Learning, Probabilities and Causality, Part 2

Charles K. Assaad, Emilie Devijver, Eric Gaussier

emilie.devijver@univ-grenoble-alpes.fr

Course overview

Session 1 Friday 25 November (E. Devijver)
Introduction to causal graphical models

Session 2 Friday 2 December (E. Gaussier - C. Assaad)
Structural equation models, structural causal models
Causal discovery: constraint-based methods

Session 3 Friday 9 December (E. Devijver)
Lab on introduction to graphs and PC algorithm

Session 4 Friday 16 December (C. Assaad, E. Gaussier)
Causal discovery: noise-based, score-based and other methods

Session 5 Friday 6 January (C. Assaad, E. Devijver)
Backdoor and frontdoor criteria
Do-calculus

Session 6 Friday 13 January (C. Assaad, E. Gaussier)
History
Lab on Simpson's paradox

Table of content

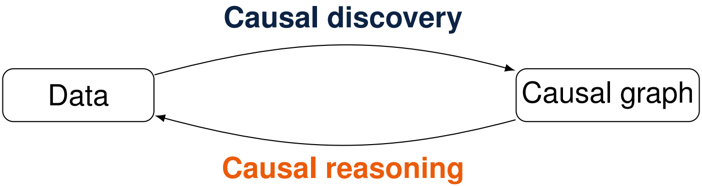
Causality

Does Obesity Shorten Life? Or is it the Soda? (Pearl, 2018)

Many applications in machine learning, medecine (science in general), root cause analysis, ...

Causality

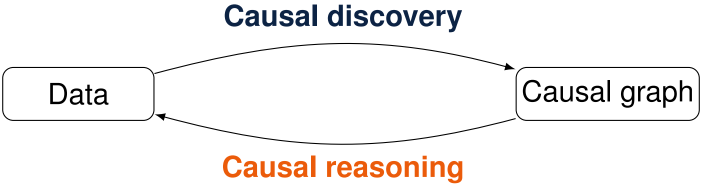
Does Obesity Shorten Life? Or is it the Soda? (Pearl, 2018)



Many applications in machine learning, medecine (science in general), root cause analysis, ...

Causality

Does Obesity Shorten Life? Or is it the Soda? (Pearl, 2018)



Many applications in machine learning, medecine (science in general), root cause analysis, ...

Causes and effects

The same causes produce the same effects ..., do they?

- Smoking causes lung cancer
- The sound of your alarm makes you wake up
- Cause: I flipped the light switch - Effect: the light came on

Probabilities are used to capture uncertainty/indeterminacy

→ Probabilistic Causal Models

Causes and effects

The same causes produce the same effects ..., do they?

- Smoking causes lung cancer
- The sound of your alarm makes you wake up
- Cause: I flipped the light switch - Effect: the light came on

Probabilities are used to capture uncertainty/indeterminacy

→ Probabilistic Causal Models

Causes and effects

The same causes produce the same effects ..., do they?

- Smoking causes lung cancer
- The sound of your alarm makes you wake up
- Cause: I flipped the light switch - Effect: the light came on

Probabilities are used to capture uncertainty/indeterminacy

→ Probabilistic Causal Models

Causes and effects

The same causes produce the same effects ..., do they?

- Smoking causes lung cancer
- The sound of your alarm makes you wake up
- Cause: I flipped the light switch - Effect: the light came on

Probabilities are used to capture uncertainty/indeterminacy

→ Probabilistic Causal Models

Causes and effects

The same causes produce the same effects ..., do they?

- Smoking causes lung cancer
- The sound of your alarm makes you wake up
- Cause: I flipped the light switch - Effect: the light came on

Probabilities are used to capture uncertainty/indeterminacy

→ Probabilistic Causal Models

(conditional) Independence

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

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

Illustration

- $Z \sim Bi(9, 0.5)$, $X \mid Z = z \sim \mathcal{N}(z, 1)$ and $Y \mid Z = z \sim \mathcal{N}(z, 1)$
- $Z \sim Bi(3, 0.5)$, $X \sim Exp(1)$ and $Y \mid X = x \sim 0.15\delta_0 + 0.85 Pois(x)$

(conditional) Independence

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

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

Illustration

- $Z \sim Bi(9, 0.5)$, $X \mid Z = z \sim \mathcal{N}(z, 1)$ and $Y \mid Z = z \sim \mathcal{N}(z, 1)$
- $Z \sim Bi(3, 0.5)$, $X \sim Exp(1)$ and $Y \mid X = x \sim 0.15\delta_0 + 0.85 Pois(x)$

