

DAG et identification

Charles Assaad charles.assaad@inserm.fr

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

Decembre 2025



- 1 Motivation and histoire
- 2 Préliminaires
- 3 DAG causal
 - Définition
 - d-separation
- 4 Identification d'un effet causal
 - Identification
 - Direct causes adjustment
 - The back-door criterion
 - L'ensemble back-door optimal
- 5 Exercices

1

Motivation and histoire

L'effet causal de A sur Y

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

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

L'effet causal de A sur Y

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

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

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

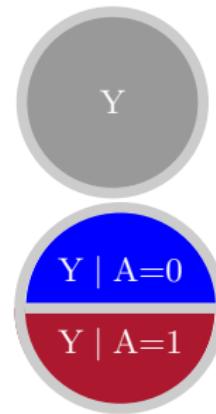
Conditionnement Vs Intervention

Population



Population

Sous-populations

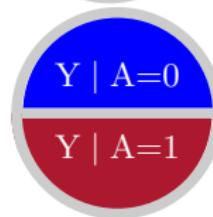


Conditionnement Vs Intervention

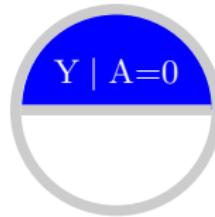
Population



Sous-populations



Conditionnement

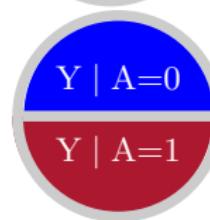


Conditionnement Vs Intervention

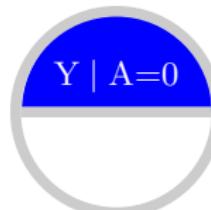
Population



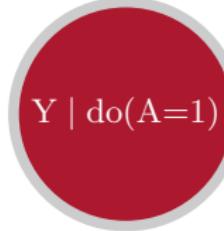
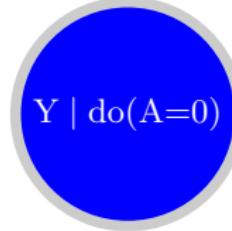
Sous-populations

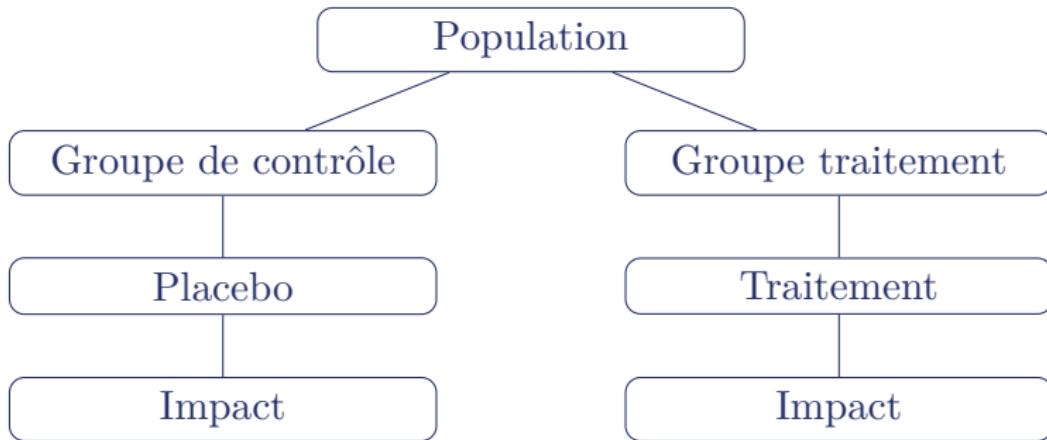


Conditionnement



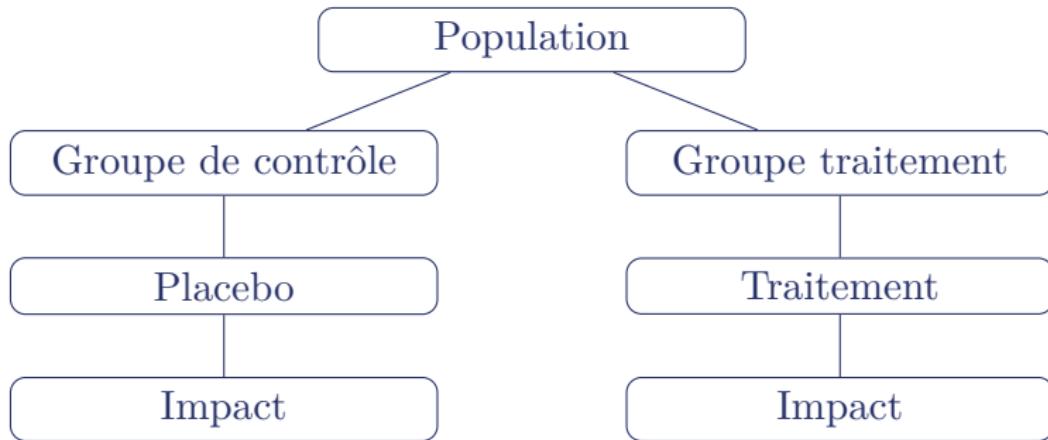
Intervention





Contrôle:

- Permet d'éliminer les biais temporels

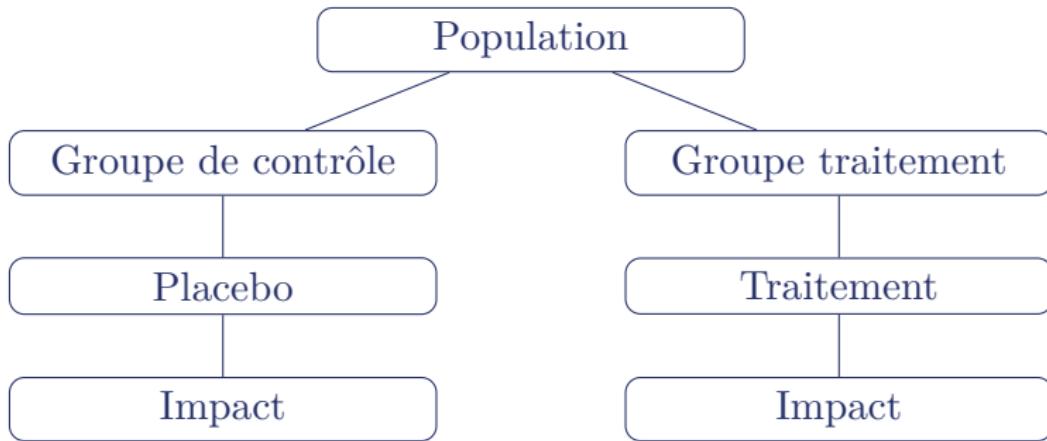


Contrôle:

