } i = 1, \cdots, n$$

(proof on board)

Proposition (Parental Markov condition (also known as local Markov condition)) $P$ is compatible with $\mathcal{G}$ *iff* $\forall X \in \mathcal{V}$ we have that

$$X \perp\!\!\!\perp Nd(X) \mid Pa(X)$$

(proof on board)

# Testing compatibility

Proposition (Ordered Markov condition) $P$ is compatible with $\mathcal{G}$ *iff* in any topological ordering $X_1, \cdots, X_n$ of $\mathcal{V}$, we have that

$$X_i \perp\!\!\!\perp X_1, \cdots, X_{i-1} \mid Pa(X_i) \qquad \text{for } i = 1, \cdots, n$$

(proof on board)

Proposition (Parental Markov condition (also known as local Markov condition)) $P$ is compatible with $\mathcal{G}$ *iff* $\forall X \in \mathcal{V}$ we have that

$$X \perp\!\!\!\perp Nd(X) \mid Pa(X)$$

(proof on board)

Proposition (Conditioning on common ancestors) For disjoint $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, if $An(\mathcal{X}) \cap An(\mathcal{Y}) \subseteq \mathcal{Z}$ and $An(\mathcal{Z}) \subseteq (\mathcal{Z})$, then

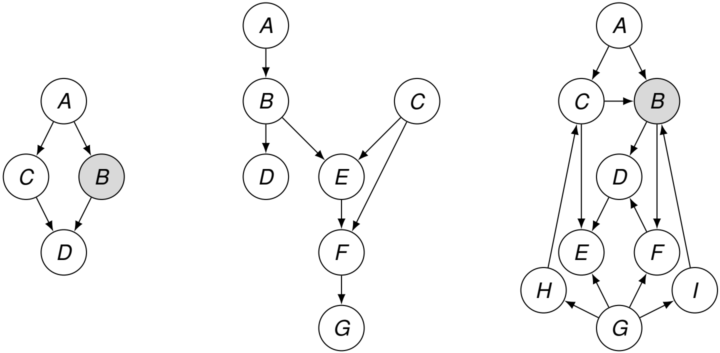
$$P(\mathcal{X}, \mathcal{Y} \mid \mathcal{Z}) = P(\mathcal{X} \mid \mathcal{Z})P(\mathcal{Y} \mid \mathcal{Z}) \text{ (i.e., } \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z})$$

in any distribution $P$ compatible with $\mathcal{G}$
(proof on board)

# Table of content

# Reading conditional independencies in graphs

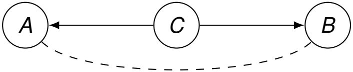
$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

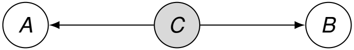


# Basic structures

# Basic structures

Fork



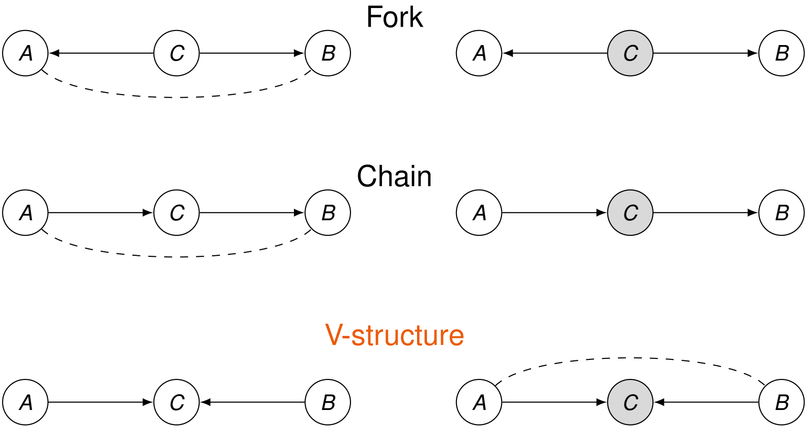



# Basic structures

Fork


Chain

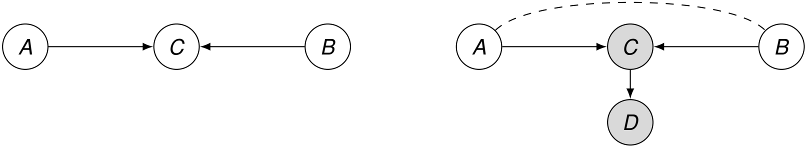
# Basic structures



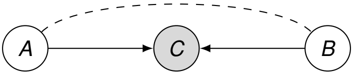
Fork


Chain

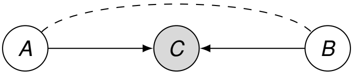

V-structure



# Artificial correlation in V-structures

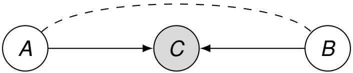


# Artificial correlation in V-structures




Example 1: $A = \begin{cases} Chicken \\ Other \backslash \{Rooster\} \end{cases} B = \begin{cases} Rooster \\ Other \backslash \{Chicken\} \end{cases} C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

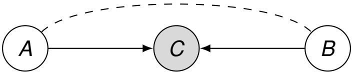
# Artificial correlation in V-structures




Example 1: $A = \begin{cases} Chicken \\ Other \backslash \{Rooster\} \end{cases} B = \begin{cases} Rooster \\ Other \backslash \{Chicken\} \end{cases} C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

If $C = Chick \implies A = Chicken$ and $B = Rooster$

# Artificial correlation in V-structures

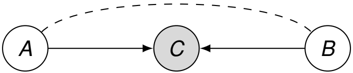



Example 1: $A = \begin{cases} Chicken \\ Other \backslash \{Rooster\} \end{cases} B = \begin{cases} Rooster \\ Other \backslash \{Chicken\} \end{cases} C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

If $C = Chick \implies A = Chicken$ and $B = Rooster$

If $C = Other \implies \begin{cases} A = Chicken \text{ and } B = Other \\ A = Other \text{ and } B = Rooster \\ A = Other \text{ and } B = Other \end{cases}$