(conditional) Independence

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

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

Illustration

- $Z \sim Bi(9, 0.5)$, $X \mid Z = z \sim \mathcal{N}(z, 1)$ and $Y \mid Z = z \sim \mathcal{N}(z, 1)$
- $Z \sim Bi(3, 0.5)$, $X \sim Exp(1)$ and $Y \mid X = x \sim 0.15\delta_0 + 0.85Pois(x)$

Probabilistic causal models (1)

A tuple $M = \langle(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))\rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables which can't be manipulated
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables
3. $\mathcal{F}$ is a set of functions s.t. f_i $(1 \leq i \leq n)$ specifies X_i: $X_i = f(\mathcal{E}_i)$ with $\mathcal{E}_i \subseteq \mathcal{U} \cup \mathcal{V}$
4. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

Probabilistic causal models (1)

A tuple $M = \langle(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))\rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables which can't be manipulated
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables
3. $\mathcal{F}$ is a set of functions s.t. f_i $(1 \leq i \leq n)$ specifies X_i: $X_i = f(\mathcal{E}_i)$ with $\mathcal{E}_i \subseteq \mathcal{U} \cup \mathcal{V}$
4. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

Probabilistic causal models (1)

A tuple $M = \langle(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))\rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables which can't be manipulated
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables
3. $\mathcal{F}$ is a set of functions s.t. f_i $(1 \leq i \leq n)$ specifies X_i: $X_i = f(\mathcal{E}_i)$ with $\mathcal{E}_i \subseteq \mathcal{U} \cup \mathcal{V}$
4. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

Probabilistic causal models (1)

A tuple $M = \langle(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))\rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables which can't be manipulated
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables
3. $\mathcal{F}$ is a set of functions s.t. f_i $(1 \leq i \leq n)$ specifies X_i: $X_i = f(\mathcal{E}_i)$ with $\mathcal{E}_i \subseteq \mathcal{U} \cup \mathcal{V}$
4. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

Probabilistic causal models (1)

A tuple $M = \langle(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))\rangle$ with

1. $\mathcal{U}$ is a set of unobserved background variables which can't be manipulated
2. $\mathcal{V} = \{X_1, ..., X_n\}$ is a set of observed variables
3. $\mathcal{F}$ is a set of functions s.t. f_i $(1 \leq i \leq n)$ specifies X_i: $X_i = f(\mathcal{E}_i)$ with $\mathcal{E}_i \subseteq \mathcal{U} \cup \mathcal{V}$
4. $P(\mathcal{U})$ is a joint distribution over $\mathcal{U}$

Probabilistic causal models (2)

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

$$
\begin{aligned}
P(\mathcal{V}) &= \sum_{u \in D_U} P(\mathcal{V}, u) \\
&= \sum_{u \in D_U} P(\mathcal{V} \mid u) P(u) \\
&= \sum_{u \in D_U} \prod_{i=1}^{n} P(x_i \mid x_1, \dots, x_{i-1}, u) P(u)
\end{aligned}
$$

Probabilistic causal models (3)

Induced graph *The graph $\mathcal{G}(M)$ induced by a probabilistic causal model M has vertices $\mathcal{V}$ and an edge $X_i \rightarrow X_j$ whenever f_i depends on X_j. In addition, 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 (causal sufficiency)*

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

Probabilistic causal models (3)

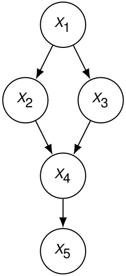
Induced graph *The graph $\mathcal{G}(M)$ induced by a probabilistic causal model M has vertices $\mathcal{V}$ and an edge $X_i \rightarrow X_j$ whenever f_i depends on X_j. In addition, G contains a bidirected edge, denoted $X_i \leftarrow\!\!-\!\!-\!\!\rightarrow X_j$, whenever f_i and f_j depend on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

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

Example

X_1: season (can take on 4 values)
X_2: rain (binary yes/no)
X_3: sprinkler (binary on/off)
X_4: wet (binary yes/no)
X_5: slippery (binary yes/no)

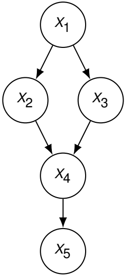


With no confounders:

$$P(\mathcal{V}) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1) P(X_4 \mid X_2, X_3)P(X_5 \mid X_4)$$

