DAG et identification

Charles Assaad charles.assaad@inserm.fr

L'Institut Pierre Louis d'Epidémiologie et de Santé Publique, Inserm, Sorbonne Université

Decembre 2024

SANTÉ SORBONNE UNIVERSITÉ

Sommaire

SANTÉ SORBONNE UNIVERSITÉ

1

Motivation and histoire

Effet causal ou effet total

$$\text{L'effet causal de } \mathbb{A} \text{ sur } \mathbb{Y}$$
$$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$$

L'opérateur $do()$ est une manière de désigner des interventions.

Effet causal ou effet total

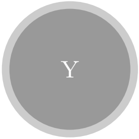
$$\text{L'effet causal de } \mathbb{A} \text{ sur } \mathbb{Y}$$
$$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$$

L'opérateur $do()$ est une manière de désigner des interventions.

$\mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}))$ peut être remplacer par $\mathbb{E}(\mathbb{Y}^{\mathbb{A}=a})$.

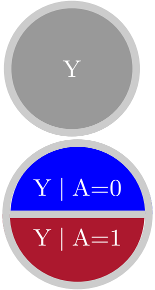
Conditionnement Vs Intervention

Population



Population

Sous-populations

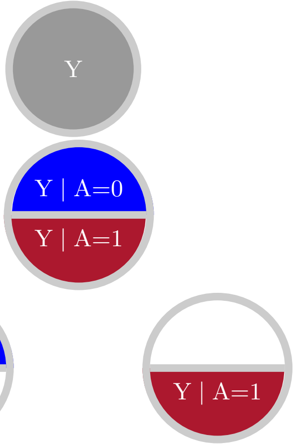


Conditionnement Vs Intervention

Population

Sous-populations

Conditionnement



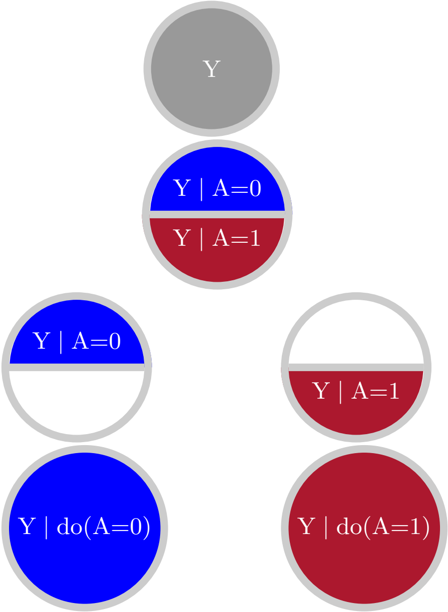
Conditionnement Vs Intervention

Population

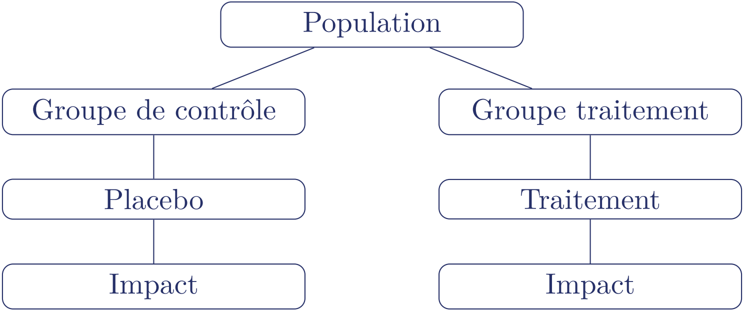
Sous-populations

Conditionnement

Intervention



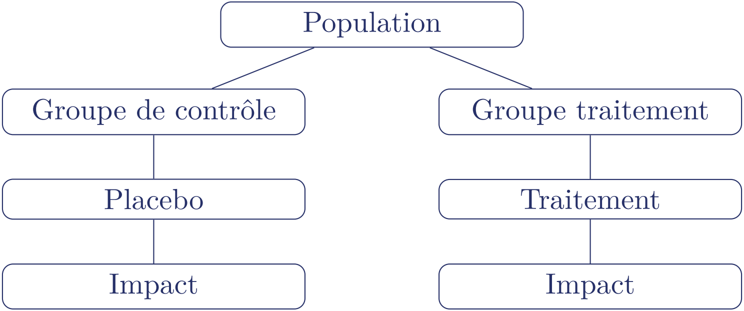
Etalon d'or: Essai randomisé contrôlés ou essai clinique



Contrôle:

- Permet d'éliminer les biais temporels

Etalon d'or: Essai randomisé contrôlés ou essai clinique



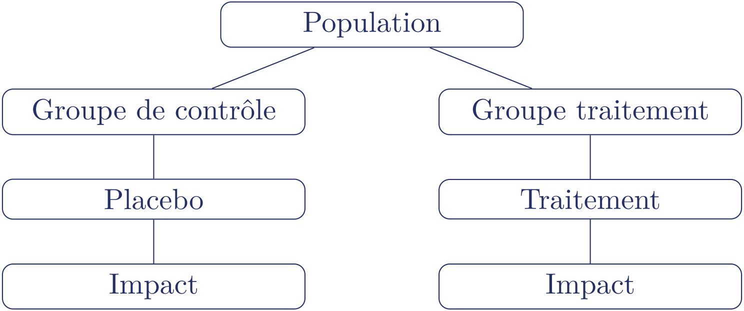
Contrôle:

- Permet d'éliminer les biais temporels

Sélection aléatoire:

- Permet d'éliminer les biais des facteurs extérieurs

Etalon d'or: Essai randomisé contrôlés ou essai clinique



Contrôle:

- Permet d'éliminer les biais temporels

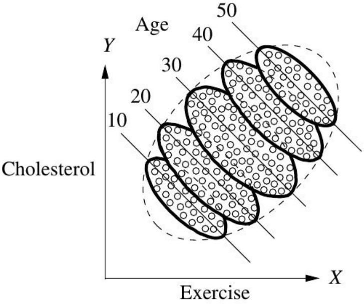
Sélection aléatoire:

- Permet d'éliminer les biais des facteurs extérieurs

Limitations:

- Coûteux
- Non éthique
- Infaisable

Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol ?



Causal inference

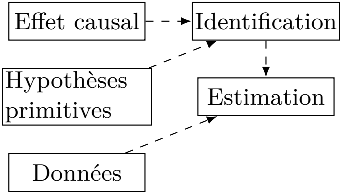
Effet causal → Identification → Estimation

Causal inference

Effet causal $\dashrightarrow$ Identification $\dashrightarrow$ Estimation

Causal inference

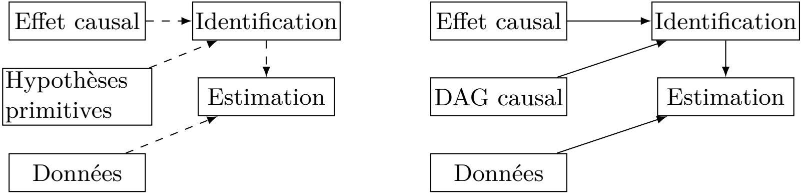
Two main frameworks



Causal inference

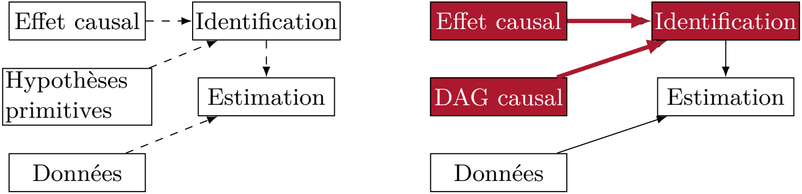
Effet causal → Identification → Estimation

Two main frameworks



Causal inference

Two main frameworks



Causal inference

Effet causal → Identification → Estimation

Two main frameworks

Effet causal → Identification

Hypothèses primitives → Identification

Identification → Estimation

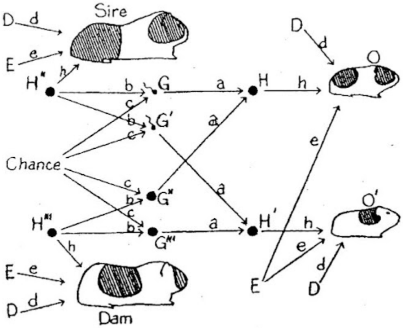
Données → Estimation

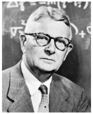
Effet causal → Identification

DAG causal → Identification

Identification → Estimation

Données → Estimation





Sewal Wright

Effet causal → Identification → Estimation

Two main frameworks

Effet causal → Identification

Hypothèses primitives → Identification

Données → Estimation

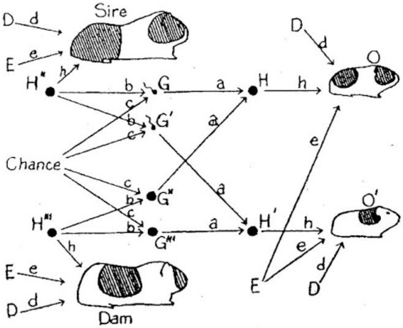
Identification → Estimation

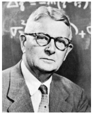
Effet causal → Identification

DAG causal → Identification

Données → Estimation

Identification → Estimation





Sewal Wright

Judea Pearl

2

Préliminaires

Probabilité

A et B sont indépendants conditionnellement à C, noté $A \perp\!\!\!\perp_P B \mid C$, si et seulement si :

$$P(A, B \mid C) = P(A \mid C)P(B \mid C)$$
$$\text{or } P(A \mid B, C) = P(A \mid C) \, if \, P(B, C) > 0$$

Théorème de Bayes :

$$P(B \mid A) = \frac{P(B, A)}{P(A)}$$

Marginalisation:

$$\sum_a P(B, A) = P(B)$$

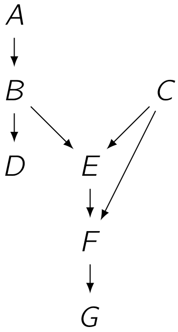
Graphe orienté

Un graphe $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ est appelé graphe orienté si, et seulement si :

- $\mathbb{V}$ représente l'ensemble des nœuds (chaque nœud correspond généralement à une variable aléatoire),
- $\mathbb{E}$ désigne l'ensemble des arêtes.
- $\forall (X, Y) \in \mathbb{E}$, Il existe une flèche orientée de X vers Y.