- Permet d'éliminer les biais temporels

Sélection aléatoire:

- Permet d'éliminer les biais des facteurs extérieurs



Contrôle:

- Permet d'éliminer les biais temporels

Sélection aléatoire:

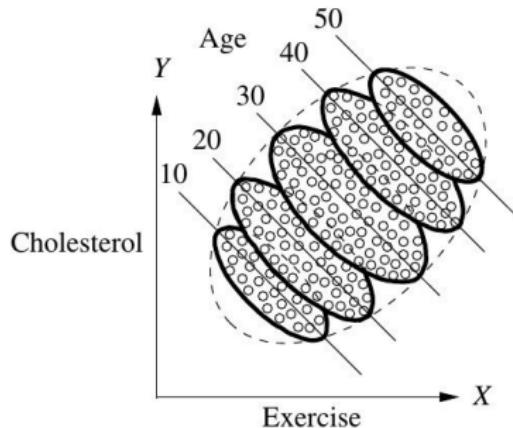
- Permet d'éliminer les biais des facteurs extérieurs

Limitations:

- Coûteux
- Non éthique
- Infaisable

Paradoxe de Simpson

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 ?

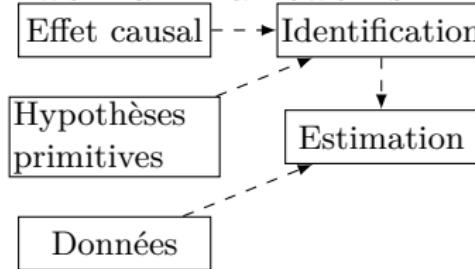






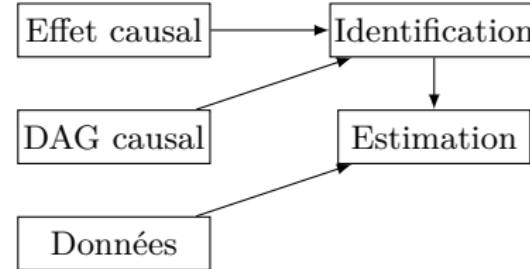
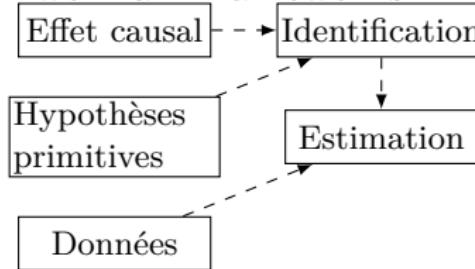


Two main frameworks



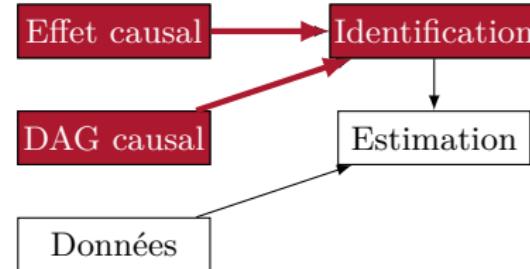
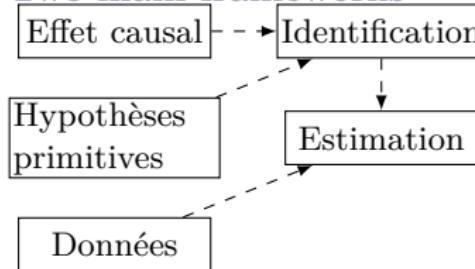


Two main frameworks





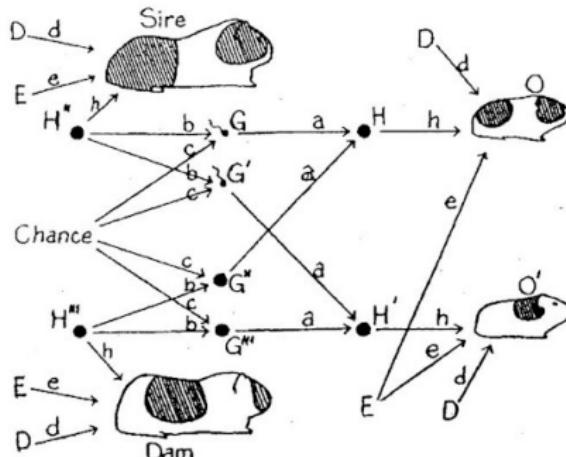
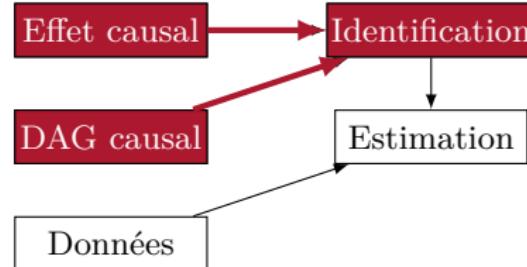
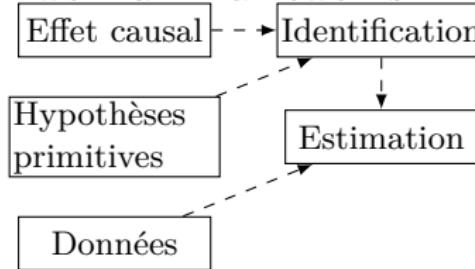
Two main frameworks