Example

X_1: season (can take on 4 values)
X_2: rain (binary yes/no)
X_3: sprinkler (binary on/off)
X_4: wet (binary yes/no)
X_5: slippery (binary yes/no)

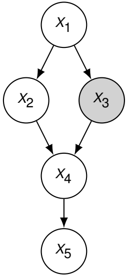


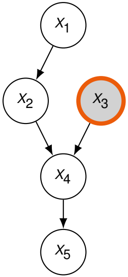
With no confounders:

$$P(\mathcal{V}) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1) P(X_4 \mid X_2, X_3)P(X_5 \mid X_4)$$

Example (cont'd)

Conditioning vs intervention



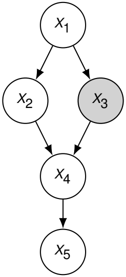


Example (cont'd)

Conditioning

$$P(X_1, X_2, X_4, X_5 \mid X_3 = off) = \frac{P(X_1, X_2, X_4, X_5, X_3 = off)}{\sum_{x_1} P(X_1 = x_1) P(X_3 = off \mid X_1 = x_1)}$$

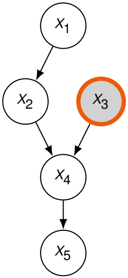
$$= \frac{P(X_1) P(X_2 \mid X_1) P(X_3 = off \mid X_1) P(X_4 \mid X_2, X_3 = off) P(X_5 \mid X_4)}{\sum_{x_1} P(X_1 = x_1) P(X_3 = off \mid X_1 = x_1)}$$



Example (cont'd)

Intervention

$$P_{X_3=off}(X_1, X_2, X_4, X_5) = P(X_1)P(X_2 \mid X_1)P(X_4 \mid X_2, X_3 = off)P(X_5 \mid X_4)$$



Example (cont'd)

Conditioning vs intervention

$P(X_1, X_2, X_4, X_5 \mid X_3 = off)$ vs $P_{X_3=off}(X_1, X_2, X_4, X_5)$

Identification (identifiability)

Example (cont'd)

Conditioning vs intervention

$P(X_1, X_2, X_4, X_5 \mid X_3 = off)$ vs $P_{X_3=off}(X_1, X_2, X_4, X_5)$

Identification (identifiability)

Example (cont'd)

Conditioning vs intervention

$$P(X_1, X_2, X_4, X_5 \mid X_3 = off) \text{ vs } P_{X_3=off}(X_1, X_2, X_4, X_5)$$

Identification (identifiability)

Interventional theory

Causation in the interventional theory

- *A* causes *B* if and only if there is a possible intervention on *A* which changes *B*
- An intervention on *A* must completely disrupt the causal relation between *A* and its previous causes so that the value of *A* is entirely fixed by this intervention

Interventional theory

Causation in the interventional theory

- *A* causes *B* if and only if there is a possible intervention on *A* which changes *B*
- An intervention on *A* must completely disrupt the causal relation between *A* and its previous causes so that the value of *A* is entirely fixed by this intervention

Interventional theory

Causation in the interventional theory

- *A* causes *B* if and only if there is a possible intervention on *A* which changes *B*
- An intervention on *A* must completely disrupt the causal relation between *A* and its previous causes so that the value of *A* is entirely fixed by this intervention

Causal discovery vs causal inference

Causal discovery From observational data, infer causal graph with or without hidden confounders (hidden common causes) - courses 2, 3 and 4

Causal inference Reasoning on the causal graph through interventions (and asking counterfactual questions) - courses 5 and 6

Causal discovery vs causal inference

Causal discovery From observational data, infer causal graph with or without hidden confounders (hidden common causes) - courses 2, 3 and 4

Causal inference Reasoning on the causal graph through interventions (and asking counterfactual questions) - courses 5 and 6

Causal discovery vs causal inference

Causal discovery From observational data, infer causal graph with or without hidden confounders (hidden common causes) - courses 2, 3 and 4

Causal inference Reasoning on the causal graph through interventions (and asking counterfactual questions) - courses 5 and 6