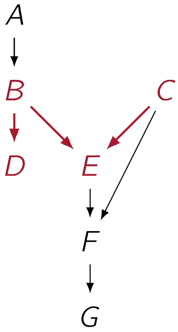
Graphe orienté: notions

Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Graphe orienté: notions

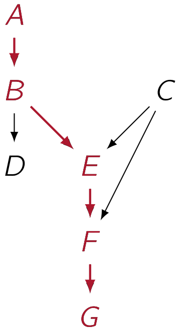
Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

Graphe orienté: notions

Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

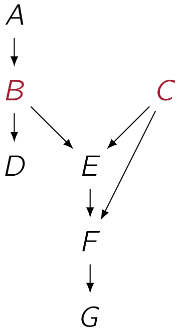


Chemin: $D \leftarrow B \rightarrow E \leftarrow C$

Chemin orienté: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Graphe orienté: notions

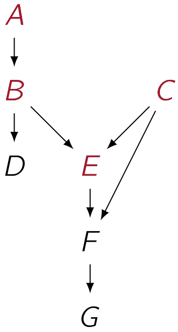
Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



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

Graphe orienté: notions

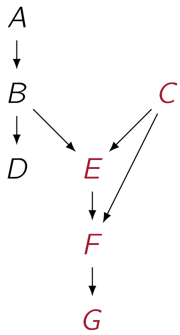
Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Chemin: $D \leftarrow B \rightarrow E \leftarrow C$
Chemin orienté: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancêtres: $An(E) = \{A, B, C, E\}$

Graphe orienté: notions

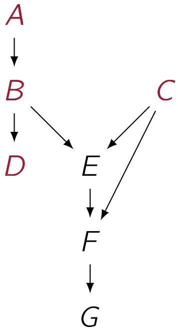
Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Chemin: $D \leftarrow B \rightarrow E \leftarrow C$
Chemin orienté: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancêtres: $An(E) = \{A, B, C, E\}$
Descendants: $De(C) = \{C, E, F, G\}$

Graphe orienté: notions

Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:



Chemin: $D \leftarrow B \rightarrow E \leftarrow C$
Chemin orienté: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
Parents: $Pa(E) = \{B, C\}$
Ancêtres: $An(E) = \{A, B, C, E\}$
Descendants: $De(C) = \{C, E, F, G\}$
Non-descendants: $Nd(E) = \{A, B, C, D\}$

3

DAG causal

- Définition
- d-separation

3

DAG causal

Définition

d-separation

DAG

Un graphe orienté $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ est appelé graphe orienté acyclique (DAG) si et seulement si

$$\forall X \in \mathbb{V}, An(X) \cap De(X) = \{X\},$$

c'est-à-dire qu'il n'existe aucun cycle dans $\mathcal{G}$.

DAG

Un graphe orienté $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ est appelé graphe orienté acyclique (DAG) si et seulement si

$$\forall X \in \mathbb{V}, An(X) \cap De(X) = \{X\},$$

c'est-à-dire qu'il n'existe aucun cycle dans $\mathcal{G}$.

Dorénavant, nous nous concentrerons uniquement sur les DAGs.

DAG

Un graphe orienté $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ est appelé graphe orienté acyclique (DAG) si et seulement si

$$\forall X \in \mathbb{V}, An(X) \cap De(X) = \{X\},$$

c'est-à-dire qu'il n'existe aucun cycle dans $\mathcal{G}$.

Dorénavant, nous nous concentrerons uniquement sur les DAGs.

Une distribution $P(\mathbb{V})$ est compatible avec un DAG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ si

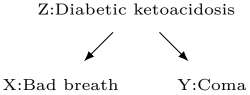
$$P(\mathbb{V}) = \prod_{X \in \mathbb{V}} P(X \mid Pa(X)).$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

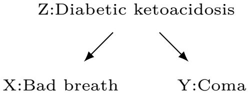
DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.



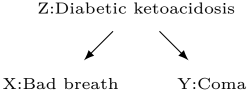
DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.



Causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.



Causal

$$P(x, y, z) = P(y \mid z)P(x \mid z)P(z)$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

$$\begin{tikzcd} & \text{Z:Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\ \text{X:Bad breath} & & \text{Y:Coma} \end{tikzcd}$$

$$\begin{tikzcd} & \text{Z:Diabetic ketoacidosis} \arrow[dr] & \\ \text{X:Bad breath} \arrow[ur] & & \text{Y:Coma} \end{tikzcd}$$

Causal

$$P(x, y, z) = P(y \mid z)P(x \mid z)P(z)$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

$$\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\
\text{X:Bad breath} & & \text{Y:Coma}
\end{tikzcd}$$

Causal

$$\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dr] & \\
\text{X:Bad breath} \arrow[ur] & & \text{Y:Coma}
\end{tikzcd}$$

Pas causal

$$P(x, y, z) = P(y \mid z)P(x \mid z)P(z)$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

$$\begin{tikzcd} & \text{Z:Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\ \text{X:Bad breath} & & \text{Y:Coma} \end{tikzcd}$$

Causal

$$\begin{tikzcd} & \text{Z:Diabetic ketoacidosis} \arrow[dr] & \\ \text{X:Bad breath} \arrow[ur] & & \text{Y:Coma} \end{tikzcd}$$

Pas causal

$$P(x, y, z) = P(y \mid z)P(x \mid z)P(z)$$

$$P(x, y, z) = P(y \mid z)P(z \mid x)P(x)$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

$$
\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\
\text{X:Bad breath} & & \text{Y:Coma}
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dr] & \\
\text{X:Bad breath} \arrow[ur] & & \text{Y:Coma}
\end{tikzcd}
$$

Causal   Pas causal

$$
\begin{aligned}
P(x, y, z) &= P(y \mid z)P(x \mid z)P(z) \\
&= P(y \mid z)P(z, x) && \text{Théorème de Bayes} \\
&= P(y \mid z)\frac{P(z, x)}{P(x)}P(x) \\
P(x, y, z) &= P(y \mid z)P(z \mid x)P(x) && \text{Théorème de Bayes}
\end{aligned}
$$

DAG Vs DAG causal

Tous les DAGs compatibles avec une distribution ne sont pas nécessairement causaux.

$$\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dl] \arrow[dr] & \\
\text{X:Bad breath} & & \text{Y:Coma}
\end{tikzcd}$$

Causal

$$\begin{tikzcd}
 & \text{Z:Diabetic ketoacidosis} \arrow[dr] & \\
\text{X:Bad breath} \arrow[ur] & & \text{Y:Coma}
\end{tikzcd}$$

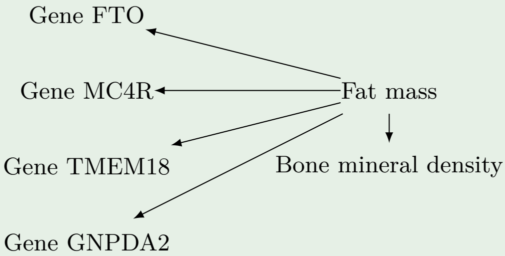
Pas causal

$$
\begin{aligned}
P(x,y,z) &= P(y \mid z)P(x \mid z)P(z) \\
&= P(y \mid z)P(z,x) && \text{Théorème de Bayes} \\
&= P(y \mid z)\frac{P(z,x)}{P(x)}P(x) \\
P(x,y,z) &= P(y \mid z)P(z \mid x)P(x) && \text{Théorème de Bayes}
\end{aligned}
$$

$\Longrightarrow$ Il est impossible de déterminer lequel des deux DAGs est causal en se basant uniquement sur la distribution observée !

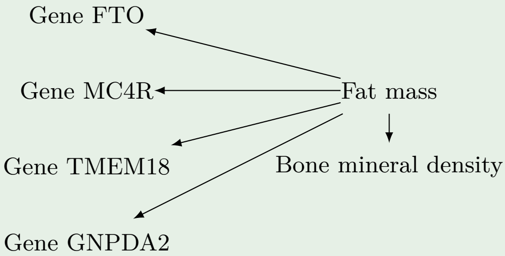
DAG Vs DAG causal: exemples

Example



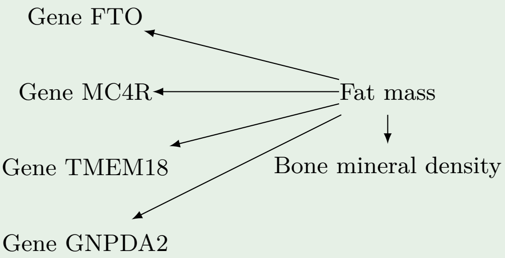
Est ce que ce DAG est causal?

Example

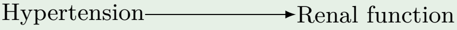


Est ce que ce DAG est causal? Non

Example



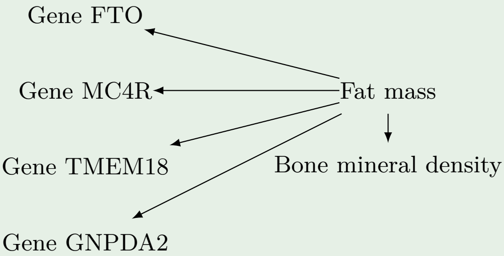
Est ce que ce DAG est causal? Non



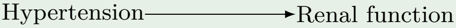
Est ce que ce DAG est causal?

DAG Vs DAG causal: exemples

Example



Est ce que ce DAG est causal? Non



Est ce que ce DAG est causal? Oui

Un DAG est causal si et seulement si:

- il est compatible avec la distribution observée $P(\mathbb{V})$ et
- il est compatible avec toute distribution interventionnelle $P(\mathbb{V} \backslash \mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V} \backslash \mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(\mathbb{v}_i \mid Pa(\mathbb{v}_i))$$

DAG causal

Un DAG est causal si et seulement si:

- il est compatible avec la distribution observée $P(\mathbb{V})$ et
- il est compatible avec toute distribution interventionnelle $P(\mathbb{V}\backslash\mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V}\backslash\mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(\mathbb{v}_i \mid Pa(\mathbb{v}_i))$$

Autrement dit, on suppose que derrière le DAG cache un générateur causal: $\forall Y \in \mathbb{V} \qquad Y := f_y(Pa(Y), \xi_y)$.

DAG causal

Un DAG est causal si et seulement si:

- il est compatible avec la distribution observée $P(\mathbb{V})$ et
- il est compatible avec toute distribution interventionnelle $P(\mathbb{V} \backslash \mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V} \backslash \mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(\mathbb{v}_i \mid Pa(\mathbb{v}_i))$$

Autrement dit, on suppose que derrière le DAG cache un générateur causal: $\forall Y \in \mathbb{V} \qquad Y := f_y(Pa(Y), \xi_y)$.

Dans ce qui suit, nous nous limiterons exclusivement aux DAGs causaux.