# Artificial correlation in V-structures



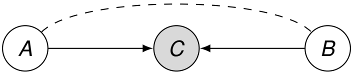

Example 1: $A = \begin{cases} Chicken \\ Other \backslash \{Rooster\} \end{cases}$ $B = \begin{cases} Rooster \\ Other \backslash \{Chicken\} \end{cases}$ $C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

If $C = Chick \implies A = Chicken$ and $B = Rooster$

If $C = Other \implies \begin{cases} A = Other \text{ and } B = Other \\ A = Other \text{ and } B = Rooster \\ A = Other \text{ and } B = Other \end{cases}$

Example 2: $A, B \sim U(-1, 1)$ $\qquad \xi_c \sim N(0, \frac{1}{2})$ $\qquad C = 2AB + \xi_c$

# Artificial correlation in V-structures

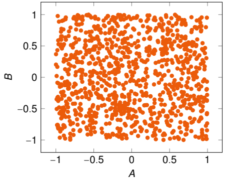



Example 1: $A = \begin{cases} Chicken \\ Other \backslash \{Rooster\} \end{cases} B = \begin{cases} Rooster \\ Other \backslash \{Chicken\} \end{cases} C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

If $C = Chick \implies A = Chicken$ and $B = Rooster$

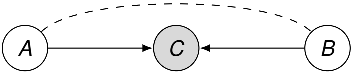
If $C = Other \implies \begin{cases} A = Chicken \text{ and } B = Other \\ A = Other \text{ and } B = Rooster \\ A = Other \text{ and } B = Other \end{cases}$

Example 2: $A, B \sim U(-1, 1)$ $\quad \xi_c \sim N(0, \frac{1}{2})$ $\quad C = 2AB + \xi_c$




$Corr(A; B) = 0.002$

# Artificial correlation in V-structures

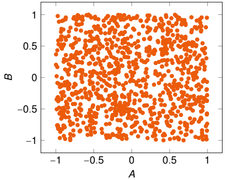


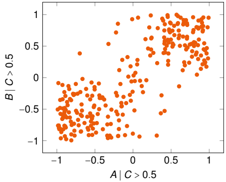

Example 1: $A = \begin{cases} Chicken \\ Other \setminus \{Rooster\} \end{cases}$ $B = \begin{cases} Rooster \\ Other \setminus \{Chicken\} \end{cases}$ $C = A \& B = \begin{cases} Chick \\ Other \end{cases}$

$$\text{If } C = Chick \implies A = Chicken \text{ and } B = Rooster$$

$$\text{If } C = Other \implies \begin{cases} A = Chicken \text{ and } B = Other \\ A = Other \text{ and } B = Rooster \\ A = Other \text{ and } B = Other \end{cases}$$

Example 2: $A, B \sim U(-1, 1)$ $\xi_C \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_C$




$Corr(A; B) = 0.002$ $Corr(A; B \mid C > 0.5) = 0.8$

# Blocked paths

Collider[1] A triple such that $X \to Z \leftarrow Y$. If the two parent vertices are not adjacent, the collider is a v-structure (also called unshielded collider or immorality)

[1] We also refer to $Z$ as the collider

# Blocked paths

Collider[1] A triple such that $X \to Z \leftarrow Y$. If the two parent vertices are not adjacent, the collider is a v-structure (also called unshielded collider or immorality)

Active and blocked paths A path is said to be blocked by a set of vertices $\mathcal{Z} \in \mathcal{V}$ if:

- it contains a chain $A \to B \to C$ or a fork $A \leftarrow B \to C$ and $B \in \mathcal{Z}$, or
- it contains a collider $A \to B \leftarrow C$ such that no descendant of $B$ is in $\mathcal{Z}$

[1] We also refer to $Z$ as the collider

# Blocked paths

Collider[1] A triple such that $X \rightarrow Z \leftarrow Y$. If the two parent vertices are not adjacent, the collider is a v-structure (also called unshielded collider or immorality)

Active and blocked paths A path is said to be blocked by a set of vertices $\mathcal{Z} \in \mathcal{V}$ if:

- it contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ and $B \in \mathcal{Z}$, or
- it contains a collider $A \rightarrow B \leftarrow C$ such that no descendant of $B$ is in $\mathcal{Z}$

A path that is not blocked is active.

[1] We also refer to $Z$ as the collider

# Blocked paths

Collider[1] A triple such that $X \rightarrow Z \leftarrow Y$. If the two parent vertices are not adjacent, the collider is a v-structure (also called unshielded collider or immorality)

Active and blocked paths A path is said to be blocked by a set of vertices $\mathcal{Z} \in \mathcal{V}$ if:

- it contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ and $B \in \mathcal{Z}$, or
- it contains a collider $A \rightarrow B \leftarrow C$ such that no descendant of $B$ is in $\mathcal{Z}$

A path that is not blocked is active. A path is active if every triple along the path is active, and blocked if a single triple is blocked

[1] We also refer to $Z$ as the collider

# d-separation

d-separation (also known as the global Markov condition) Given disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, we say that $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ if every path between a node in $\mathcal{X}$ and a node in $\mathcal{Y}$ is blocked by $\mathcal{Z}$ and we write $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$.

# d-separation

d-separation (also known as the global Markov condition) Given disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, we say that $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ if every path between a node in $\mathcal{X}$ and a node in $\mathcal{Y}$ is blocked by $\mathcal{Z}$ and we write $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$.

# d-separation

d-separation (also known as the global Markov condition) Given disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, we say that $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ if every path between a node in $\mathcal{X}$ and a node in $\mathcal{Y}$ is blocked by $\mathcal{Z}$ and we write $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$.