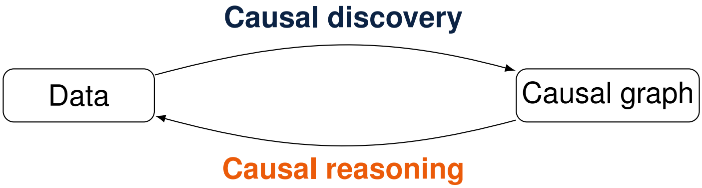
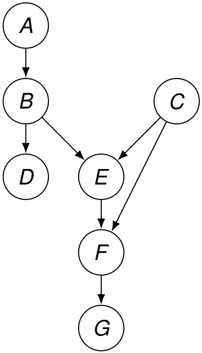


Table of content

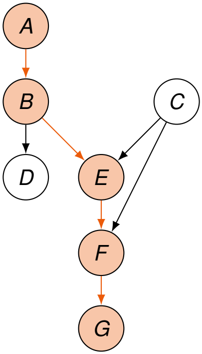
Basic graph concepts

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



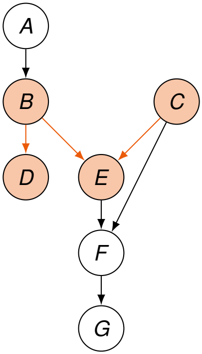
Basic graph concepts (cont'd)

Directed path: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$ ($A \rightsquigarrow G$)



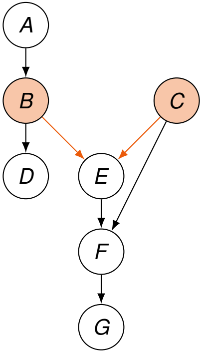
Basic graph concepts (cont'd)

Path (trail): $D \leftarrow B \rightarrow E \leftarrow C$ ($D \circ\!\!\sim\!\!\sim\!\!\sim\!\!\circ C$)



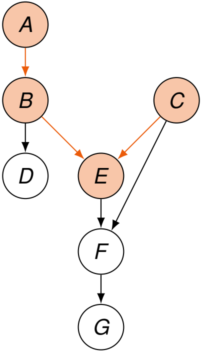
Basic graph concepts (cont'd)

Parents, ancestors: $Pa(E) = \{B, C\}$,
An: transitive closure of the parents relation



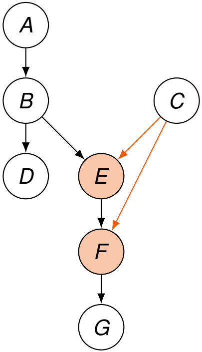
Basic graph concepts (cont'd)

Parents, ancestors: $Pa(E) = \{B, C\}$, $An(E) = \{A, B, C, E\}$
An: transitive closure of the parents relation



Basic graph concepts (cont'd)

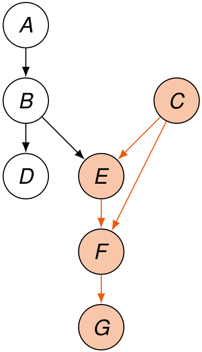
Children, descendants: $Ch(C) = \{E, F\}$,
De: transitive closure of the children relation



Basic graph concepts (cont'd)

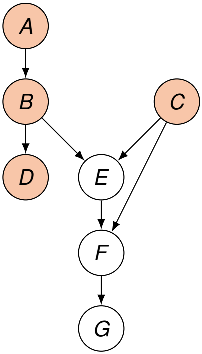
Children, descendants: $Ch(C) = \{E, F\}$, $De(C) = \{C, E, F, G\}$

De: transitive closure of the children relation



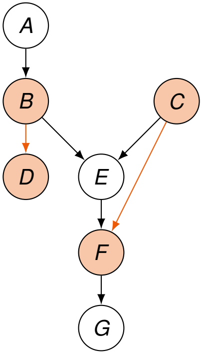
Basic graph concepts (cont'd)

Upwards-closed sets: a subset of nodes $\mathcal{S}$ is upward-closed (or ancestral) if $\forall S \in \mathcal{S}, An(S) \subseteq \mathcal{S}$



Basic graph concepts (cont'd)

Induced subgraph $\mathcal{G}[\mathcal{S}]$: $\mathcal{G}[\{B, C, D, F\}]$



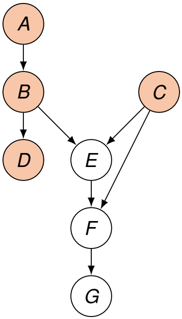
Bayesian networks and compatibility

A Bayesian network is a DAG (directed acyclic graph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *along with a joint distribution* $P(\mathcal{V})$ *that admits the factorization* $P(\mathcal{V}) = \prod_{X \in \mathcal{V}} P(X \mid Pa_G(X))$

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 $P(\mathcal{V}) = \prod_{X \in \mathcal{V}} P(X \mid Pa(X))$. We denote by $\mathcal{P}(\mathcal{V})$ the set of distributions compatible with $\mathcal{G}$.