Code: Define a DAG using DAGitty package

```
#Load the DAGitty package
library(dagitty)
#Define the DAG
dag <- dagitty("Z->X Z->Y")
#Display the DAG
print(dag)
#Visualize the DAG
plot(dag)
```

Résumé de la cette section

- Tous les DAGs ne sont pas causaux.
- Un DAG causal doit être compatible avec la distribution observée ainsi qu'avec toutes les distributions interventionnelles.
- Le package DAGitty permet de créer et de visualiser des DAGs dans R.

SANTÉ SORBONNE UNIVERSITÉ

3

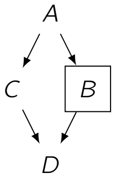
DAG causal

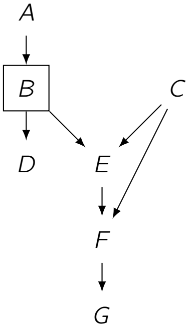
Définition

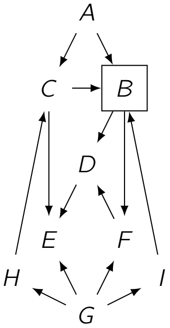
d-separation

Reading conditional independencies in graphs

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







Basic structures

Basic structures

Fourchette: contenant un facteur de confusion

$$A \longleftarrow C \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B$$

$$A \longleftarrow \boxed{C} \longrightarrow B$$

$$A \perp\!\!\!\perp_P B \mid C$$

Basic structures

Fourchette: contenant un facteur de confusion

$$A \longleftarrow C \longrightarrow B \qquad\qquad A \longleftarrow \boxed{C} \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \qquad\qquad A \perp\!\!\!\perp_P B \mid C$$

Chaîne: contenant mediateur ou cause intermédiaire

$$A \longrightarrow C \longrightarrow B \qquad\qquad A \longrightarrow \boxed{C} \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \qquad\qquad A \perp\!\!\!\perp_P B \mid C$$

Basic structures

Fourchette: contenant un facteur de confusion

$$A \longleftarrow C \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B$$

$$A \longleftarrow \boxed{C} \longrightarrow B$$

$$A \perp\!\!\!\perp_P B \mid C$$

Chaîne: contenant mediateur ou cause intermédiaire

$$A \longrightarrow C \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B$$

$$A \longrightarrow \boxed{C} \longrightarrow B$$

$$A \perp\!\!\!\perp_P B \mid C$$

V-structure: contenant un collider

$$A \longrightarrow C \longleftarrow B$$

$$A \perp\!\!\!\perp_P B$$

$$A \longrightarrow \boxed{C} \longleftarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \mid C$$

Basic structures

Fourchette: contenant un facteur de confusion

$$A \longleftarrow C \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B$$

$$A \longleftarrow \boxed{C} \longrightarrow B$$

$$A \perp\!\!\!\perp_P B \mid C$$

Chaîne: contenant mediateur ou cause intermédiaire

$$A \longrightarrow C \longrightarrow B$$

$$A \overset{?}{\perp\!\!\!\perp}_P B$$

$$A \longrightarrow \boxed{C} \longrightarrow B$$

$$A \perp\!\!\!\perp_P B \mid C$$

V-structure: contenant un collider

$$A \longrightarrow C \longleftarrow B$$

$$A \perp\!\!\!\perp_P B$$

$$\begin{tikzcd} A \arrow[r] & C \arrow[d] & B \arrow[l] \\ & \boxed{D} & \end{tikzcd}$$

$$A \overset{?}{\perp\!\!\!\perp}_P B \mid C$$

Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example

$$A \longrightarrow \boxed{C} \longleftarrow B$$

Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example

$$A \longrightarrow \boxed{C} \longleftarrow B$$

$$A = \begin{cases} \text{Mère porteuse} \\ \text{Mère non porteuse} \end{cases} \qquad B = \begin{cases} \text{Père porteur} \\ \text{Père non porteur} \end{cases}$$

$$C = (A \text{ or } B) = \begin{cases} \text{Enfant porteur} \\ \text{Enfant non porteur} \end{cases}$$

Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example

$$
\begin{tikzcd}
A \arrow[r] \arrow[rr, bend left, dash, dashed] & \boxed{C} & B \arrow[l]
\end{tikzcd}
$$

$$A = \begin{cases} \text{Mère porteuse} \\ \text{Mère non porteuse} \end{cases} \qquad B = \begin{cases} \text{Père porteur} \\ \text{Père non porteur} \end{cases}$$

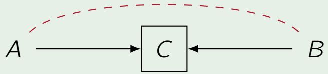
$$C = (A \text{ or } B) = \begin{cases} \text{Enfant porteur} \\ \text{Enfant non porteur} \end{cases}$$

Si C = Enfant porteur $\implies$

$$\begin{cases} \text{Si } A = \text{Mère non porteuse alors } B = \text{Père porteur} \\ \text{Si } B = \text{Père non porteur alors } A = \text{Mère porteuse} \end{cases}$$

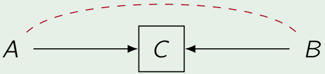
Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example



Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example



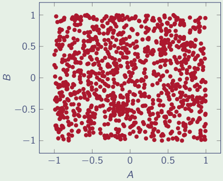
$A, B \sim U(-1, 1)$ $\xi_c \sim N(0, \frac{1}{2})$ $C = 2AB + \xi_c$

Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example

$$A \longrightarrow \boxed{C} \longleftarrow B$$

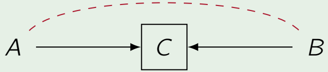
$A, B \sim U(-1, 1)$ $\qquad$ $\xi_c \sim N(0, \frac{1}{2})$ $\qquad$ $C = 2AB + \xi_c$



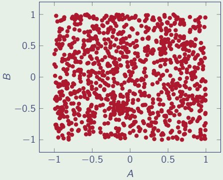
$Corr(A; B) = 0.002$

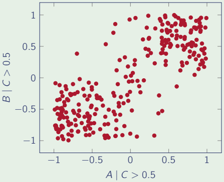
Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example



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





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

Chemins bloqués

Un chemin est dit être bloqué par un ensemble de noeuds $\mathbb{Z} \in \mathbb{V}$ si

- Il contient une chaîne $A \to B \to C$ ou une fourchette $A \leftarrow B \to C$ et $B \in \mathbb{Z}$, ou
- Il contient un collider $A \to B \leftarrow C$ tel que B ou ses descendants ne soient pas dans $\mathbb{Z}$.

Chemins bloqués

Un chemin est dit être bloqué par un ensemble de noeuds $\mathbb{Z} \in \mathbb{V}$ si

- Il contient une chaîne $A \rightarrow B \rightarrow C$ ou une fourchette $A \leftarrow B \rightarrow C$ et $B \in \mathbb{Z}$, ou
- Il contient un collider $A \rightarrow B \leftarrow C$ tel que B ou ses descendants ne soient pas dans $\mathbb{Z}$.

Un chemin qui n'est pas bloqué est dit actif.

Chemins bloqués

Un chemin est dit être bloqué par un ensemble de noeuds $\mathbb{Z} \in \mathbb{V}$ si

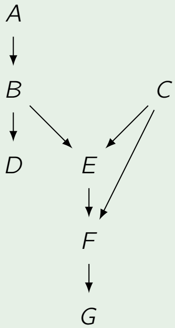
- Il contient une chaîne $A \to B \to C$ ou une fourchette $A \leftarrow B \to C$ et $B \in \mathbb{Z}$, ou
- Il contient un collider $A \to B \leftarrow C$ tel que B ou ses descendants ne soient pas dans $\mathbb{Z}$.

Un chemin qui n'est pas bloqué est dit actif.

Un chemin est actif si chaque triplet le long du chemin est actif, et bloqué si un seul triplet est bloqué.

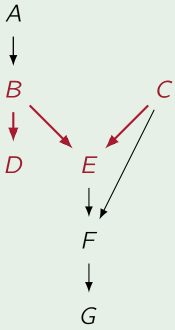
Chemins bloqués: exemples

Example



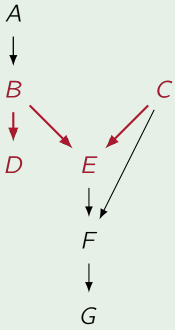
Chemins bloqués: exemples

Example



Est ce que le chemin $< D, B, E, C >$ est bloqué?

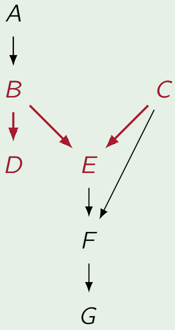
Example



Est ce que le chemin $< D, B, E, C >$ est bloqué?
Oui

Chemins bloqués: exemples

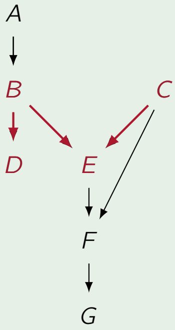
Example



Est ce que le chemin $< D, B, E, C >$ est bloqué?
Oui
Est ce que le chemin $< D, B, E, C >$ est bloqué par E?

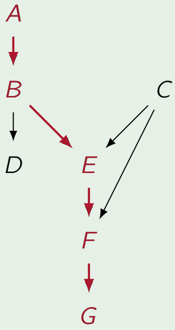
Chemins bloqués: exemples

Example



Est ce que le chemin $< D, B, E, C >$ est bloqué?
Oui
Est ce que le chemin $< D, B, E, C >$ est bloqué par E? Non

Example



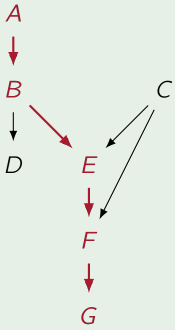
Est ce que le chemin $< D, B, E, C >$ est bloqué? Oui

Est ce que le chemin $< D, B, E, C >$ est bloqué par E? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué?

Chemins bloqués: exemples

Example

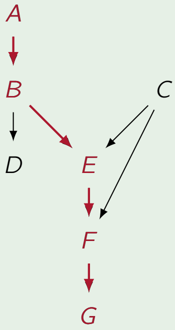


Est ce que le chemin $< D, B, E, C >$ est bloqué? Oui

Est ce que le chemin $< D, B, E, C >$ est bloqué par E? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué? Non

Example



Est ce que le chemin $< D, B, E, C >$ est bloqué? Oui

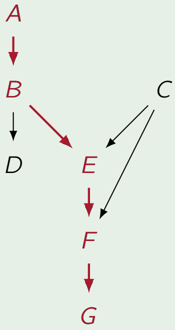
Est ce que le chemin $< D, B, E, C >$ est bloqué par E? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué par E?