If one of the above path is not blocked, we say that $\mathcal{X}$ and $\mathcal{Y}$ are d-connected given $\mathcal{Z}$

# d-separation and conditional independence

d-separation characterizes the conditional independencies of distributions compatible with a given DAG

Theorem (probabilistic implications of d-separation)

(i) *Soundness* $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$ in every distribution $P$ compatible with $\mathcal{G}$

(ii) *Completeness* If $\mathcal{X} \not\perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$, then there exists a distribution $P$ compatible with $\mathcal{G}$ such that $\mathcal{X} \not\perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$

Proof in (Pearl, 1988)

# d-separation and conditional independence

d-separation characterizes the conditional independencies of distributions compatible with a given DAG

Theorem (probabilistic implications of d-separation)

(i) *Soundness* $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ in every distribution $P$ compatible with $\mathcal{G}$

(ii) *Completeness* If $\mathcal{X} \not\perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$, then there exists a distribution $P$ compatible with $\mathcal{G}$ such that $\mathcal{X} \not\perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$

Proof in (Pearl, 1988)

# d-separation and conditional independence

d-separation characterizes the conditional independencies of distributions compatible with a given DAG

Theorem (probabilistic implications of d-separation)

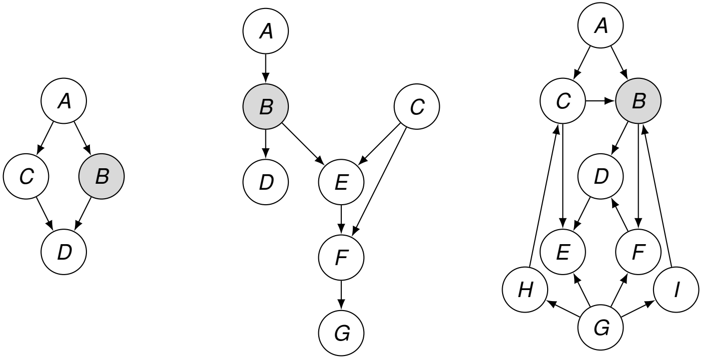
(i) *Soundness* $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$ in every distribution $P$ compatible with $\mathcal{G}$

(ii) *Completeness* If $\mathcal{X} \not\perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$, then there exists a distribution $P$ compatible with $\mathcal{G}$ such that $\mathcal{X} \not\perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$

Proof in (Pearl, 1988)

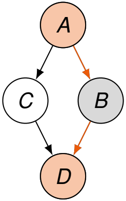
# Reading conditional independencies in graphs using d-separation

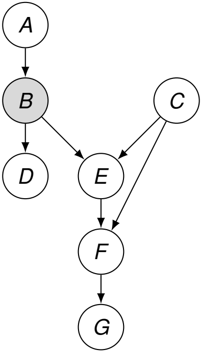
$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$

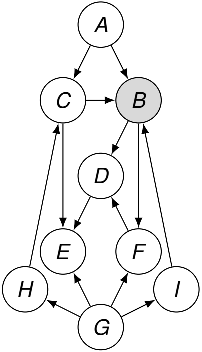


# Reading conditional independencies in graphs using d-separation

$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$

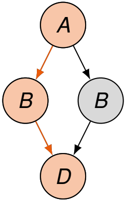


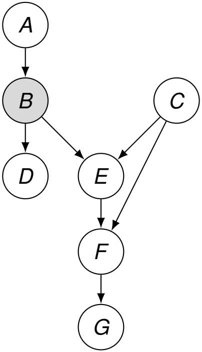



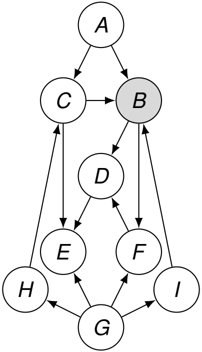



# Reading conditional independencies in graphs using d-separation

$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$

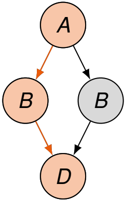


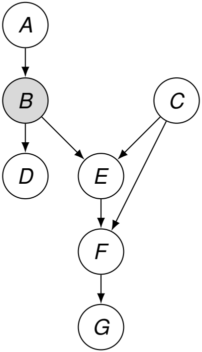
# Reading conditional independencies in graphs using d-separation

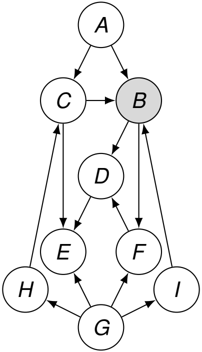
$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$




$< A, B, D >$ is not blocked

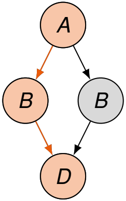
$$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$



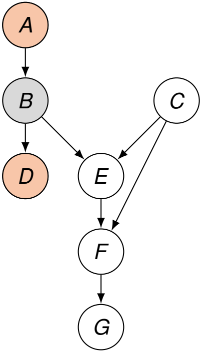
# Reading conditional independencies in graphs using d-separation

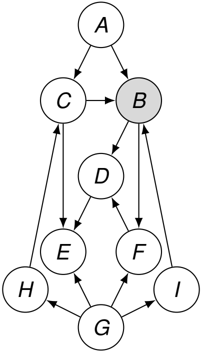
$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$




$< A, B, D >$ is not blocked

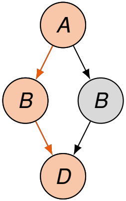
$$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$



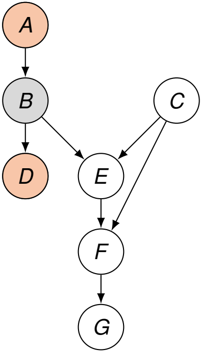
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



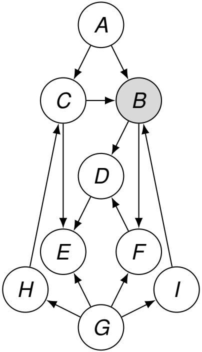

$< A, B, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



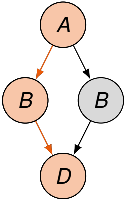

All paths are blocked

$\implies A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$



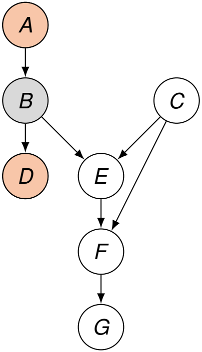
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