Observation

Upwards-closed set If P is compatible with $\mathcal{G}$ and $\mathcal{S} \subseteq \mathcal{V}$ is upwards-closed, then $P(\mathcal{S})$ is compatible with $\mathcal{G}[\mathcal{S}]$, *i.e.*, $P(\mathcal{S}) = \prod_{S \in \mathcal{S}} P(S \mid Pa(S))$

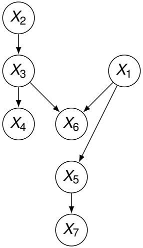


Markov conditions

Ordered Markov condition P is compatible with $\mathcal{G}$ *iff* in any topological ordering each X_i is independent of its non-descendants given its parents

Topological ordering: for any edge $X_i \rightarrow X_j$, $i < j$

Parental Markov condition P is compatible with $\mathcal{G}$ *iff* every variable is independent of its non-descendants given its parents



Conditioning on common ancestors

Property 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

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

in any distribution P compatible with $\mathcal{G}$

Causal Bayesian networks

Causal Markov condition Every Markovian causal model M induces a distribution that is compatible with the induced graph $\mathcal{G}[M]$

Causal Bayesian network (Pearl 2000) Let $P(\mathcal{V})$ be a probability distribution and let $P_s(\mathcal{V})$ 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 $P_s(\mathcal{V})$. A DAG $\mathcal{G}$ is said to be a *causal Bayesian network* compatible with $\mathcal{P}_*$ *iff* for every $P_s(\mathcal{V}) \in \mathcal{P}_*$:

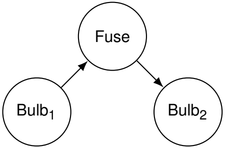
(i) $P_s(\mathcal{V})$ is Markov relative to $\mathcal{G}$

(ii) $P_s(s_i) = 1$ or all $S_i \in \mathcal{S}$ whenever s_i is consistent with $\mathcal{S} = s$

(ii) $P_s(x_i \mid Pa(X_i)) = P(x_i \mid Pa(X_i))$ for all $X_i \notin \mathcal{S}$ whenever $Pa(X_i)$ is consistent with $\mathcal{S} = s$

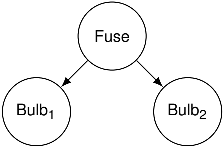
Causal Bayesian networks: example

Bayesian networks vs causal graph



Non causal graph

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

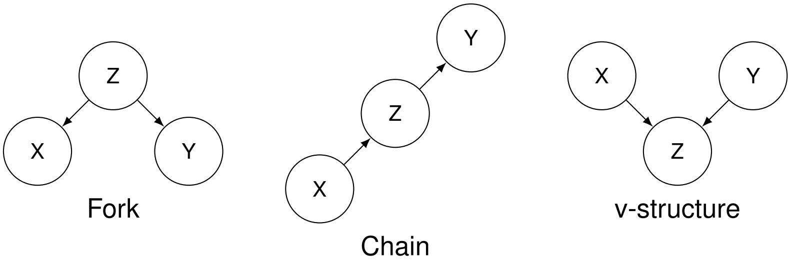


Causal graph

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

- $F \sim \mathcal{U}\{0,1\}$
- $P(Bulb_1 = 1 \mid Fuse = 1) = 1 - \epsilon_1,\ P(Bulb_1 = 0 \mid Fuse = 1) = \epsilon_1$
- $P(Bulb_2 = 1 \mid Fuse = 1) = 1 - \epsilon_2,\ P(Bulb_2 = 0 \mid Fuse = 1) = \epsilon_2$
- $P(Bulb_1 = 1 \mid Fuse = 0) = P(Bulb_2 = 1 \mid Fuse = 0) = 0$
- $P(Bulb_1 = 0 \mid Fuse = 0) = P(Bulb_2 = 0 \mid Fuse = 0) = 1$
- $\epsilon_1, \epsilon_2 \sim \mathcal{U}[0; 0.1]$