Causal inference



Two main frameworks

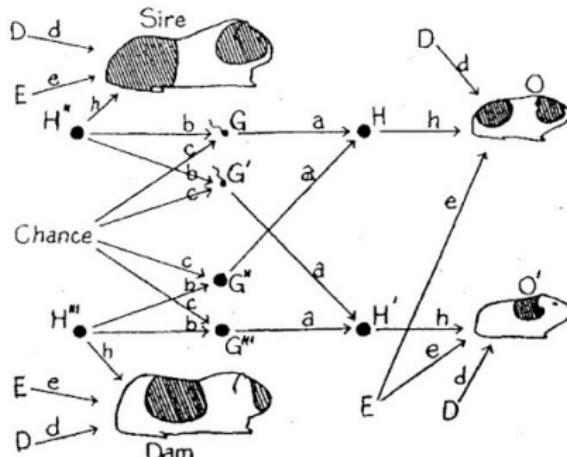
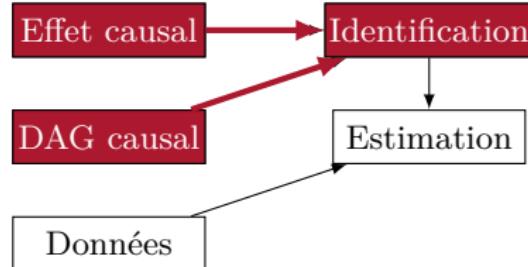
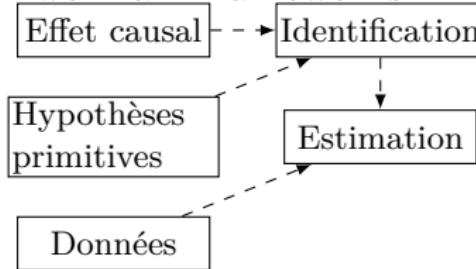


Sewall
Wright

Causal inference



Two main frameworks



Sewall
Wright



Judea
Pearl

2

Préliminaires

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

$$P(A, B | C) = P(A | C)P(B | C)$$

Théorème de Bayes :

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

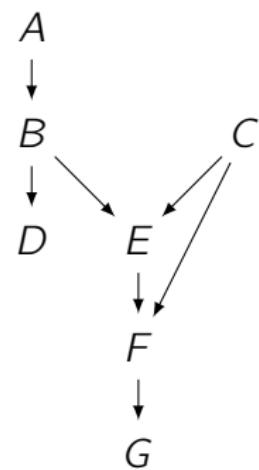
Marginalisation:

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

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

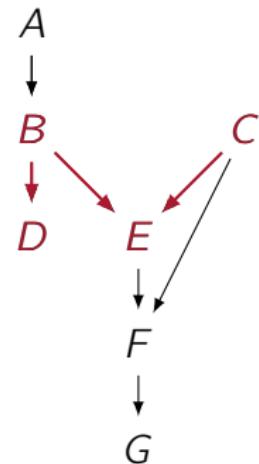
- \mathbb{V} représente l'ensemble des nœuds (chaque nœud correspond généralement à une variable aléatoire),
- \mathbb{E} désigne l'ensemble des arêtes.
- $\forall (X, Y) \in \mathbb{E}$, Il existe une flèche orientée de X vers Y .

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



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

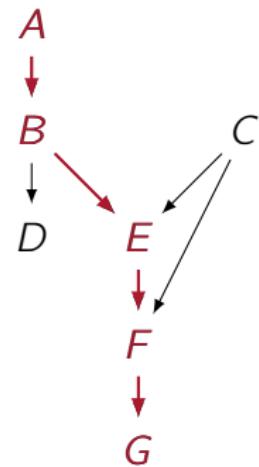
Chemin: $D \leftarrow B \rightarrow E \leftarrow C$



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

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

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

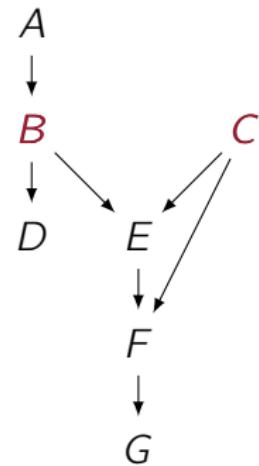


Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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

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

Parents: $Pa(E) = \{B, C\}$



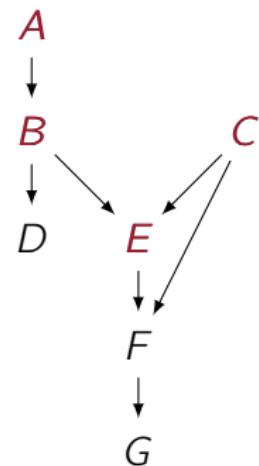
Considérons le graphe orienté suivant $\mathcal{G} = (\mathbb{V}, \mathbb{E})$:

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: $An(E) = \{A, B, C, E\}$



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

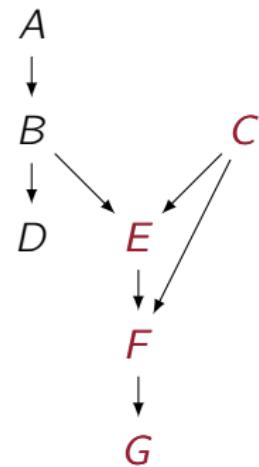
Chemin: $D \leftarrow B \rightarrow E \leftarrow C$

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

Parents: $Pa(E) = \{B, C\}$

Ancêtres: $An(E) = \{A, B, C, E\}$

Descendants: $De(C) = \{C, E, F, G\}$



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

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

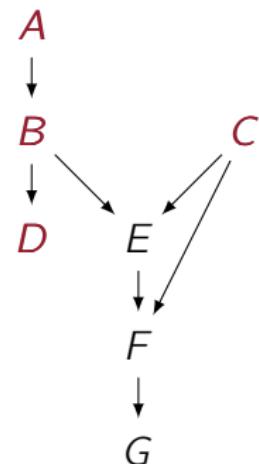
Chemin orienté: $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

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

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} .

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.

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

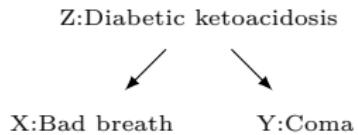
$$P(\mathbb{V}) = \prod_{X \in \mathbb{V}} P(X | Pa(X)).$$

DAG Vs DAG causal

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

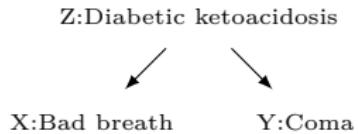
DAG Vs DAG causal

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



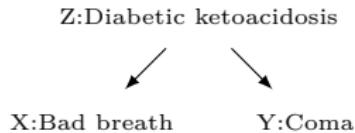
DAG Vs DAG causal

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.