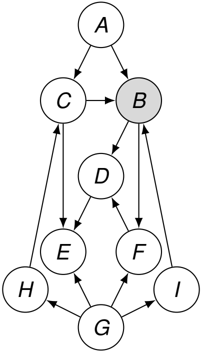

$< A, B, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



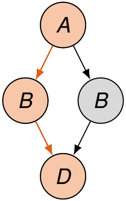

All paths are blocked

$\Longrightarrow A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$



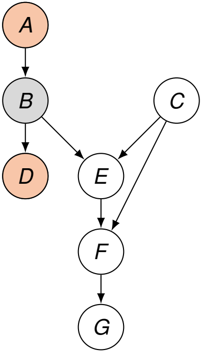
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

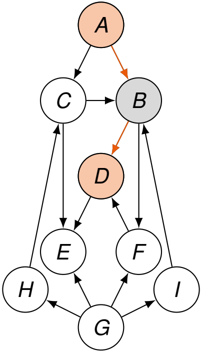



$< A, B, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



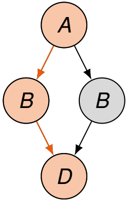

All paths are blocked




$\implies A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$

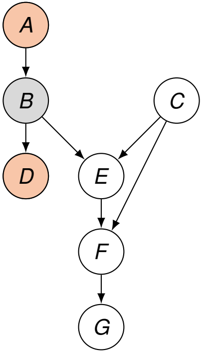
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

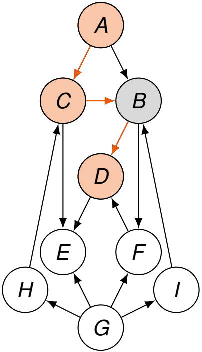



$< A, B, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



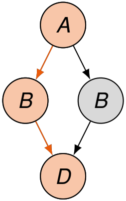

All paths are blocked




$\implies A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$

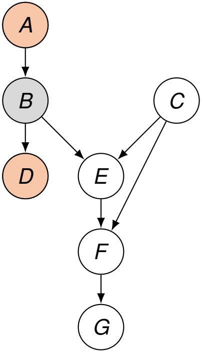
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

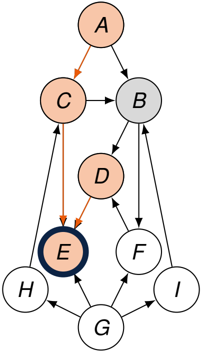



$< A, B, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



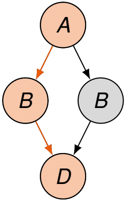

All paths are blocked




$\implies A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$

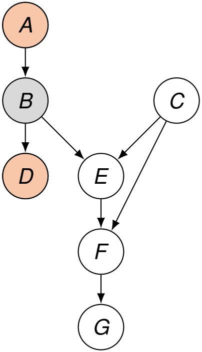
# Reading conditional independencies in graphs using d-separation

$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$

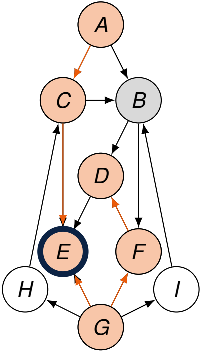



$< A, B, D >$ is not blocked

$$\Longrightarrow \quad A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$



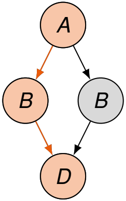

All paths are blocked




$$\Longrightarrow \quad A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$$

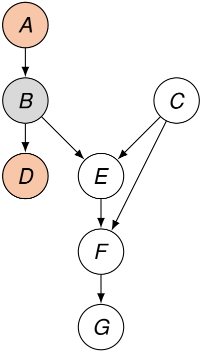
# Reading conditional independencies in graphs using d-separation

$$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$

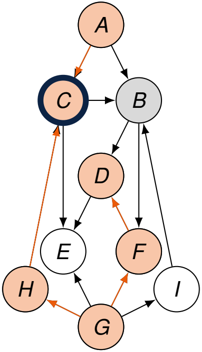



$< A, B, D >$ is not blocked

$$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$$



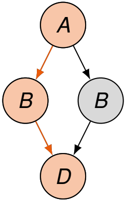

All paths are blocked




$$\Longrightarrow A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$$

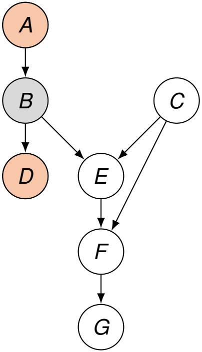
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

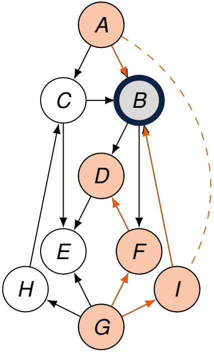



$< A, B, D >$ is not blocked

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



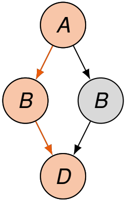

All paths are blocked




$\Longrightarrow A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$

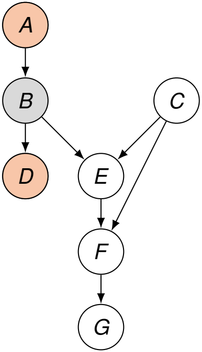
# Reading conditional independencies in graphs using d-separation

$A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$



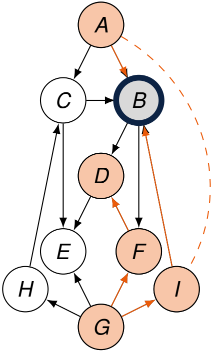

$< A, B, D >$ is not blocked

$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$




All paths are blocked

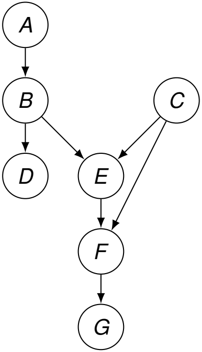
$\implies A \perp\!\!\!\perp_{\mathcal{P}} D \mid B$




$< A, I, G, F, D >$ is not blocked

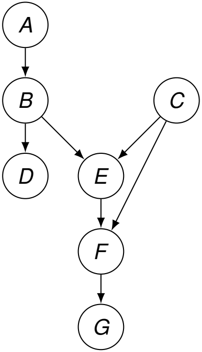
$\implies A \overset{?}{\perp\!\!\!\perp}_{\mathcal{P}} D \mid B$

# More examples



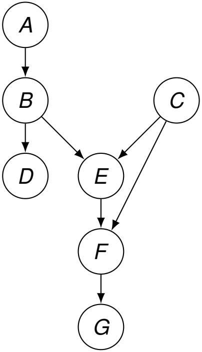

- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid C, E$?
- $B \perp\!\!\!\perp_P E \mid F$?