Chemins bloqués: exemples

Example



Est ce que le chemin $< D, B, E, C >$ est bloqué? Oui

Est ce que le chemin $< D, B, E, C >$ est bloqué par E? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué? Non

Est ce que le chemin $< A, B, E, F, G >$ est bloqué par E? Oui

d-séparation

Étant donné des ensembles disjoints $\mathbb{A}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, on dit que $\mathbb{A}$ et $\mathbb{Y}$ sont d-séparés par $\mathbb{Z}$ si chaque chemin entre un nœud de $\mathbb{A}$ et un nœud de $\mathbb{Y}$ est bloqué par $\mathbb{Z}$, et on écrit $\mathbb{A} \perp\!\!\!\perp_G \mathbb{Y} \mid \mathbb{Z}$.

d-séparation

Étant donné des ensembles disjoints $\mathbb{A}, \mathbb{Y}, \mathbb{Z} \subseteq \mathbb{V}$, on dit que $\mathbb{A}$ et $\mathbb{Y}$ sont d-séparés par $\mathbb{Z}$ si chaque chemin entre un nœud de $\mathbb{A}$ et un nœud de $\mathbb{Y}$ est bloqué par $\mathbb{Z}$, et on écrit $\mathbb{A} \perp\!\!\!\perp_G \mathbb{Y} \mid \mathbb{Z}$.

Si l'un des chemins ci-dessus n'est pas bloqué, on dit que $\mathbb{A}$ et $\mathbb{Y}$ sont d-connectés par $\mathbb{Z}$.

d-séparation et indépendances conditionnelles

Theorem

(i) Si $\mathbb{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$, alors $\mathbb{A} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ dans toute distribution P compatible avec $\mathcal{G}$.

Preuve dans (Pearl, 1988)

d-séparation et indépendances conditionnelles

Theorem

(i) Si $\mathbb{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$, alors $\mathbb{A} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ dans toute distribution P compatible avec $\mathcal{G}$.

(ii) Si $\mathbb{A} \not\!\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$, alors il existe une distribution P compatible avec $\mathcal{G}$ telle que $\mathbb{A} \not\!\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$.

Preuve dans (Pearl, 1988)

d-séparation et indépendances conditionnelles

Theorem

(i) Si $\mathbb{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$, alors $\mathbb{A} \perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$ dans toute distribution P compatible avec $\mathcal{G}$.

(ii) Si $\mathbb{A} \not\perp\!\!\!\perp_{\mathcal{G}} \mathbb{Y} \mid \mathbb{Z}$, alors il existe une distribution P compatible avec $\mathcal{G}$ telle que $\mathbb{A} \not\perp\!\!\!\perp_{P} \mathbb{Y} \mid \mathbb{Z}$.

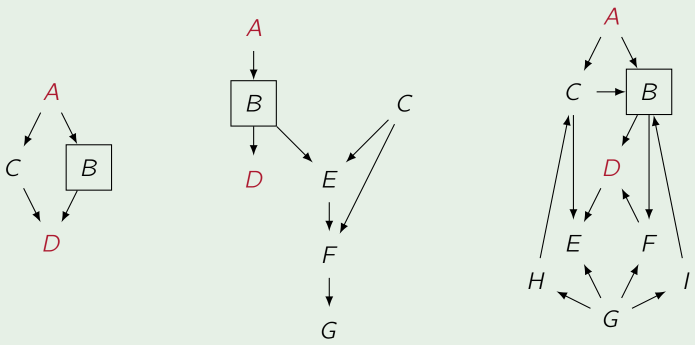
Preuve dans (Pearl, 1988)

(i) $\implies$ La d-séparation caractérise les indépendances conditionnelles des distributions compatibles avec un DAG donné.

(ii) $\implies$ Lors de la construction d'un DAG causal, si vous hésitez à inclure $A \to Y$ dans le DAG, ajoutez-le (tant que le graphe reste acyclique) !

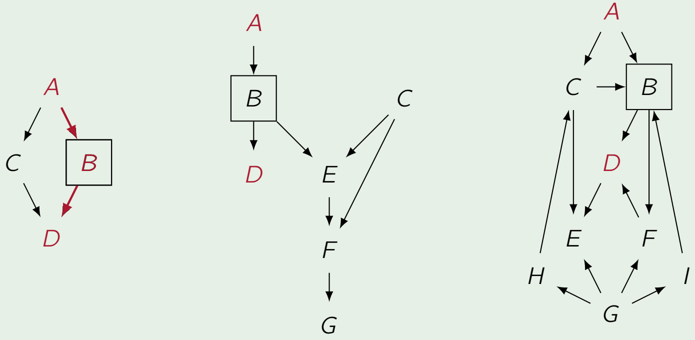
Example

$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



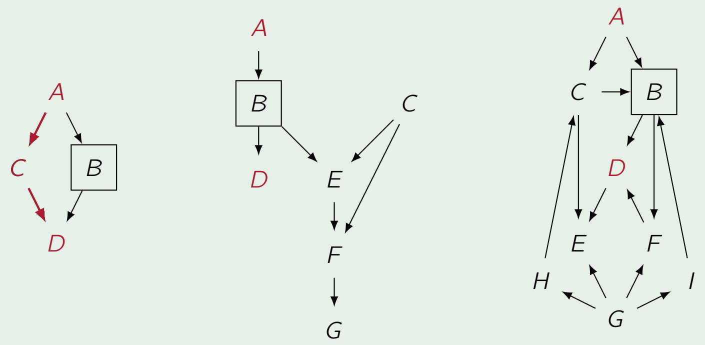
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Example

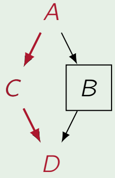
$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$



Lire les indépendances conditionnelles à l'aide de la d-séparation

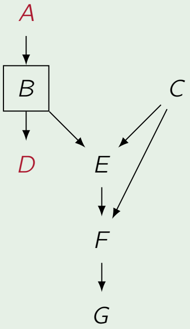
Example

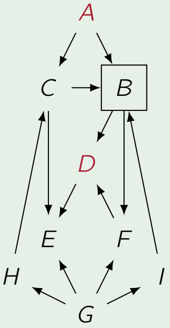
$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



$< A, C, D >$ n'est pas bloqué

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$

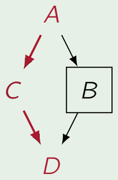


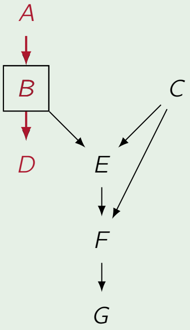


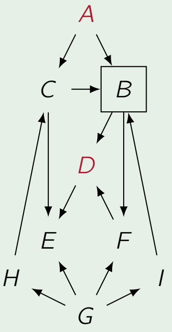
Lire les indépendances conditionnelles à l'aide de la d-séparation

Example

$$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$







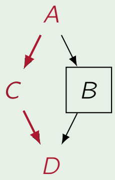
$< A, C, D >$ n'est pas bloqué

$$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$$

Lire les indépendances conditionnelles à l'aide de la d-séparation

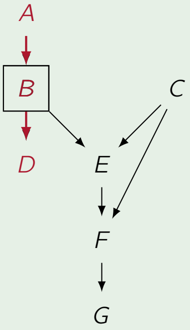
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



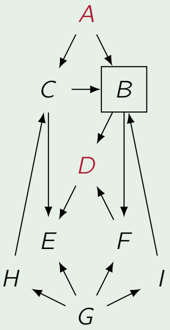
$< A, C, D >$ n'est pas bloqué

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



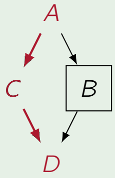
Tous les chemins sont bloqués

$\implies A \perp\!\!\!\perp_P D \mid B$



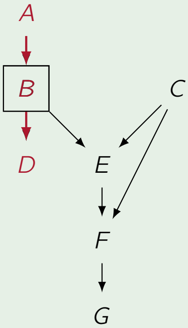
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



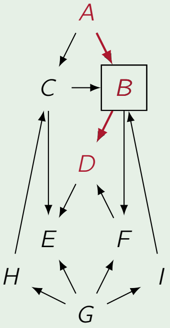
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

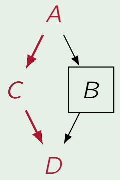
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

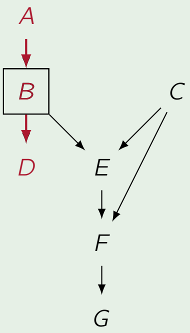
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



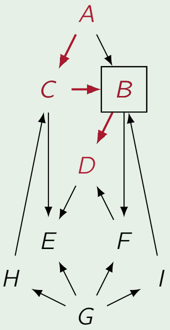
$< A, C, D >$ n'est pas bloqué

$\implies A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

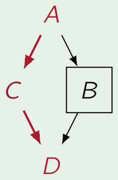
$\implies A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

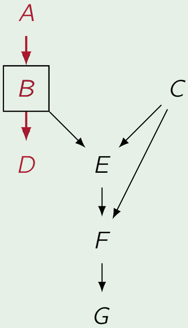
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



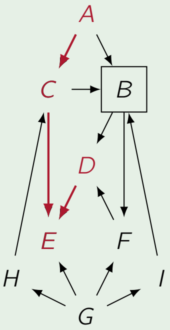
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

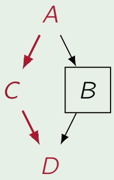
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

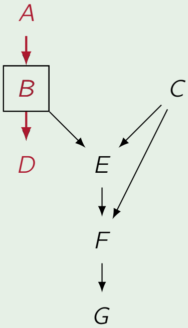
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



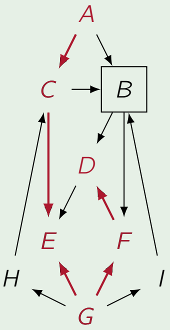
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow \; A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

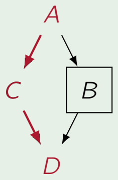
$\Longrightarrow \; A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

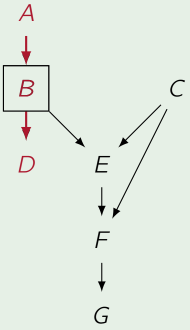
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



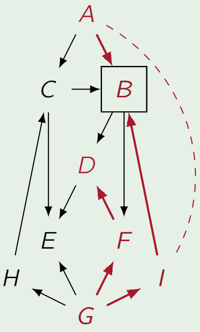
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

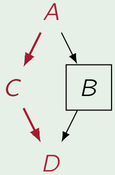
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

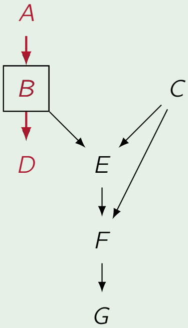
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



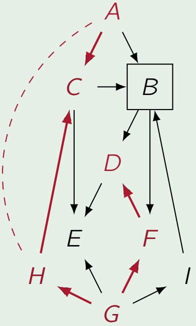
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