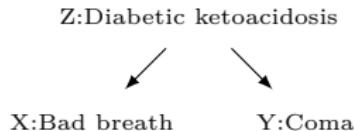


Causal

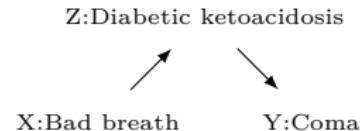
$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

DAG Vs DAG causal

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



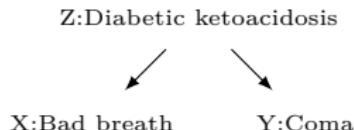
Causal



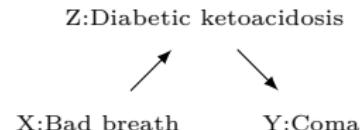
$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

DAG Vs DAG causal

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



Causal

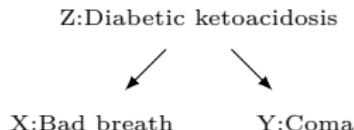


Pas causal

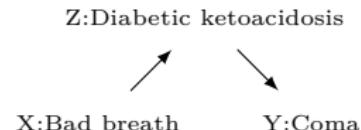
$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

DAG Vs DAG causal

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



Causal



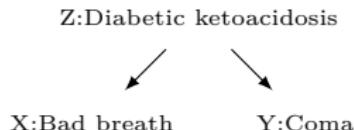
Pas causal

$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

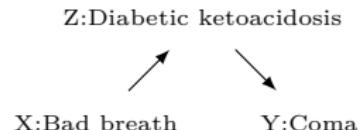
$$P(x, y, z) = P(y | z)P(z | x)P(x)$$

DAG Vs DAG causal

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



Causal



Pas causal

$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

$$= P(y | z)P(z, x)$$

$$= P(y | z) \frac{P(z, x)}{P(x)} P(x)$$

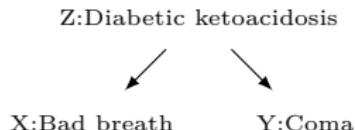
Théorème de Bayes

$$P(x, y, z) = P(y | z)P(z | x)P(x)$$

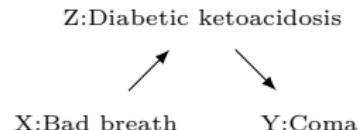
Théorème de Bayes

DAG Vs DAG causal

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



Causal



Pas causal

$$P(x, y, z) = P(y | z)P(x | z)P(z)$$

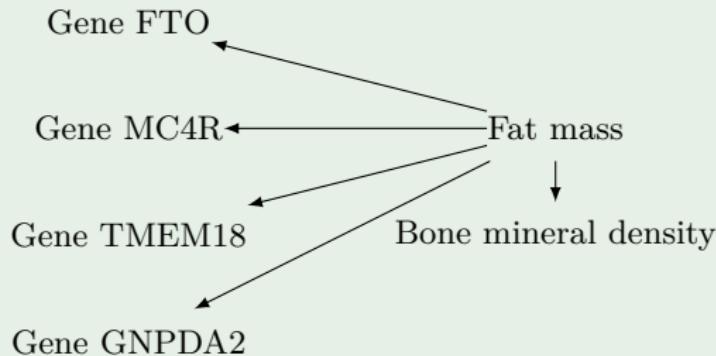
$$= P(y | z)P(z, x) \qquad \text{Théorème de Bayes}$$

$$= P(y | z) \frac{P(z, x)}{P(x)} P(x)$$

$$P(x, y, z) = P(y | z)P(z | x)P(x) \qquad \text{Théorème de Bayes}$$

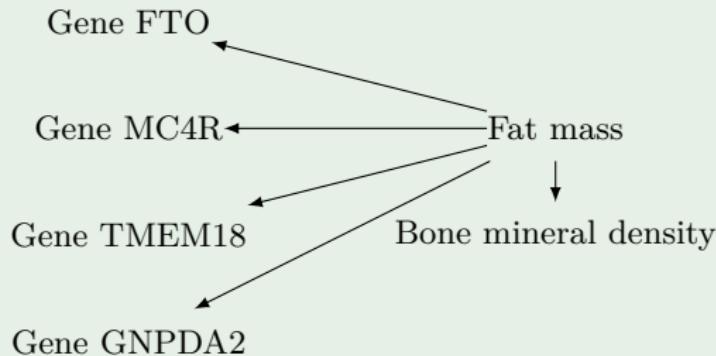
⇒ Il est impossible de déterminer lequel des deux DAGs est causal en se basant uniquement sur la distribution observée !

Example



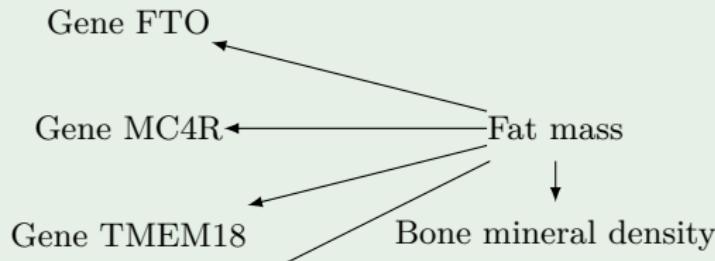
Est ce que ce DAG est causal?

Example

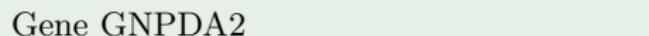


Est ce que ce DAG est causal? **Non**

Example

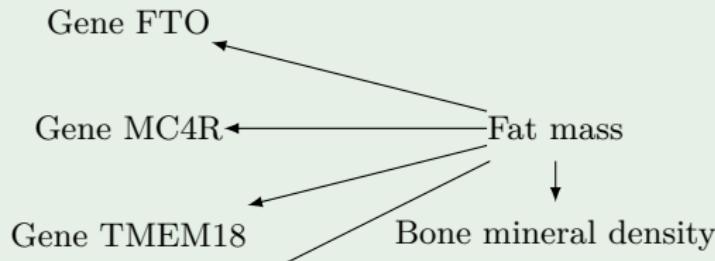


Est ce que ce DAG est causal? **Non**

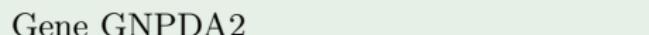


Est ce que ce DAG est causal?