# More examples



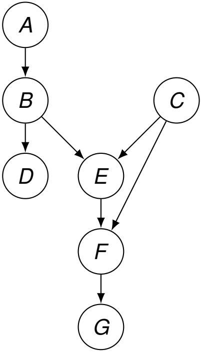

- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid C, E$?
- $B \perp\!\!\!\perp_P E \mid F$?

# More examples




- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid C, E$?
- $B \perp\!\!\!\perp_P E \mid F$?

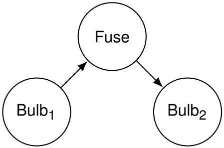
# More examples

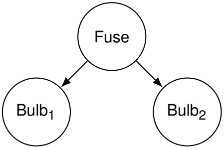



- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid C, E$?
- $B \perp\!\!\!\perp_P E \mid F$?

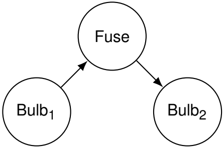
# Table of content

# Bayesian networks vs causal graph



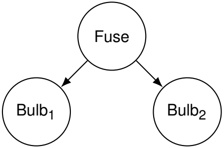



# Bayesian networks vs causal graph




$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

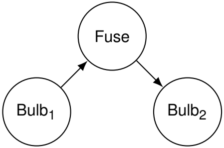
Bayesian network




$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

Bayesian network

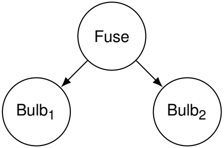
# Bayesian networks vs causal graph




$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

Bayesian network

Not a causal graph

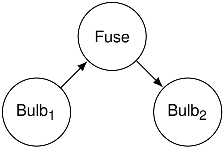



$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

Bayesian network

Causal graph

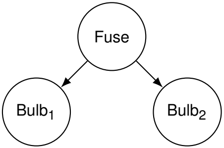
# Bayesian networks vs causal graph




$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

Bayesian network

Not a causal graph




$\text{Bulb}_1 \perp\!\!\!\perp_P \text{Bulb}_2 \mid \text{Fuse}$

Bayesian network

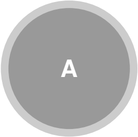
Causal graph

Oracle for conditional independence

Oracle for intervention

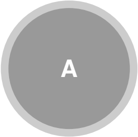
# Conditioning vs Intervening (1/2)

Population

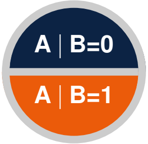


# Conditioning vs Intervening (1/2)

Population




Sub-populations

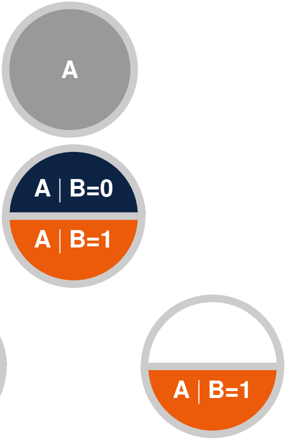


# Conditioning vs Intervening (1/2)

Population

Sub-populations

Conditioning



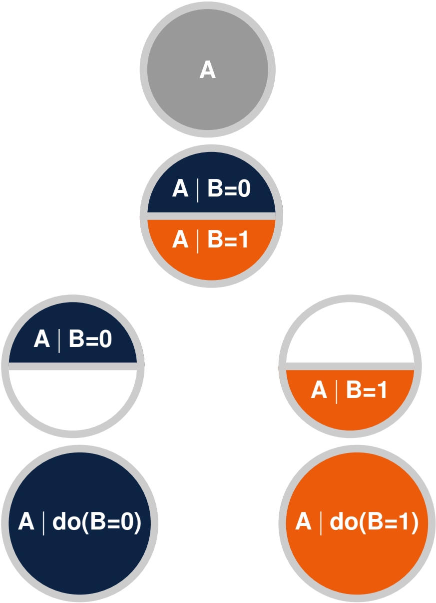
# Conditioning vs Intervening (1/2)

Population

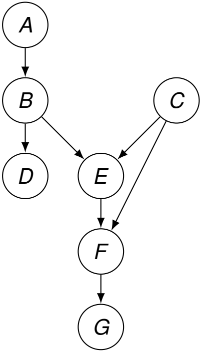
Sub-populations

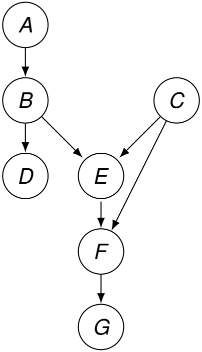
Conditioning

Intervening

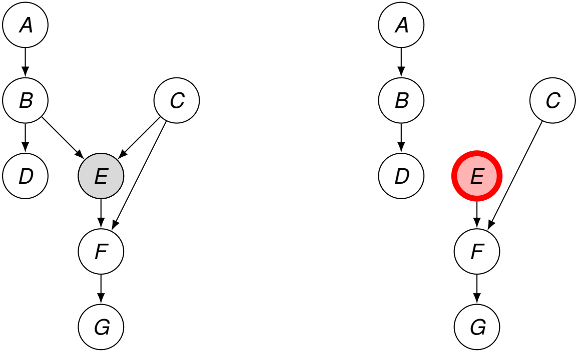


# Conditioning vs Intervening (2/2)



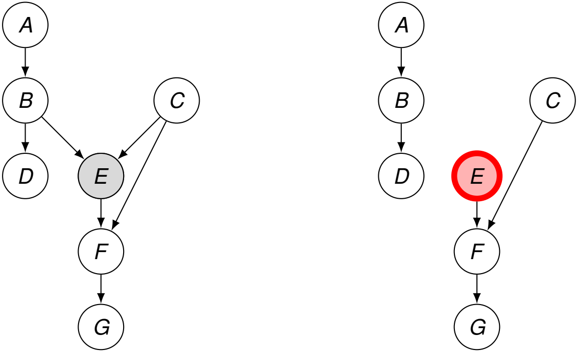
# Conditioning vs Intervening (2/2)