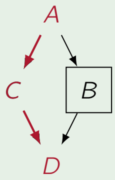
$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



Lire les indépendances conditionnelles à l'aide de la d-séparation

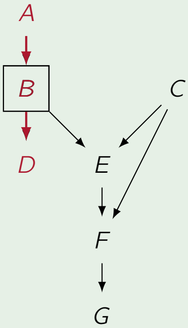
Example

$A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



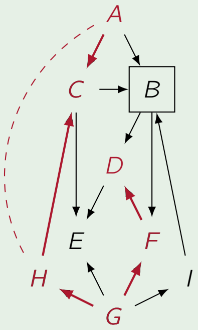
$< A, C, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$



Tous les chemins sont bloqués

$\Longrightarrow A \perp\!\!\!\perp_P D \mid B$



$< A, B, I, G, F, D >$ n'est pas bloqué

$\Longrightarrow A \overset{?}{\perp\!\!\!\perp}_P D \mid B$

Code: d-seperation using DAGitty package

```
#Load the DAGitty package
library(dagitty)
#Define the DAG
dag <- dagitty("Z->X␣Z->Y")
#Are X and Y d-separated given Z?
res <- dseparated(dag, "X", "Y", "Z")
print(res)
```

Résumé de la cette section

- La d-séparation est un outil simple permettant d'identifier toutes les indépendances conditionnelles représentées par un DAG.
- Lors de la construction d'un DAG causal, si vous hésitez à inclure $A \rightarrow Y$ dans le DAG, ajoutez-le (tant que le graphe reste acyclique) !
  - Dans les DAGs, toutes les informations se trouvent dans l'absence d'arêtes et dans la direction des flèches, plutôt que dans la simple présence des arêtes.
- Le package DAGitty permet de tester la d-séparation dans R.

4

Identification d’un effet causal

- Identification
- Direct causes adjustment
- The back-door criterion

4

Identification d'un effet causal

Identification

Direct causes adjustment

The back-door criterion

Rappel:

Un effet causal de $\mathbb{A}$ sur $\mathbb{Y}$

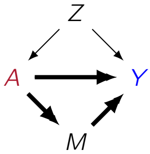
$$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$$

Identification d'un effet causal

Rappel:

Un effet causal de $\mathbb{A}$ sur $\mathbb{Y}$

$$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$$



Identification d'un effet causal

Rappel:

Un effet causal de $\mathbb{A}$ sur $\mathbb{Y}$
$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$

$$\begin{array}{ccc} & Z & \\ \swarrow & & \searrow \\ A & \longrightarrow & Y \\ \searrow & & \nearrow \\ & M & \end{array}$$

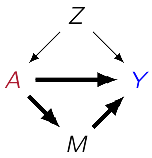
L'effet causal de $\mathbb{A}$ sur $\mathbb{Y}$ est dit identifiable à partir de P dans $\mathcal{G}$ s'il peut être calculé de manière unique à partir de $P(\mathbb{V})$.

Autrement dit, si $P(\mathbb{Y} = \mathbb{y} \mid do(\mathbb{A} = \mathbb{a}))$ peut être calculé de manière unique à partir de $P(\mathbb{V})$.

Rappel:

Un effet causal de $\mathbb{A}$ sur $\mathbb{Y}$

$= \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a})) - \mathbb{E}(\mathbb{Y} \mid do(\mathbb{A} = \mathbb{a}'))$



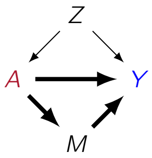
L'effet causal de $\mathbb{A}$ sur $\mathbb{Y}$ est dit identifiable à partir de P dans $\mathcal{G}$ s'il peut être calculé de manière unique à partir de $P(\mathbb{V})$.

Autrement dit, si $P(\mathbb{Y} = \mathbb{y} \mid do(\mathbb{A} = \mathbb{a}))$ peut être calculé de manière unique à partir de $P(\mathbb{V})$.

Dans ce qui suit, nous allons simplifier les notations en utilisant $P(\mathbb{y} \mid do(\mathbb{a}))$ au lieu de $P(\mathbb{Y} = \mathbb{y} \mid do(\mathbb{A} = \mathbb{a}))$.

Inférence causale : principaux défis (1/3)

Biais liés à des facteurs de confusion:



Biais liés à des facteurs de confusion:

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ A \arrow[rr] \arrow[dr] & & Y \\ & M \arrow[ur] & \end{tikzcd}$$

Nous devons éliminer tous les biais de confusion en ajustant sur les facteurs de confusion.

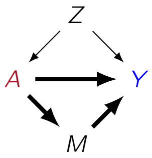
Biais liés à des facteurs de confusion:

$$
\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
A \arrow[rr] \arrow[dr] & & Y \\
 & M \arrow[ur] &
\end{tikzcd}
$$

Nous devons éliminer tous les biais de confusion en ajustant sur les facteurs de confusion.
Devons-nous toujours ajuster sur toutes les variables disponibles?

Inférence causale : principaux défis (1/3)

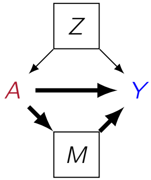
Biais liés à des facteurs de confusion:



Nous devons éliminer tous les biais de confusion en ajustant sur les facteurs de confusion.
Devons-nous toujours ajuster sur toutes les variables disponibles? Non!

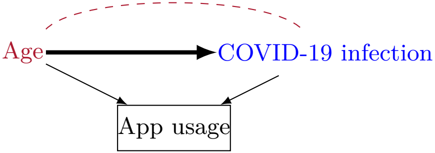
Inférence causale : principaux défis (2/3)

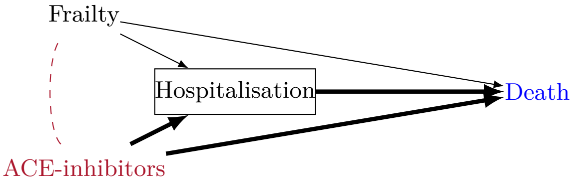
Biais liés à un ajustement incorrect sur des médiateurs:



Inférence causale : principaux défis (3/3)

Biais liés à un ajustement sur des colliders:



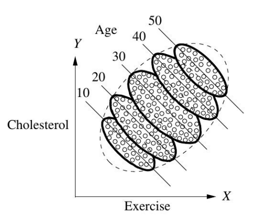


4

Identification d'un effet causal

Example

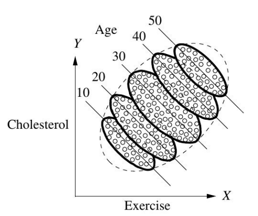
Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol $P(c \mid do(e))$?

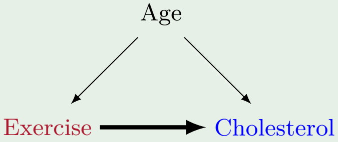


Paradoxe de Simpson 1 avec une solution simple

Example

Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol $P(c \mid do(e))$?

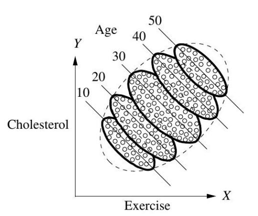




Paradoxe de Simpson 1 avec une solution simple

Example

Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol $P(c \mid do(e))$?



$$\begin{tikzcd} & \text{Age} \arrow[dl] \arrow[dr] & \\ \text{Exercise} \arrow[rr] & & \text{Cholesterol} \end{tikzcd}$$

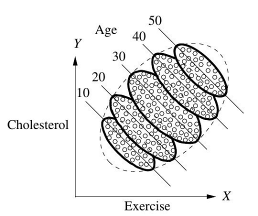
$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

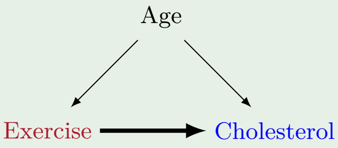
$$P(a, c \mid do(e)) = P(a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

Paradoxe de Simpson 1 avec une solution simple

Example

Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol $P(c \mid do(e))$?





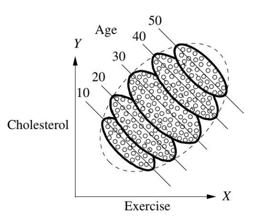
$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

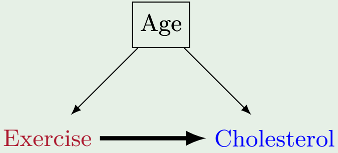
$$P(a, c \mid do(e)) = P(a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

$$P(c \mid do(e)) = \sum_a P(a)P(c \mid a, e) \qquad \text{(Marginalisation)}$$

Example

Dans une étude, nous mesurons l'exercice hebdomadaire et les niveaux de cholestérol pour différents groupes d'âge. Quel est l'effet causal de l'exercice sur le cholestérol $P(c \mid do(e))$?





$$P(a, e, c) = P(a)P(e \mid a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

$$P(a, c \mid do(e)) = P(a)P(c \mid a, e) \qquad \text{(Compatibilité)}$$

$$P(c \mid do(e)) = \sum_a P(a)P(c \mid a, e) \qquad \text{(Marginalisation)}$$

Une solution simple

Theorem

Étant donné un DAG causal $\mathcal{G}$ dans lequel un sous-ensemble $\mathbb{V}$ de variables est mesuré, $P(y \mid do(x))$ est identifiable lorsque $\{X \cup Y \cup Pa(X)\} \subseteq \mathbb{V}$, et est donné par:

$$P(y \mid do(x)) = \sum_{Z \in Pa(X)} P(y \mid x, z) P(z)$$

Preuve dans (Pearl, 2000)

Limitations of the direct causes adjustment

- Parfois; l'ensemble des parents est trop grand.

- Parfois, une partie des parents n'est pas mesurée.

4

Identification d'un effet causal

Identification

Direct causes adjustment

The back-door criterion

Le critère du back-door

Un ensemble de variables $\mathbb{Z}$ satisfait le critère du back-door par rapport à une paire ordonnée de variables (A, Y) dans un DAG $\mathcal{G}$ si :

- Aucun nœud de $\mathbb{Z}$ n'est un descendant de A; et
- $\mathbb{Z}$ bloque tous les chemins entre A et Y qui contiennent une flèche pointant vers A.

Le critère du back-door

Un ensemble de variables $\mathbb{Z}$ satisfait le critère du back-door par rapport à une paire ordonnée de variables (A, Y) dans un DAG $\mathcal{G}$ si :

- Aucun nœud de $\mathbb{Z}$ n'est un descendant de A; et
- $\mathbb{Z}$ bloque tous les chemins entre A et Y qui contiennent une flèche pointant vers A.