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 $P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(v_i \mid Pa(v_i))$$

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 $P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(v_i \mid Pa(v_i))$$

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

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 $P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s}))$, c'est-à-dire,

$$P(\mathbb{V} \setminus \mathbb{S} \mid do(\mathbb{s})) = \prod_{i \notin \mathbb{S}} P(v_i \mid Pa(v_i))$$

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

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->A \u222a Z->Y")
#Display the DAG
print(dag)
#Visualize the DAG
plot(dag)
```

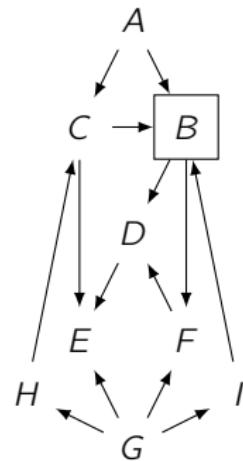
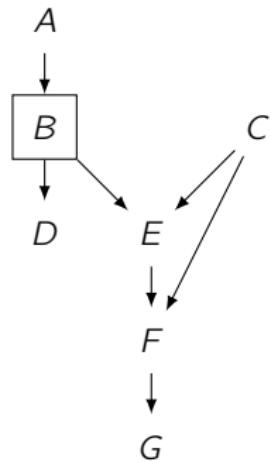
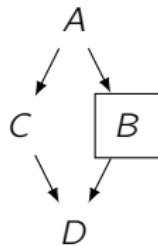
3

DAG causal

Définition
d-separation

Reading conditional independencies in graphs

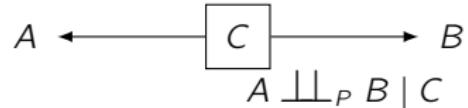
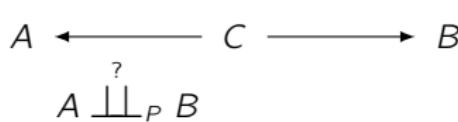
$$A \perp\!\!\!\perp_D ? \mid B$$



Basic structures

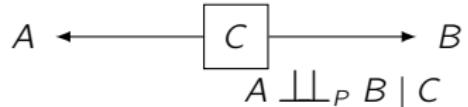
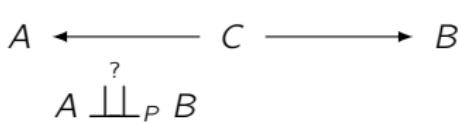
Basic structures

Fourchette: contenant un facteur de confusion

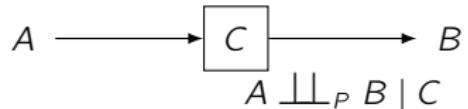
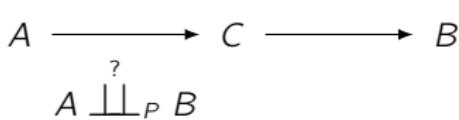


Basic structures

Fourchette: contenant un facteur de confusion

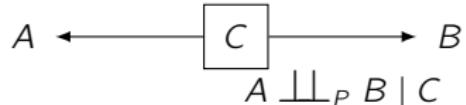
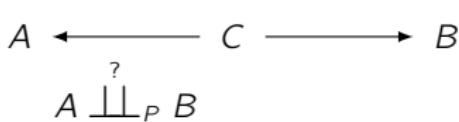


Chaîne: contenant mediateur ou cause intermédiaire

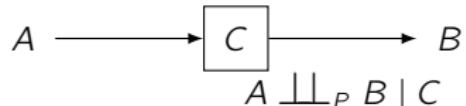
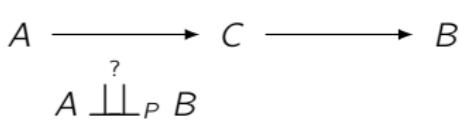


Basic structures

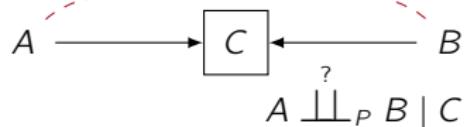
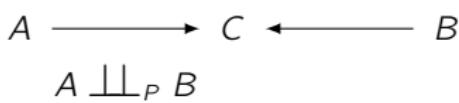
Fourchette: contenant un facteur de confusion



Chaîne: contenant mediateur ou cause intermédiaire

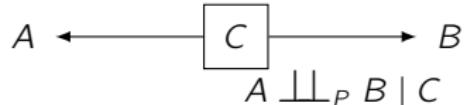
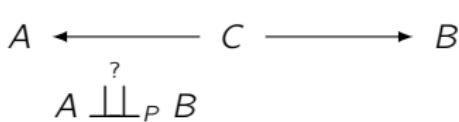


V-structure: contenant un collider

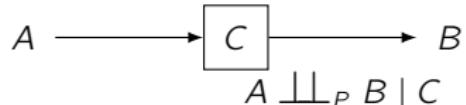
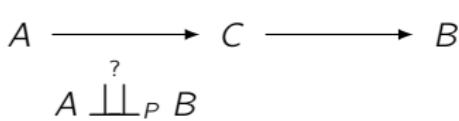


Basic structures

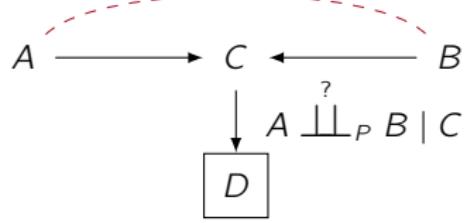
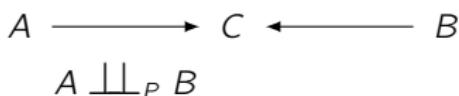
Fourchette: contenant un facteur de confusion



Chaîne: contenant mediateur ou cause intermédiaire

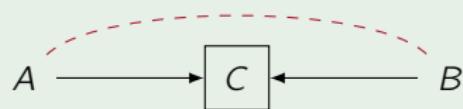


V-structure: contenant un collider



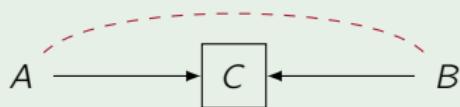
Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example



Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example



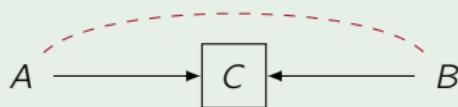
$$A = \begin{cases} \text{Mère porteuse} \\ \text{Mère non porteuse} \end{cases}$$

$$B = \begin{cases} \text{Père porteur} \\ \text{Père non porteur} \end{cases}$$

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

Corrélation artificielle en conditionnant sur des colliders : Exemple 1

Example



$$A = \begin{cases} \text{Mère porteuse} \\ \text{Mère non porteuse} \end{cases}$$

$$B = \begin{cases} \text{Père porteur} \\ \text{Père non porteur} \end{cases}$$

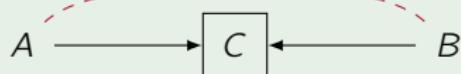
$$C = (A \text{ or } B) = \begin{cases} \text{Enfant porteur} \\ \text{Enfant non porteur} \end{cases}$$

Si $C = \text{Enfant porteur} \implies$

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

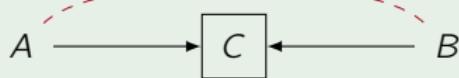
Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example



Corrélation artificielle en conditionnant sur des colliders :
Exemple 2

Example



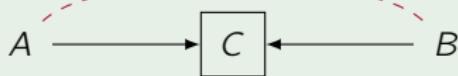
$$A, B \sim U(-1, 1)$$

$$\xi_c \sim N(0, \frac{1}{2})$$

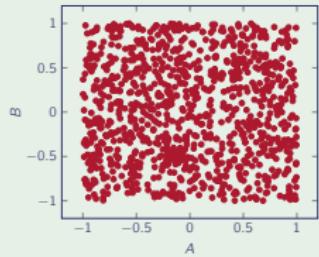
$$C = 2AB + \xi_c$$

Corrélation artificielle en conditionnant sur des colliders :
Exemple 2

Example



$$A, B \sim U(-1, 1)$$



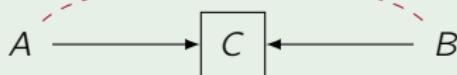
$$\xi_c \sim N(0, \frac{1}{2})$$

$$C = 2AB + \xi_c$$

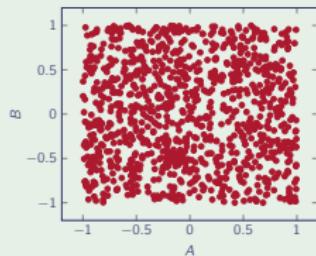
$$\text{Corr}(A; B) = 0.002$$

Corrélation artificielle en conditionnant sur des colliders : Exemple 2

Example

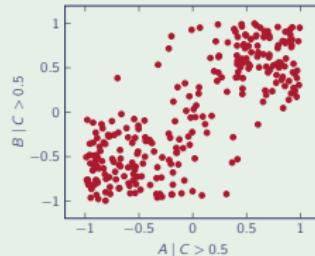


$$A, B \sim U(-1, 1)$$



$$\text{Corr}(A; B) = 0.002$$

$$\xi_c \sim N(0, \frac{1}{2})$$



$$\text{Corr}(A; B | C > 0.5) = 0.8$$

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} .

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**.

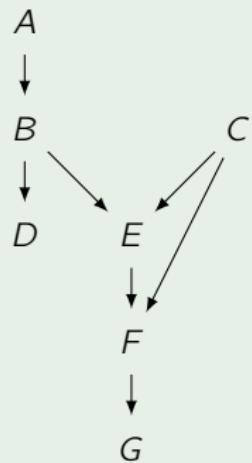
Un chemin est dit être **bloqué** par un ensemble de noeuds $\mathbb{Z} \in \mathbb{V}$ si

- 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} .

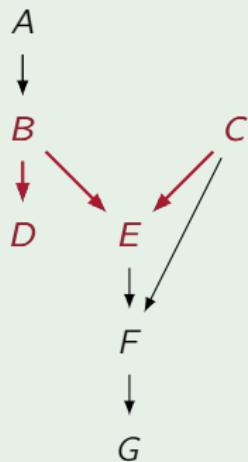
Un chemin qui n'est pas bloqué est dit **actif**.

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

Example

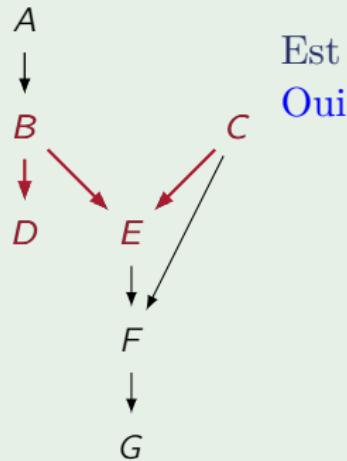


Example



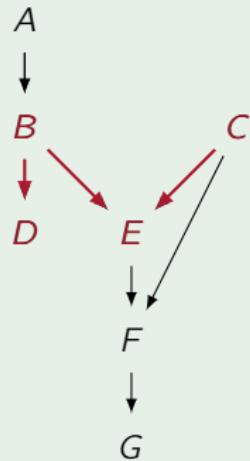
Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué?

Example



Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué?
Oui

Example

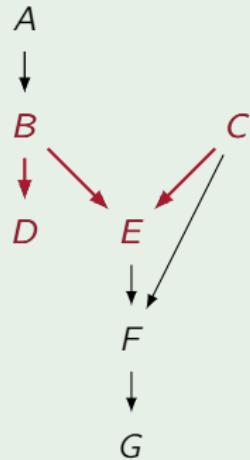


Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué?

Oui

Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué par E ?

Example

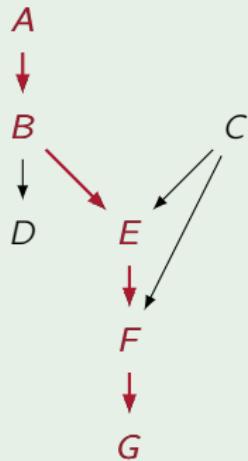


Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué?

Oui

Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué par E ? Non

Example



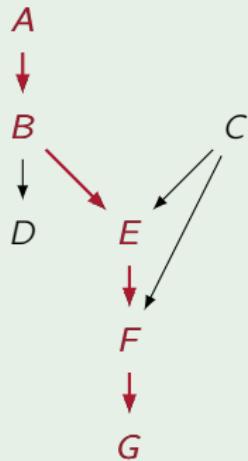
Est ce que le chemin $< D, B, E, C >$ est bloqué?

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?

Example



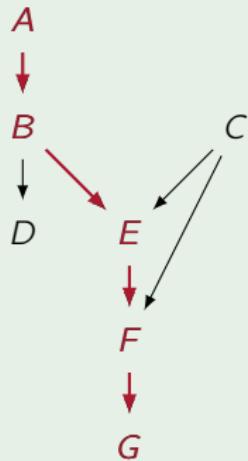
Est ce que le chemin $< D, B, E, C >$ est bloqué?

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?

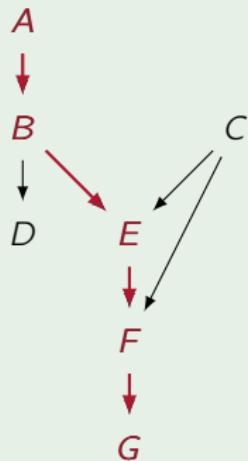
Oui

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 ?

Example



Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué?

Oui

Est ce que le chemin $\langle D, B, E, C \rangle$ est bloqué par E ? Non

Est ce que le chemin $\langle A, B, E, F, G \rangle$ est bloqué? Non

Est ce que le chemin $\langle A, B, E, F, G \rangle$ est bloqué par E ? Oui

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 $\mathbb{A} \perp\!\!\!\perp_G \mathbb{Y} | \mathbb{Z}$.

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 $\mathbb{A} \perp\!\!\!\perp_G \mathbb{Y} | \mathbb{Z}$.

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} .

Theorem

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

Preuve dans (Pearl, 1988)

Theorem

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

Preuve dans (Pearl, 1988)

Theorem

- (i) Si $A \perp\!\!\!\perp_{\mathcal{G}} Y | Z$, alors $A \perp\!\!\!\perp_P Y | Z$ dans toute distribution P compatible avec \mathcal{G} .
- (ii) Si $A \not\perp\!\!\!\perp_{\mathcal{G}} Y | Z$, alors il existe une distribution P compatible avec \mathcal{G} telle que $A \not\perp\!\!\!\perp_P Y | Z$.

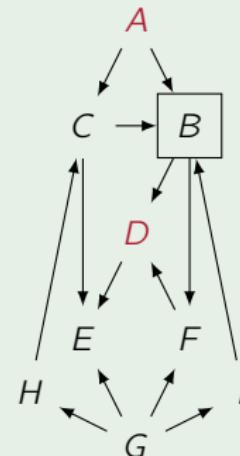
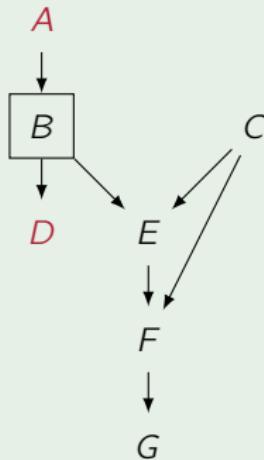
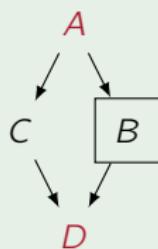
Preuve dans (Pearl, 1988)

- (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 $A \rightarrow Y$ dans le DAG, ajoutez-le (tant que le graphe reste acyclique) !

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