Note that there are two types of interventions:

- Structural (or hard) intervention
- Parametric (or soft) intervention

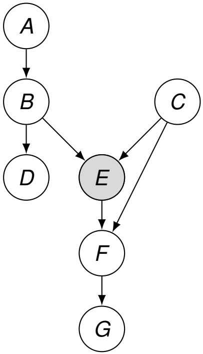
# Conditioning vs Intervening (2/2)

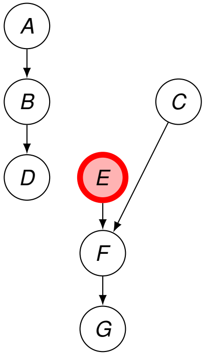



Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

# Conditioning vs Intervening (2/2)



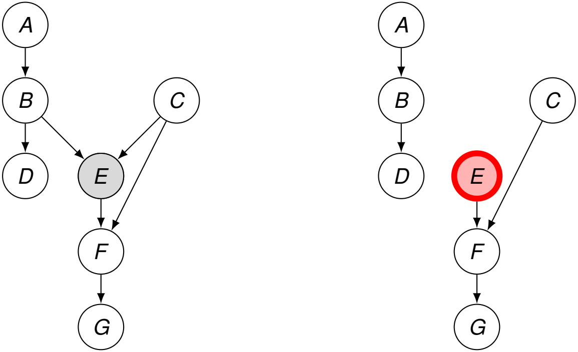




Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

The operator $do()$ is a way to denote (hard) interventions

# Conditioning vs Intervening (2/2)

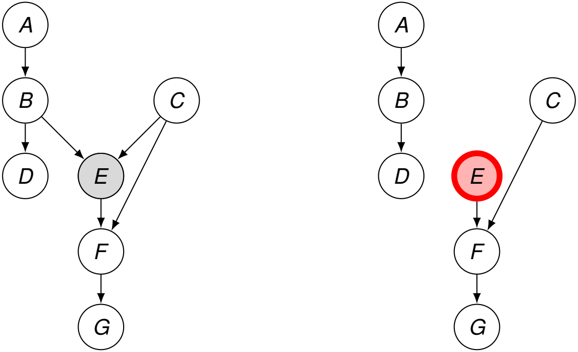



Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

The operator $do()$ is a way to denote (hard) interventions
For example $P(a, b, c, d, f, g \mid do(e))$

# Conditioning vs Intervening (2/2)




Note that there are two types of interventions:

- Structural (or hard) intervention (we will focus on this)
- Parametric (or soft) intervention

The operator $do()$ is a way to denote (hard) interventions
For example $P(a, b, c, d, f, g \mid do(e))$ or $P_{E=e}(a, b, c, d, f, g)$

# From association to causation (1/2)

Reminder: parental Markov condition

$$\forall X \in \mathcal{V}, \qquad X \perp\!\!\!\perp Nd(X) \mid Pa(X)$$

# From association to causation (1/2)

Reminder: parental Markov condition

$$\forall X \in \mathcal{V}, \qquad X \perp\!\!\!\perp Nd(X) \mid Pa(X)$$

Causal Markov condition

$$\forall X \in \mathcal{V}, \qquad X \perp\!\!\!\perp NotEffects(X) \mid DirectCauses(X)$$

# From association to causation (2/2)

Reminder: Bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

# From association to causation (2/2)

Reminder: Bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

$$\Pr(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_i \Pr(\mathbf{V}_i = \mathbf{v}_i \mid Pa(\mathbf{V}_i))$$

# From association to causation (2/2)

Reminder: Bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

$$\Pr(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_i \Pr(\mathbf{V}_i = \mathbf{v}_i \mid Pa(\mathbf{V}_i))$$

**Truncated factorization** (also known as the manipulation theorem) If we intervene on a subset $S \subset \mathbf{V}$, then

$$\Pr_{\{S=s\}}(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_{i \notin S} \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

if $\mathbf{v}_1, \cdots, \mathbf{v}_d$ are values consistant with the intervention, else,

$$\Pr_{\{S=s\}}(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = 0$$

# Causal Bayesian networks

Causal Bayesian network Let $P(\mathcal{V})$ be a probability distribution and let $P(\mathcal{V} \mid do(s))$ denote the distribution resulting from the intervention that sets a subset $\mathcal{S}$ of variables to constants $s$. Let $\mathcal{P}_*$ denote the set of all interventional distributions $P(\mathcal{V} \mid do(s))$. A DAG $\mathcal{G}$ is said to be a causal Bayesian network compatible with $\mathcal{P}_*$ *iff* $\mathcal{G}$ and $\mathcal{P}_*$ satisfy the truncated factorization.

# Applications

Causal discovery

- It is possible to infer a causal graph from observational data?
- How?

# Applications

Causal discovery

- It is possible to infer a causal graph from observational data?
- How?

Causal reasoning:

- Given a causal graph, is it possible to estimate the effect of an intervention from observational data?
- How?

# Applications

Causal discovery

- It is possible to infer a causal graph from observational data?
- How?

Causal reasoning:

- Given a causal graph, is it possible to estimate the effect of an intervention from observational data?
- How?

Identifiability: The causal effect of an intervention $do(x)$ on a set of variables $Y$ such that $Y \cap X = \varnothing$ is said to be identifiable from $P$ in $\mathcal{G}$ if $P(Y \mid do(x))$ is uniquely computable from $P(\mathcal{V})$.