Peut être étendu à un ensemble $\mathbb{A}$ et un ensemble $\mathbb{Y}$.

Le critère du back-door: intuition

- Pourquoi aucun noeud de $\mathbb{Z}$ ne doit être un descendant de A?

- Pourquoi $\mathbb{Z}$ doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A?

Le critère du back-door: intuition

- Pourquoi aucun noeud de $\mathbb{Z}$ ne doit être un descendant de A?

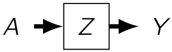
$$A \rightarrow \boxed{Z} \rightarrow Y$$

Pour éviter de bloquer sur des causes intermediaires.

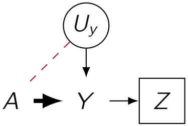
- Pourquoi $\mathbb{Z}$ doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A?

Le critère du back-door: intuition

- Pourquoi aucun noeud de $\mathbb{Z}$ ne doit être un descendant de A?



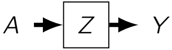
Pour éviter de bloquer sur des causes intermediaires.



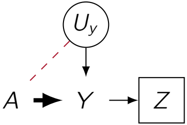
Pour éviter d'introduire un biais de confusion artificiel en conditionnant sur un collider.

- Pourquoi $\mathbb{Z}$ doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A?

- Pourquoi aucun noeud de $\mathbb{Z}$ ne doit être un descendant de A?



Pour éviter de bloquer sur des causes intermediaires.

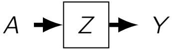


Pour éviter d'introduire un biais de confusion artificiel en conditionnant sur un collider. Pour éviter les biais de sélection!

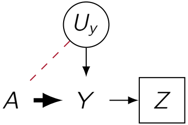
- Pourquoi $\mathbb{Z}$ doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A?

Le critère du back-door: intuition

- Pourquoi aucun noeud de $\mathbb{Z}$ ne doit être un descendant de A?



Pour éviter de bloquer sur des causes intermediaires.



Pour éviter d'introduire un biais de confusion artificiel en conditionnant sur un collider. Pour éviter les biais de sélection!

- Pourquoi $\mathbb{Z}$ doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A?

Pour éliminer les biais provenant des facteurs de confusions.

Le critère du back-door: en utilisant la d-séparation

Un ensemble de variables $\mathbb{Z}$ satisfait le critère du back-door par rapport à une paire ordonnée de variables (A, Y) dans un DAG $\mathcal{G}$ si :

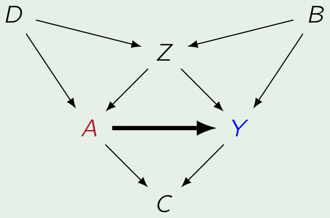
- $A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid \mathbb{Z}$,

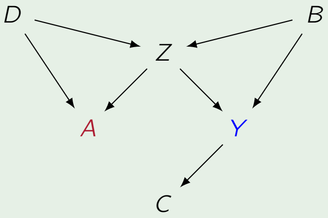
Tel que $\mathcal{G}_{\underline{A}}$ soit un graphe identique à $\mathcal{G}$, à l'exception des flèches sortantes de A qui sont supprimées.

Le critère du back-door: exemple

Example

$\mathcal{G}$ $\quad$ $P(y \mid do(a))$ $\quad$ $\mathcal{G}_{\underline{A}}$

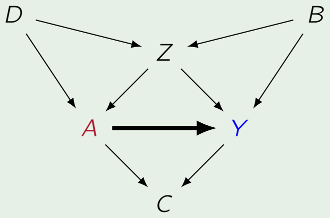


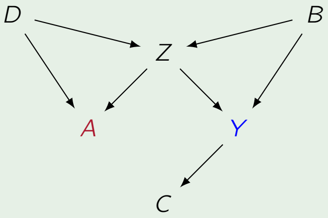


Le critère du back-door: exemple

Example

$\mathcal{G}$ $\qquad$ $P(y \mid do(a))$ $\qquad$ $\mathcal{G}_{\underline{A}}$

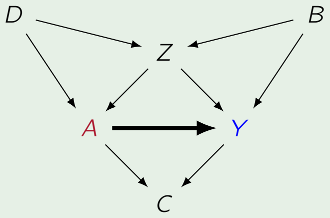


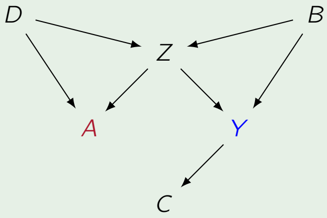


$$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z \ ?$$

Example

$\mathcal{G}$ $\quad$ $P(y \mid do(a))$ $\quad$ $\mathcal{G}_{\underline{A}}$



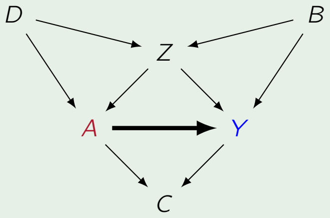


$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

Le critère du back-door: exemple

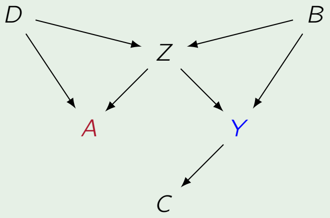
Example

$\mathcal{G}$



$P(y \mid do(a))$

$\mathcal{G}_{\underline{A}}$



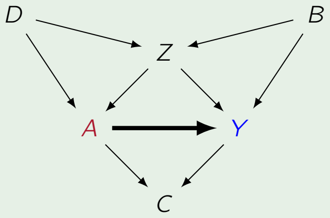
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

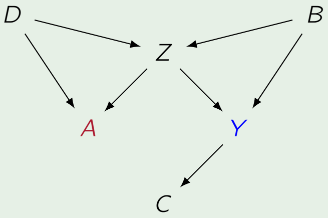
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$?

Le critère du back-door: exemple

Example

$\mathcal{G}$ $\quad$ $P(y \mid do(a))$ $\quad$ $\mathcal{G}_{\underline{A}}$



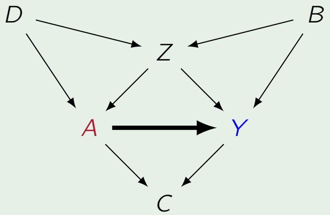


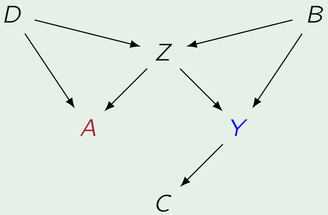
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

Example

$\mathcal{G}$ $\quad\quad P(y \mid do(a)) \quad\quad \mathcal{G}_{\underline{A}}$





$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

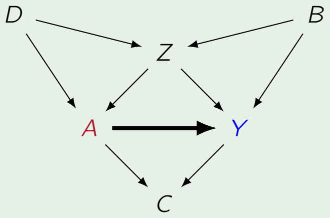
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$?

Le critère du back-door: exemple

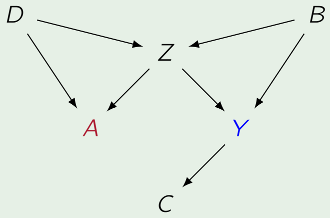
Example

$\mathcal{G}$



$P(y \mid do(a))$

$\mathcal{G}_{\underline{A}}$



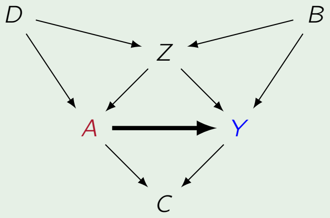
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

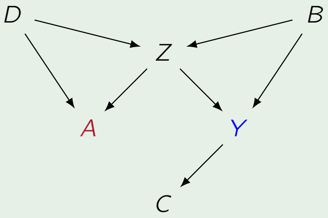
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

Example

$\mathcal{G}$ $\quad P(y \mid do(a))$ $\quad \mathcal{G}_{\underline{A}}$





$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

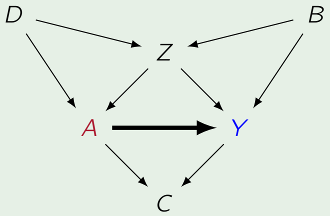
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$?

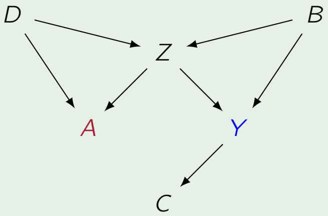
Example

$\mathcal{G}$



$P(y \mid do(a))$

$\mathcal{G}_{\underline{A}}$



$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

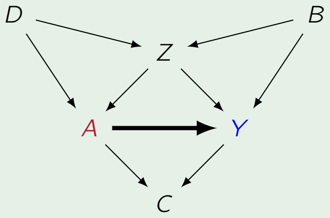
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

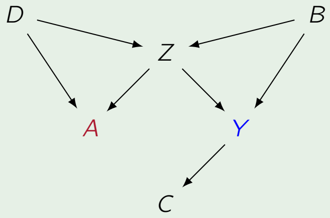
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$? Non

Le critère du back-door: exemple

Example

$\mathcal{G}$ $\quad\quad\quad\quad$ $P(y \mid do(a))$ $\quad\quad$ $\mathcal{G}_{\underline{A}}$





$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

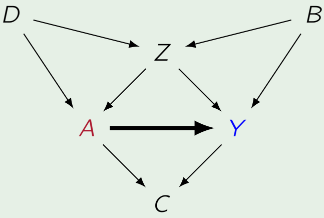
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

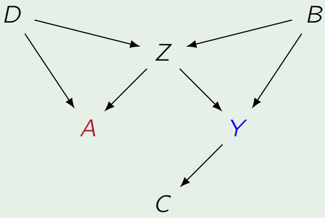
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D, B$?

Example

$\mathcal{G}$ $\qquad\qquad$ $P(y \mid do(a))$ $\qquad$ $\mathcal{G}_{\underline{A}}$





$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

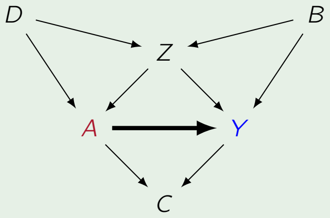
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

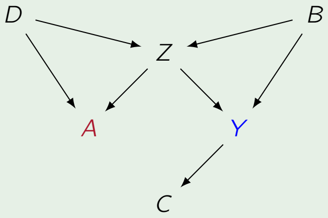
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D, B$? Non

Example

$\mathcal{G}$ $\quad P(y \mid do(a))$ $\quad \mathcal{G}_{\underline{A}}$





$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

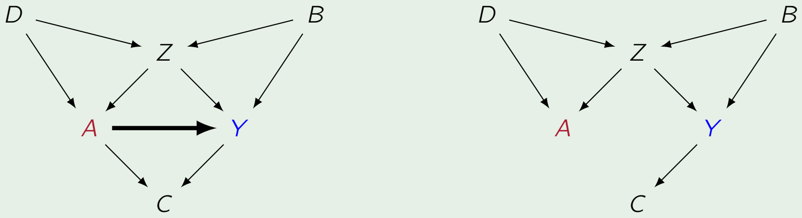
$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D, B$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z, D$?