Forks, chains and v-structures



Exploiting (in)dependencies in observational data

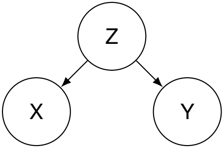
$X, Y \sim U(-1, 1)$

$Corr(X, Y) = 0.002$

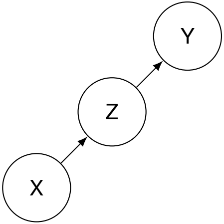
$Z = 2XY + \xi_C,\ \xi_C \sim N(0, \frac{1}{2})$

$Corr(X; Y \mid Z > 0.5) = 0.8$

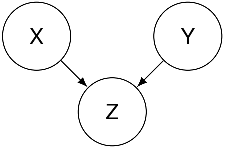
Forks, chains and v-structures



Fork



Chain



v-structure

Exploiting (in)dependencies in observational data

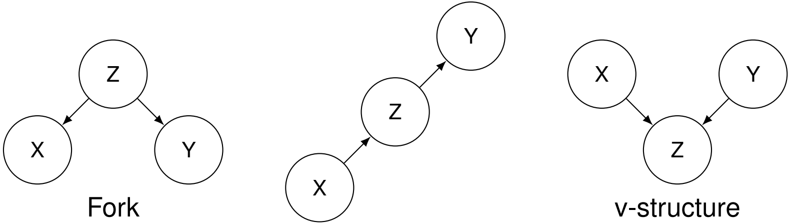
$X, Y \sim U(-1, 1)$

$Z = 2XY + \xi_c, \ \xi_c \sim N(0, \frac{1}{2})$

$Corr(X, Y) = 0.002$

$Corr(X; Y \mid Z > 0.5) = 0.8$

Forks, chains and v-structures



Fork

Chain

v-structure

Exploiting (in)dependencies in observational data

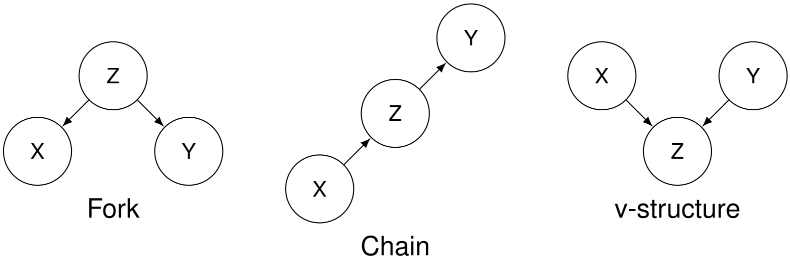
$$X, Y \sim U(-1, 1)$$

$$Corr(X, Y) = 0.002$$

$$Z = 2XY + \xi_c, \ \xi_c \sim N(0, \frac{1}{2})$$

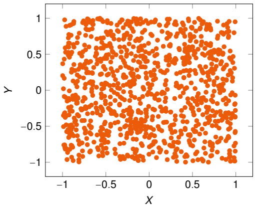
$$Corr(X; Y \mid Z > 0.5) = 0.8$$

Forks, chains and v-structures



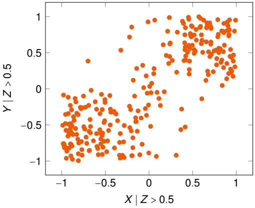
Exploiting (in)dependencies in observational data

$$X, Y \sim U(-1, 1)$$



$$Corr(X, Y) = 0.002$$

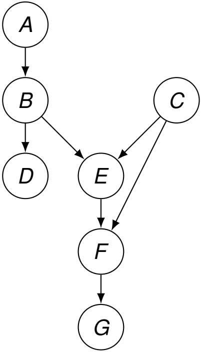
$$Z = 2XY + \xi_C, \ \xi_C \sim N(0, \frac{1}{2})$$



$$Corr(X; Y \mid Z > 0.5) = 0.8$$

Reading conditional independencies in graphs

What conditional independencies hold in a distribution P compatible with a given graph $\mathcal{G}$?

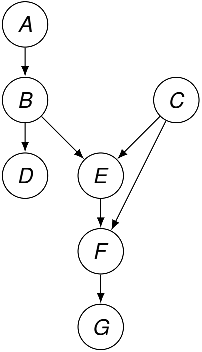


- $A \perp\!\!\!\perp_P D \mid B$
- $E \not\perp\!\!\!\perp_P F \mid C$
- $B \perp\!\!\!\perp_P F \mid E$?