# Table of content

# Linear structural causal models

Linear structural causal model It consists on a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \beta_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of $y$, $\xi_y$ represent errors due to ommitted factors and $\beta_{xy}$ which are known as a structural coefficient represents the strength of the causal relation.

# Linear structural causal models

Linear structural causal model It consists on a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \beta_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of $y$, $\xi_y$ represent errors due to ommitted factors and $\beta_{xy}$ which are known as a structural coefficient represents the strength of the causal relation.

**Differences between regression and causal coefficients**: Suppose the following linear structural causal model.

$$M : \begin{cases} a := \xi_a \\ b := \beta_{ab} a + \xi_b \\ c := \beta_{ac} a + \xi_c \end{cases}$$

# Linear structural causal models

Linear structural causal model It consists on a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \beta_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of $y$, $\xi_y$ represent errors due to ommitted factors and $\beta_{xy}$ which are known as a structural coefficient represents the strength of the causal relation.

**Differences between regression and causal coefficients**: Suppose the following linear structural causal model. What is the regression coefficient when we regress $c$ on $b$?

$$M : \begin{cases} a := \xi_a \\ b := \beta_{ab} a + \xi_b \\ c := \beta_{ac} a + \xi_c \end{cases}$$

# Linear structural causal models

Linear structural causal model It consists on a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \beta_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of $y$, $\xi_y$ represent errors due to ommitted factors and $\beta_{xy}$ which are known as a structural coefficient represents the strength of the causal relation.

**Differences between regression and causal coefficients**:
Suppose the following linear structural causal model. What is the regression coefficient when we regress $c$ on $b$?

$$M : \begin{cases} a := \xi_a \\ b := \beta_{ab} a + \xi_b \\ c := \beta_{ac} a + \xi_c \end{cases}$$

$$\begin{aligned} b &= \hat{\beta}_{cb} c + \xi_b \\ &= \hat{\beta}_{cb}(\beta_{ac} a + \xi_c) + \xi_b \\ \hat{\beta}_{cb} &= \frac{b - \xi_b}{\beta_{ac} a + \xi_c} \end{aligned}$$

# Linear structural causal models

Linear structural causal model It consists on a set of structural equations of the form:

$$y := \sum_{x \in Pa(y)} \beta_{xy} x + \xi_y$$

where $Pa(y)$ are direct causes of $y$, $\xi_y$ represent errors due to ommitted factors and $\beta_{xy}$ which are known as a structural coefficient represents the strength of the causal relation.

**Differences between regression and causal coefficients**: Suppose the following linear structural causal model. What is the regression coefficient when we regress $c$ on $b$?

$$M : \begin{cases} a := \xi_a \\ b := \beta_{ab} a + \xi_b \\ c := \beta_{ac} a + \xi_c \end{cases}$$

$$\begin{aligned} b &= \hat{\beta}_{cb} c + \xi_b \\ &= \hat{\beta}_{cb}(\beta_{ac} a + \xi_c) + \xi_b \\ \hat{\beta}_{cb} &= \frac{b - \xi_b}{\beta_{ac} a + \xi_c} \neq 0 \end{aligned}$$

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$
3. $\mathcal{F}$ is a set of functions s.t. $f_i$ $(1 \leq i \leq n)$ specifies $X_i$:
   $X_i = f(\mathcal{S}_i)$ with $\mathcal{S}_i \subseteq \mathcal{U} \cup \mathcal{V}$

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$
3. $\mathcal{F}$ is a set of functions s.t. $f_i$ $(1 \leq i \leq n)$ specifies $X_i$: $X_i = f(\mathcal{S}_i)$ with $\mathcal{S}_i \subseteq \mathcal{U} \cup \mathcal{V}$

Probabilistic causal models A pair $\langle M, P \rangle$ with

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, \dots, X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$
3. $\mathcal{F}$ is a set of functions s.t. $f_i$ $(1 \leq i \leq n)$ specifies $X_i$: $X_i = f(\mathcal{S}_i)$ with $\mathcal{S}_i \subseteq \mathcal{U} \cup \mathcal{V}$

Probabilistic causal models A pair $\langle M, P \rangle$ with

1. $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ is a structural causal model

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$
3. $\mathcal{F}$ is a set of functions s.t. $f_i$ $(1 \leq i \leq n)$ specifies $X_i$: $X_i = f(\mathcal{S}_i)$ with $\mathcal{S}_i \subseteq \mathcal{U} \cup \mathcal{V}$

Probabilistic causal models A pair $\langle M, P \rangle$ with

1. $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ is a structural causal model
2. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

# Non-parametric structural causal models

Structural causal model A triple $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables (also known as exogenous variables or error terms) that are determined by factors outside the model
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables (also known as endogenous variables) that are determined by variables in the model - that is, variables in $\mathcal{U} \cup \mathcal{V}$
3. $\mathcal{F}$ is a set of functions s.t. $f_i$ ($1 \leq i \leq n$) specifies $X_i$: $X_i = f(\mathcal{S}_i)$ with $\mathcal{S}_i \subseteq \mathcal{U} \cup \mathcal{V}$

Probabilistic causal models A pair $\langle M, P \rangle$ with

1. $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$ is a structural causal model
2. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

$P(\mathcal{U})$ and $\mathcal{F}$ induce a joint distribution $P(\mathcal{V})$ over $\mathcal{V}$.