Example

$\mathcal{G}$ $\quad P(y \mid do(a))$ $\quad \mathcal{G}_{\underline{A}}$



$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid B$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid C$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid D, B$? Non

$A \perp\!\!\!\perp_{\mathcal{G}_{\underline{A}}} Y \mid Z, D$? Oui

Theorem

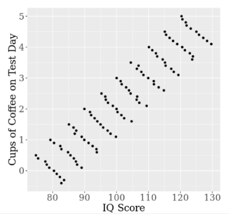
Si $\mathbb{Z}$ satisfait le critère du back-door par rapport à (A, Y) et si $P(a, z) > 0$, alors l'effet causal de A sur Y est identifiable et est donné par

$$P(y \mid do(a)) = \sum_z P(y \mid a, z) P(z).$$

Preuve dans (Pearl, 2000)

Example

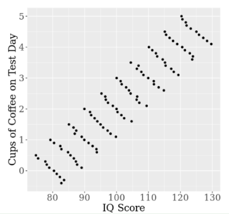
Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.

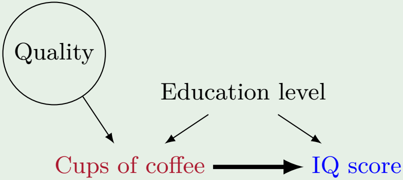


Quel est l'effet du nombre de tasses de café sur le score de QI, $P(i \mid do(c))$?

Example

Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.



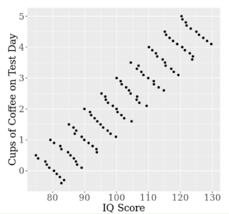


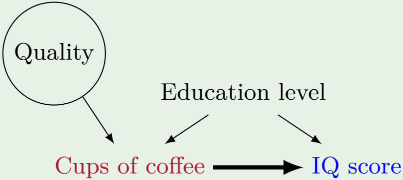
Quel est l'effet du nombre de tasses de café sur le score de QI, $P(i \mid do(c))$?

Le critère du back-door en action

Example

Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.





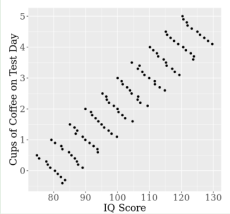
Quel est l'effet du nombre de tasses de café sur le score de QI, $P(i \mid do(c))$?

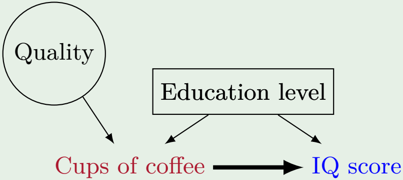
$$P(i \mid do(c)) = \sum_{e} P(i \mid c, e) P(e)$$

Le critère du back-door en action

Example

Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.





Quel est l'effet du nombre de tasses de café sur le score de QI, $P(i \mid do(c))$?

$$P(i \mid do(c)) = \sum_e P(i \mid c, e)P(e)$$

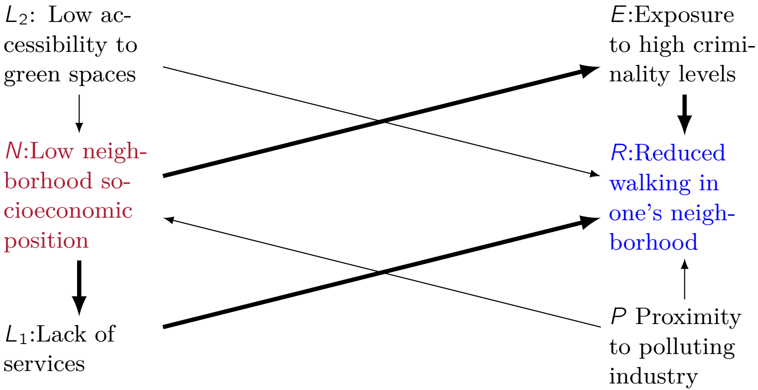
```r
#Load the DAGitty package
library(dagitty)
#Define the DAG
dag <- dagitty("Z->X Z->Y X->Y")
#Sets that satisfy the back-door criterion for (X, Y)?
res <- adjustmentSets(dag, exposure="X", outcome="Y")
print(res)
```

Une application réelle où le critère du back-door est appliqué

Dans cette étude, nous cherchons à estimer l'effet de la position socioéconomique du quartier (N) sur la réduction de la marche dans son propre quartier (R). En d'autres termes, $P(r \mid do(n))$?

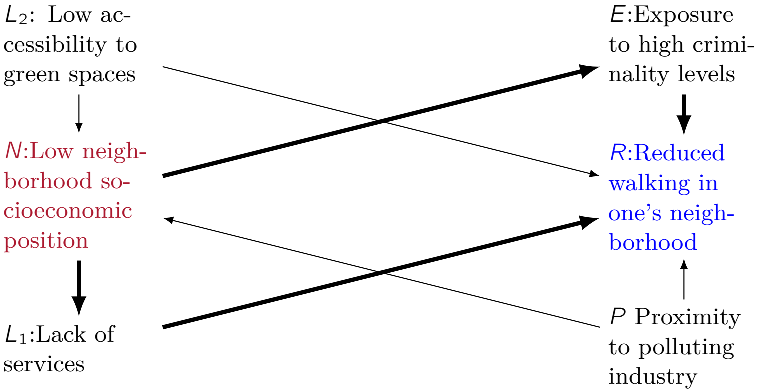
Une application réelle où le critère du back-door est appliqué

Dans cette étude, nous cherchons à estimer l'effet de la position socioéconomique du quartier (N) sur la réduction de la marche dans son propre quartier (R). En d'autres termes, $P(r \mid do(n))$?



Une application réelle où le critère du back-door est appliqué

Dans cette étude, nous cherchons à estimer l'effet de la position socioéconomique du quartier (N) sur la réduction de la marche dans son propre quartier (R). En d'autres termes, $P(r \mid do(n))$?



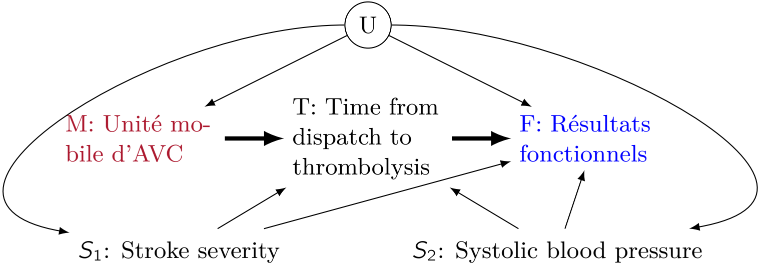
$$P(r \mid do(n)) = \sum_{l_2, p} P(r \mid n, l_2, p) P(l_2, p)$$

Une application réelle où le critère du back-door ne peut pas être appliqué

Dans cette étude, nous cherchons à estimer l'effet de l'envoi d'une unité mobile d'AVC (M) sur les résultats fonctionnels (F). En d'autres termes, $P(f \mid do(m))$?

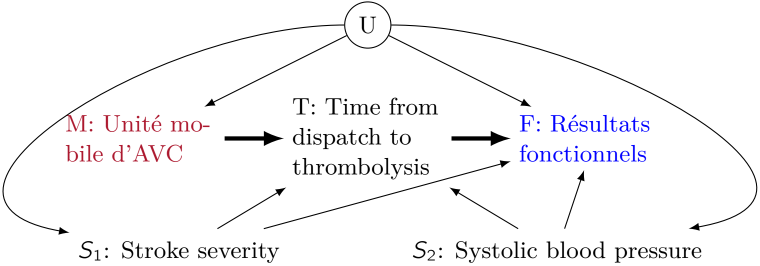
Une application réelle où le critère du back-door ne peut pas être appliqué

Dans cette étude, nous cherchons à estimer l'effet de l'envoi d'une unité mobile d'AVC (M) sur les résultats fonctionnels (F). En d'autres termes, $P(f \mid do(m))$?



Une application réelle où le critère du back-door ne peut pas être appliqué

Dans cette étude, nous cherchons à estimer l'effet de l'envoi d'une unité mobile d'AVC (M) sur les résultats fonctionnels (F). En d'autres termes, $P(f \mid do(m))$?



$$P(f \mid do(m)) = \sum_{t,s_1,s_2} P(t|m,s_1,s_2) \sum_{m'} P(f \mid t,m',s_1,s_2) P(m',s_1,s_2)$$

Le critère du back-door n'est pas complet

- Le critère du back-door n'est pas complet:

Le critère du back-door n'est pas complet

- Le critère du back-door n'est pas complet:
  - Si aucun ensemble ne satisfait le critère du back-door pour $P(y \mid do(x))$, cela ne signifie pas nécessairement que $P(y \mid do(x))$ n'est pas identifiable.
  - Il existe d'autres critères, tels que le critère du front-door et le do-calculus (complet).

Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

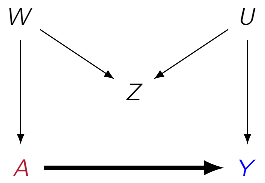
- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Contre-exemple

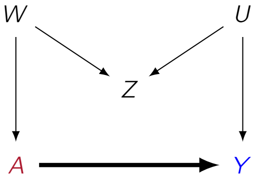


Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Contre-exemple



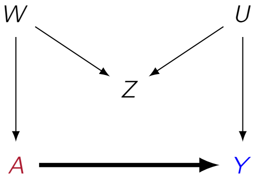
Z est associé à A et Z est associé à Y, conditionnellement à A

Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Contre-exemple



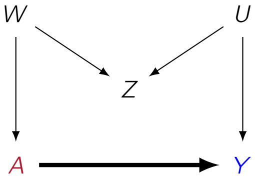
Z est associé à A et Z est associé à Y, conditionnellement à A
Faut-il ajuster sur Z?

Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Contre-exemple



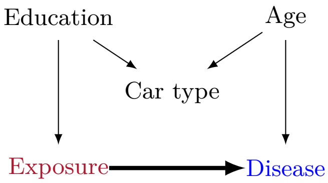
Z est associé à A et Z est associé à Y, conditionnellement à A
Faut-il ajuster sur Z? Non!

Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables observées dans un problème qui ne sont pas affectées par A. $\mathbb{Z}$ satisfait le critère d'association 1 si chaque élément Z de $\mathbb{Z}$ satisfait les conditions suivantes :

- Z est associé à A ; et
- Z est associé à Y, conditionnellement à A.