By definition: $\mathcal{I}_{prob}(P) := \{(X, Y, Z), X \perp\!\!\!\perp_P Y \mid Z\}$

Reading conditional independencies in graphs

What conditional independencies hold in a distribution P compatible with a given graph $\mathcal{G}$?

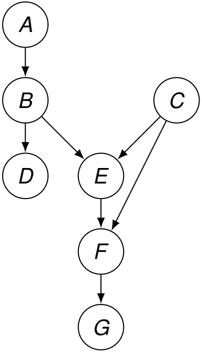


- $A \perp\!\!\!\perp_P D \mid B$
- $E \not\perp\!\!\!\perp_P F \mid C$
- $B \perp\!\!\!\perp_P F \mid E$?

By definition: $\mathcal{I}_{prob}(P) := \{(X, Y, Z), X \perp\!\!\!\perp_P Y \mid Z\}$

Reading conditional independencies in graphs

What conditional independencies hold in a distribution P compatible with a given graph $\mathcal{G}$?

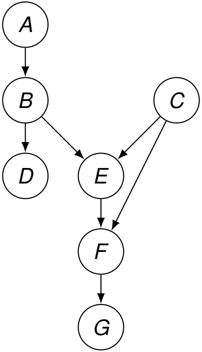


- $A \perp\!\!\!\perp_P D \mid B$
- $E \not\perp\!\!\!\perp_P F \mid C$
- $B \perp\!\!\!\perp_P F \mid E$?

By definition: $\mathcal{I}_{prob}(P) := \{(X, Y, Z), X \perp\!\!\!\perp_P Y \mid Z\}$

Reading conditional independencies in graphs

What conditional independencies hold in a distribution P compatible with a given graph $\mathcal{G}$?

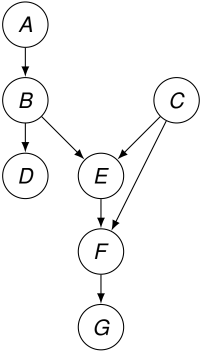


- $A \perp\!\!\!\perp_P D \mid B$
- $E \not\perp\!\!\!\perp_P F \mid C$
- $B \perp\!\!\!\perp_P F \mid E$?

By definition: $\mathcal{I}_{prob}(P) := \{(X, Y, Z), X \perp\!\!\!\perp_P Y \mid Z\}$

Reading conditional independencies in graphs

What conditional independencies hold in a distribution P compatible with a given graph $\mathcal{G}$?



- $A \perp\!\!\!\perp_P D \mid B$
- $E \not\perp\!\!\!\perp_P F \mid C$
- $B \perp\!\!\!\perp_P F \mid E$?

By definition: $\mathcal{I}_{prob}(P) := \{(X, Y, Z), X \perp\!\!\!\perp_P Y \mid Z\}$

d-Separation

Collider *A* collider is a directed graph isomorphic to $X \rightarrow Z \leftarrow Y$. We'll refer to Z in a collider as *the* collider. If the two parent vertices are not adjacent, the collider is a *v-structure* (also called *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

d-Separation

Collider A collider is a directed graph isomorphic to $X \to Z \leftarrow Y$. We'll refer to Z in a collider as *the* collider. If the two parent vertices are not adjacent, the collider is a *v-structure* (also called *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 $\mathcal{Z}$

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

d-Separation

Collider A collider is a directed graph isomorphic to $X \rightarrow Z \leftarrow Y$. We'll refer to Z in a collider as *the* collider. If the two parent vertices are not adjacent, the collider is a *v-structure* (also called *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

d-Separation (cont'd)

d-separation 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 $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$. By definition:

$$\mathcal{I}_{d-sep}(\mathcal{G}) := \{\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z} : \mathcal{X}, \mathcal{Y}, \mathcal{Z} \textit{ disjoint sets}\}$$

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 (cont'd)

d-separation 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 $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$. By definition:

$$\mathcal{I}_{d-sep}(\mathcal{G}) := \{\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z} : \mathcal{X}, \mathcal{Y}, \mathcal{Z} \textit{ disjoint sets}\}$$

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 (cont'd)

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 $\mathcal{G}$

(i) *Completeness* If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$

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

d-Separation (cont'd)

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 $\mathcal{G}$

(i) *Completeness* If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$

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

d-Separation (cont'd)

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 $\mathcal{G}$

(i) *Completeness* If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$

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

d-Separation (cont'd)