# Induced graph

Induced graph *The graph $\mathcal{G}$ induced by a structural causal model M has vertices $\mathcal{V}$ and an edge $X_i \rightarrow X_j$ whenever $f_j$ depends on $X_i$. In addition, $\mathcal{G}$ contains a bidirected edge, denoted $X_i \dashleftarrow\dashrightarrow X_j$ , whenever $f_i$ and $f_j$ depend on a common subset of $\mathcal{U}$.*

# Induced graph

Induced graph *The graph $\mathcal{G}$ induced by a structural causal model $M$ has vertices $\mathcal{V}$ and an edge $X_i \rightarrow X_j$ whenever $f_j$ depends on $X_i$. In addition, $\mathcal{G}$ contains a bidirected edge, denoted $X_i \leftarrow\!\!\dashrightarrow X_j$ , whenever $f_i$ and $f_j$ depend on a common subset of $\mathcal{U}$.*

Markovian causal model *A causal model $M$ is Markovian if the graph induced by $M$ contains no bidirected edges (the graph is a DAG)*

# Induced graph

Induced graph *The graph $\mathcal{G}$ induced by a structural causal model M has vertices $\mathcal{V}$ and an edge $X_i \rightarrow X_j$ whenever $f_j$ depends on $X_i$. In addition, $\mathcal{G}$ contains a bidirected edge, denoted $X_i \leftarrow\!\!-\!\!-\!\!\rightarrow X_j$ , whenever $f_i$ and $f_j$ depend on a common subset of $\mathcal{U}$.*

Markovian causal model *A causal model M is Markovian if the graph induced by M contains no bidirected edges (the graph is a DAG)*

Semi-Markovian causal model *A causal model M is Semi-Markovian if the graph induced by M contains bidirected edges (the graph is a ADMG)*

# Induced distribution in Markovian models

$P(\mathcal{V})$ does not depend on $\mathcal{U}$ in Markovian causal models

$$P(\mathcal{V} \cup \mathcal{U}) = \prod_{i=1}^{n} P(x_i \,|\, Pa(x_i), u_i) P(u_i)$$

$$\sum_{u} P(\mathcal{V} \cup \mathcal{U}) = \sum_{u} \prod_{i=1}^{n} P(x_i \,|\, x_1, ..., x_{i-1}, u_i) P(u_i)$$

$$P(\mathcal{V}) = \sum_{u} \prod_{i=1}^{n} \frac{P(x_i, u_i \,|\, x_1, ..., x_{i-1})}{P(u_i)} P(u_i)$$

$$= \prod_{i=1}^{n} P(x_i \,|\, x_1, ..., x_{i-1})$$

# Induced distribution in Markovian models

$P(\mathcal{V})$ does not depend on $\mathcal{U}$ in Markovian causal models

$$P(\mathcal{V} \cup \mathcal{U}) = \prod_{i=1}^{n} P(x_i \,|\, Pa(x_i), u_i) P(u_i)$$

$$\sum_{u} P(\mathcal{V} \cup \mathcal{U}) = \sum_{u} \prod_{i=1}^{n} P(x_i \,|\, x_1, ..., x_{i-1}, u_i) P(u_i)$$

$$P(\mathcal{V}) = \sum_{u} \prod_{i=1}^{n} \frac{P(x_i, u_i \,|\, x_1, ..., x_{i-1})}{P(u_i)} P(u_i)$$

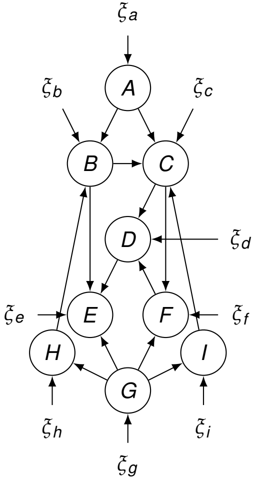
$$= \prod_{i=1}^{n} P(x_i \,|\, x_1, ..., x_{i-1})$$

# Example of a Markovian model

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

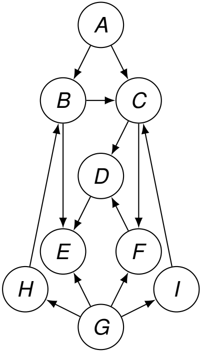
# Example of a Markovian model

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$



# Example of a Markovian model

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$



# Example of a semi-Markovian model

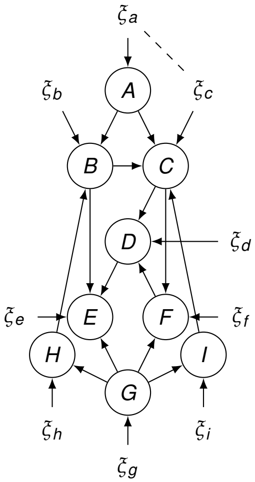
$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

$\xi_a \not\perp\!\!\!\perp \xi_c$

# Example of a semi-Markovian model

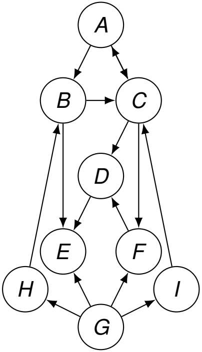
$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

$\xi_a \not\perp\!\!\!\perp \xi_c$



# Example of a semi-Markovian model

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$




$\xi_a \not\perp\!\!\!\perp \xi_c$

# SCMs and interventions

SCM

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

# SCMs and interventions

SCM

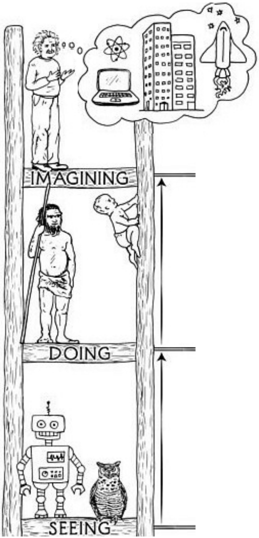
$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

Interventional SCM

$$M_c : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := c \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

# Table of content

# Conclusion




**SCMs**

---

**Causal graphs**

---

**Bayesian networks**

# References


1. *An Introduction to Causal Graphical Models*, S. Gordon (slides available at https://simons.berkeley.edu/sites/default/files/docs/18989/cau22-bcspencergordon.pdf)

2. *An Introduction to Causal Graphical Models*, V. Kumar, A. Capiln, C. Park, S. Gordon, L. Schulman (handout available at https://tinyurl.com/causalitybootcamp)

3. *Causality*, J. Pearl. Cambridge University Press, 2nd edition, 2009

4. *Probabilistic Reasoning in Intelligent Systems*, J. Pearl. The Morgan Kaufmann Series in Representation and Reasoning, 1988