Contre-exemple



Z est associé à A et Z est associé à Y, conditionnellement à A

Faut-il ajuster sur Z?

Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables dans un problème qui ne sont pas affectées par X. $\mathbb{Z}$ satisfait le critère d'association 2 si cet ensemble peut être divisé en deux sous-ensembles $\mathbb{Z}_1$ et $\mathbb{Z}_2$, lesquels satisfont les conditions suivantes :

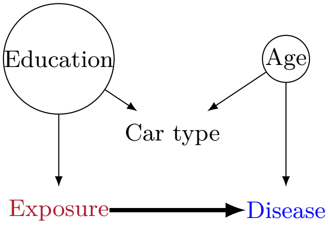
- $\mathbb{Z}_1$ est associé à A ; et
- $\mathbb{Z}_2$ est associé à Y, conditionnellement à A et $\mathbb{Z}_1$.

Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables dans un problème qui ne sont pas affectées par X. $\mathbb{Z}$ satisfait le critère d'association 2 si cet ensemble peut être divisé en deux sous-ensembles $\mathbb{Z}_1$ et $\mathbb{Z}_2$, lesquels satisfont les conditions suivantes :

- $\mathbb{Z}_1$ est associé à A ; et
- $\mathbb{Z}_2$ est associé à Y, conditionnellement à A et $\mathbb{Z}_1$.

Counter-example

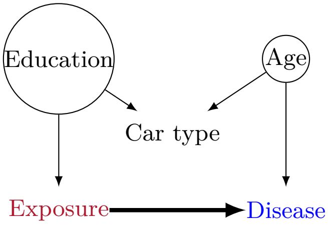


Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables dans un problème qui ne sont pas affectées par X. $\mathbb{Z}$ satisfait le critère d'association 2 si cet ensemble peut être divisé en deux sous-ensembles $\mathbb{Z}_1$ et $\mathbb{Z}_2$, lesquels satisfont les conditions suivantes :

- $\mathbb{Z}_1$ est associé à A ; et
- $\mathbb{Z}_2$ est associé à Y, conditionnellement à A et $\mathbb{Z}_1$.

Counter-example



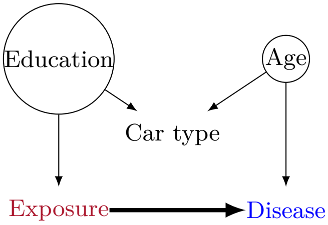
Car type est associé à Exposure

Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables dans un problème qui ne sont pas affectées par X. $\mathbb{Z}$ satisfait le critère d'association 2 si cet ensemble peut être divisé en deux sous-ensembles $\mathbb{Z}_1$ et $\mathbb{Z}_2$, lesquels satisfont les conditions suivantes :

- $\mathbb{Z}_1$ est associé à A ; et
- $\mathbb{Z}_2$ est associé à Y, conditionnellement à A et $\mathbb{Z}_1$.

Counter-example



Car type est associé à Exposure

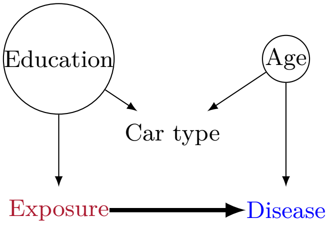
Faut-il adjuster sur Car type?

Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

Soit $\mathbb{Z}$ l'ensemble des variables dans un problème qui ne sont pas affectées par X. $\mathbb{Z}$ satisfait le critère d'association 2 si cet ensemble peut être divisé en deux sous-ensembles $\mathbb{Z}_1$ et $\mathbb{Z}_2$, lesquels satisfont les conditions suivantes :

- $\mathbb{Z}_1$ est associé à A ; et
- $\mathbb{Z}_2$ est associé à Y, conditionnellement à A et $\mathbb{Z}_1$.

Counter-example



Car type est associé à Exposure

Faut-il adjuster sur Car type? Non!

Résumé de cette section

- Le critère de back-door est un outil simple pour identifier les effets causaux.
- Si le critère de la back-door n'est pas satisfait, ce n'est pas la fin du monde !
- Les critères basés uniquement sur les associations sont incorrects !
- Le package DAGitty permet de tester le critère du back-door dans R.

5

Exercices

Exercice 1

Une étude de santé publique vise à évaluer l'efficacité d'un vaccin pour réduire le risque d'une maladie spécifique. Les chercheurs comparent le taux de maladie chez les individus vaccinés et non vaccinés.
Lorsqu'on regarde séparément les groupes d'âge (par exemple, enfants et adultes), il apparaît que les personnes vaccinées ont systématiquement moins de cas de maladie que celles non vaccinées. Cependant, lorsque les données des deux groupes sont combinées, on observe une situation inattendue : le taux de maladie global semble plus faible chez les personnes non vaccinées que chez les vaccinées.

À partir des données observées, pouvez-vous conclure si le vaccin est efficace ? Justifiez votre réponse.

Exercice 2

Supposons que le DAG suivant est non causal.

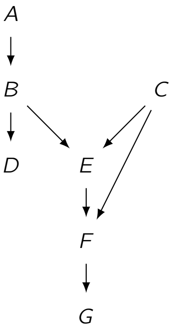
1. écrire la probabilité jointe $P(a, b, c, d)$
2. écrire la probabilité conditionelle $P(b, c, d \mid a)$
3. écrire la probabilité interventionelle $P(b, c, d \mid do(a))$

Supposons que le DAG suivant est causal.

1. écrire la probabilité jointe $P(a, b, c, d)$
2. écrire la probabilité conditionelle $P(b, c, d \mid a)$
3. écrire la probabilité interventionelle $P(b, c, d \mid do(a))$

$$A \rightarrow B \rightarrow C \rightarrow D$$

Exercice 3



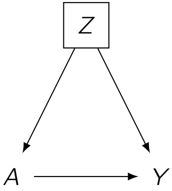
- $B \perp\!\!\!\perp_P G \mid F$?
- $A \perp\!\!\!\perp_P F \mid E$?
- $A \perp\!\!\!\perp_P F \mid C, E$?
- $B \perp\!\!\!\perp_P E \mid F$?
- $\{A, B, D\} \perp\!\!\!\perp_P C \mid G$?
- $\{A, B, D\} \perp\!\!\!\perp_P \{C, F, G\}$?
- $\{A, B, D\} \perp\!\!\!\perp_P \{C, F, G\} \mid E$?

Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?

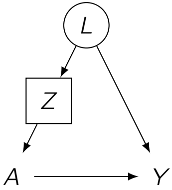
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



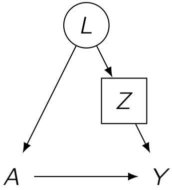
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



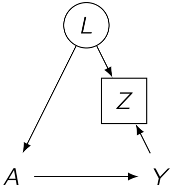
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



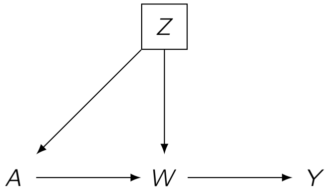
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



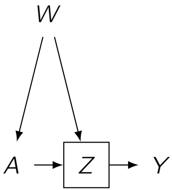
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?

$$A \rightarrow \boxed{Z} \rightarrow Y$$

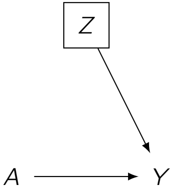
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



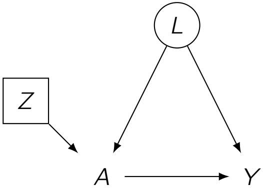
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?

$$\begin{tikzcd} & \boxed{Z} \arrow[dl] & \\ A \arrow[rr] & & Y \end{tikzcd}$$

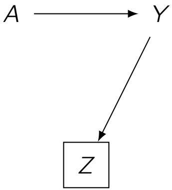
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



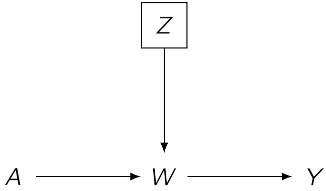
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?



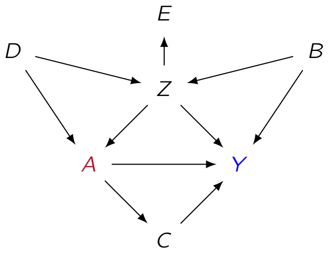
Exercise 4

Est ce que Z satisfait le critère du back-door pour (A, Y)?

$$\begin{tikzcd} A \arrow[r] & W \arrow[r] \arrow[d] & Y \\ & \boxed{Z} & \end{tikzcd}$$

Exercice 5

Considérez le DAG suivant. Listez tous les ensembles de variables qui satisfont au critère du back-door pour $P(y \mid do(a))$.



À partir du papier disponible ici :
https://pmc.ncbi.nlm.nih.gov/articles/PMC9148671/
et en supposant que les variables non mesurées (représentées par U) deviennent mesurées, répondez aux questions suivantes :

1. Dessinez un DAG causal détaillé correspondant en utilisant DAGitty sur R.
2. À l'aide de DAGitty, identifiez un ensemble de variables qui d-séparent "codeine" de U. Quel est cet ensemble?
3. Utilisez DAGitty pour trouver un ensemble de variables satisfaisant le critère du back-door pour l'effet de "Chronic pain" sur "mortality". Quel est cet ensemble?
4. En supposant une distribution positive, l'effet causal de "Chronic pain" sur "mortality" est-il identifiable? Justifiez votre réponse.

6

References

References

- Pearl J. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann. 1988.
- Pearl J. Causality: Models, Reasoning and Inference. Cambridge University Press. 2009.
- Griffith G, Morris T, Tudball M, Herbert A, Mancano G, Pike L, Sharp G, Sterne J, Palmer T, Smith GD, Tilling K, Zuccolo L, Davies NM, and Hemani G. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nature Communications. 2020.
- Chaix B, Leal C, and Evans D. Neighborhood-level Con- founding in Epidemiologic Studies Unavoidable Challenges, Uncertain Solutions. Epidemiology (Cambridge, Mass.) 2009.
- Piccininni M, Kurth T, Audebert HJ, Rohmann JL. The Effect of Mobile Stroke Unit Care on Functional Outcomes: An Application of the Front-door Formula. Epidemiology. 2023.
- Inoue K, Ritz B, Arah OA. Causal Effect of Chronic Pain on Mortality Through Opioid Prescriptions: Application of the Front-Door Formula. Epidemiology. 2022.

MERCI POUR VOTRE ATTENTION

SANTÉ
SORBONNE
UNIVERSITÉ