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 $\mathcal{G}$

(i) *Completeness* If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$

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

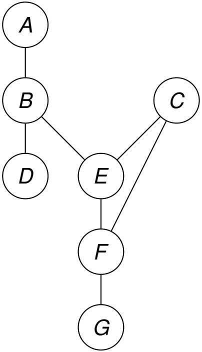
Table of content

Markov equivalence

Theorem (Markov equivalence) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same d-separations *iff* they have the same skeleton and the same v-structures

Markov equivalence

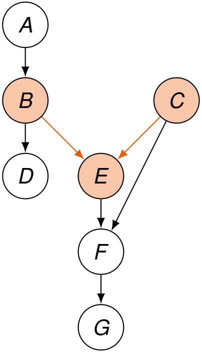
Theorem (Markov equivalence) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same d-separations *iff* they have the same skeleton and the same v-structures



- Skeleton is the undirected graph with same adjacencies

Markov equivalence

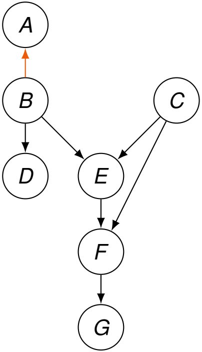
Theorem (Markov equivalence) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same d-separations *iff* they have the same skeleton and the same v-structures



- Skeleton is the undirected graph with same adjacencies
- v-structure: collider $X \rightarrow Z \leftarrow Y$ s.t. X and Y are not adjacent

Markov equivalence

Theorem (Markov equivalence) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same d-separations *iff* they have the same skeleton and the same v-structures



- Skeleton is the undirected graph with same adjacencies
- v-structure: collider $X \rightarrow Z \leftarrow Y$ s.t. X and Y are not adjacent
- Flipping some edges may not change d-separation

Table of content

Minimality and faithfulness

Causal Markov condition in practice (*i.e.* using observational data) may be too loose. In particular, one wants to impose that the graph does not contain dependencies not present in the observational data

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$

May not be sufficient to rule out special cases when the probability distribution leads to cancellation of some causal relations

Faithfulness We say that a graph $\mathcal{G}$ and a compatible probability distribution P are faithful to one another if all and only the conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$

Minimality and faithfulness

Causal Markov condition in practice (*i.e.* using observational data) may be too loose. In particular, one wants to impose that the graph does not contain dependencies not present in the observational data

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$

May not be sufficient to rule out special cases when the probability distribution leads to cancellation of some causal relations

Faithfulness We say that a graph $\mathcal{G}$ and a compatible probability distribution P are faithful to one another if all and only the conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$

Minimality and faithfulness

Causal Markov condition in practice (*i.e.* using observational data) may be too loose. In particular, one wants to impose that the graph does not contain dependencies not present in the observational data

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$

May not be sufficient to rule out special cases when the probability distribution leads to cancellation of some causal relations

Faithfulness We say that a graph $\mathcal{G}$ and a compatible probability distribution P are faithful to one another if all and only the conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$

Minimality and faithfulness

Causal Markov condition in practice (*i.e.* using observational data) may be too loose. In particular, one wants to impose that the graph does not contain dependencies not present in the observational data

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$

May not be sufficient to rule out special cases when the probability distribution leads to cancellation of some causal relations

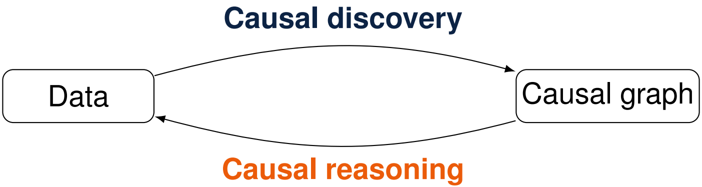
Faithfulness We say that a graph $\mathcal{G}$ and a compatible probability distribution P are faithful to one another if all and only the conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$

Conclusion

Bayesian networks, causal graphical models

Conclusion

Bayesian networks, causal graphical models



References (1)

Direct inspirations

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

References (2)

Additional readings

1. *Equivalence and Synthesis of Causal Models*, T. S. Verma, J. Pearl. Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, 1990
2. *Graphical aspects of causal models*, T. S. Verma. Technical report R-191, UCLA, 1993
3. *Probabilistic Graphical Models: Principles and Techniques*, D. Koller, N. Friedman. MIT Press, 2009
4. *Does Obesity Shorten Life? Or is it the Soda?*, J. Pearl. Journal of Causal Inference, 6(2), 2018