Example

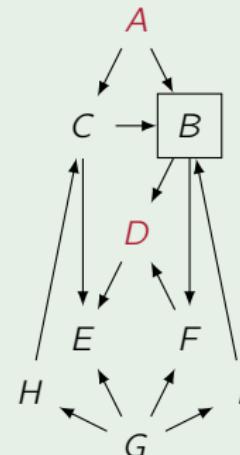
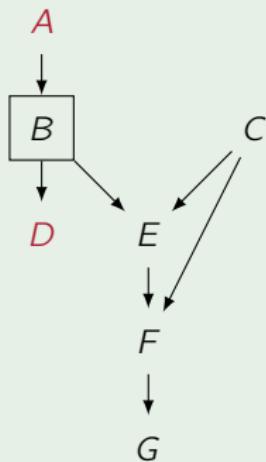
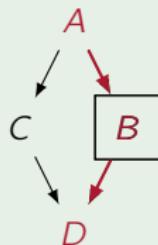
$$A \perp\!\!\!\perp_P D | B$$



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

Example

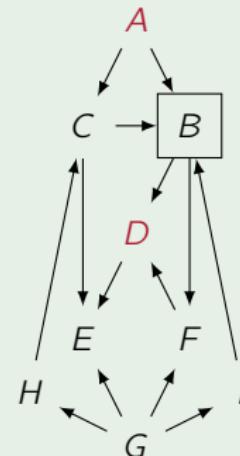
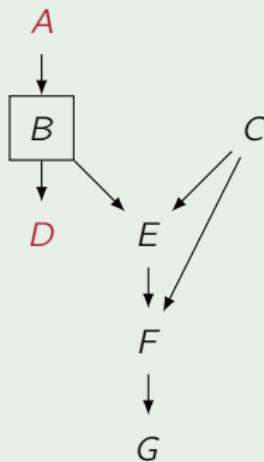
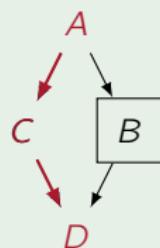
$$A \perp\!\!\!\perp_P D | B$$



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

Example

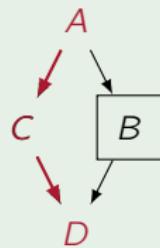
$$A \perp\!\!\!\perp_P D | B$$



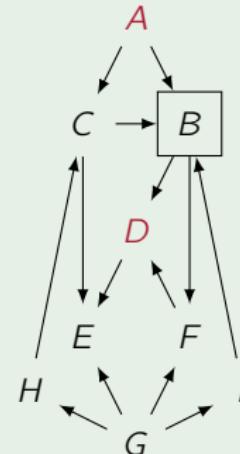
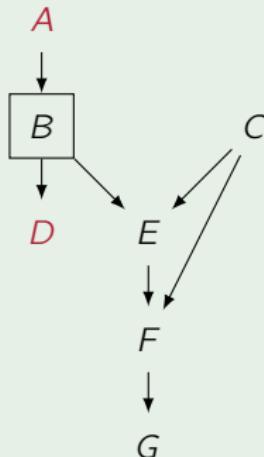
Lire les indépendances conditionnelles à l'aide de la d-separation

Example

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



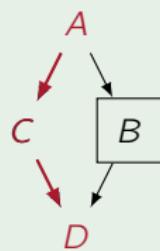
$\langle A, C, D \rangle$ n'est
pas bloqué
 $\Rightarrow A \perp\!\!\!\perp_P D \mid B$



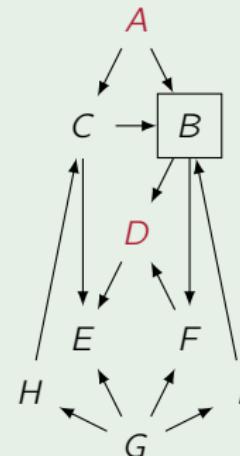
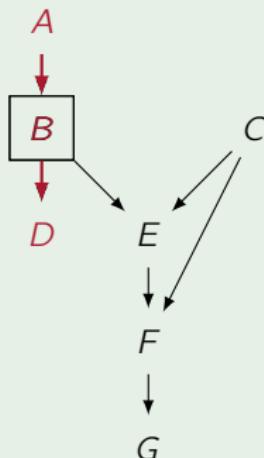
Lire les indépendances conditionnelles à l'aide de la d-separation

Example

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



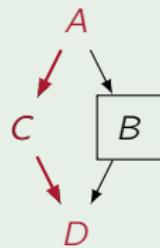
$\langle A, C, D \rangle$ n'est
pas bloqué
 $\Rightarrow A \perp\!\!\!\perp_P D \mid B$



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

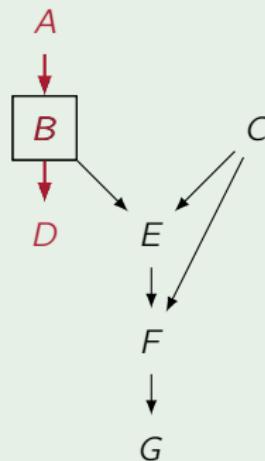
Example

$$A \perp\!\!\!\perp_P D | B$$

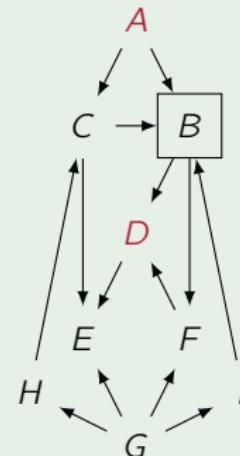


$\langle A, C, D \rangle$ n'est
pas bloqué

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



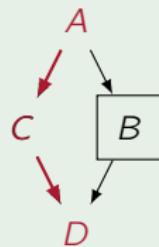
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

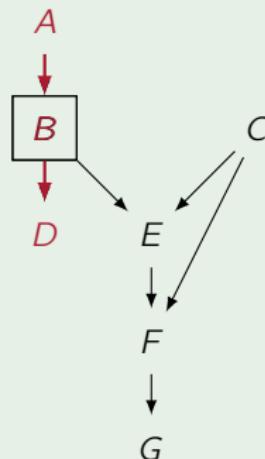
Example

$$A \perp\!\!\!\perp_P D | B$$

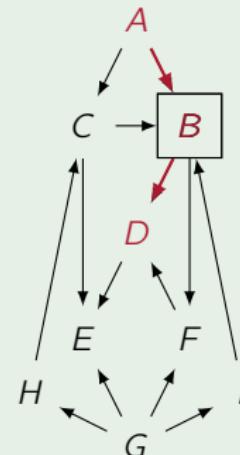


$\langle A, C, D \rangle$ n'est
pas bloqué

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



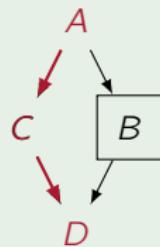
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

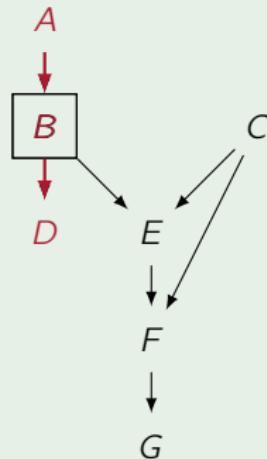
Example

$$A \perp\!\!\!\perp_P D | B$$

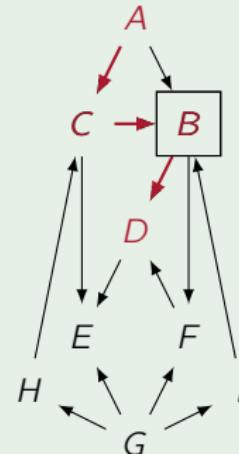


$\langle A, C, D \rangle$ n'est
pas bloqué

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



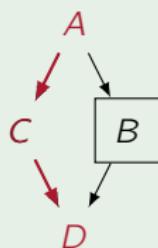
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

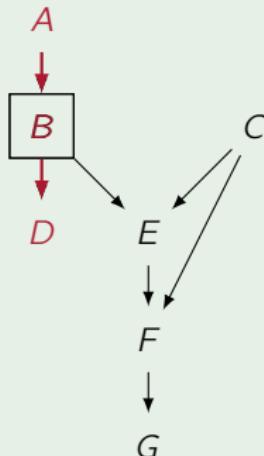
Example

$$A \perp\!\!\!\perp_P D | B$$

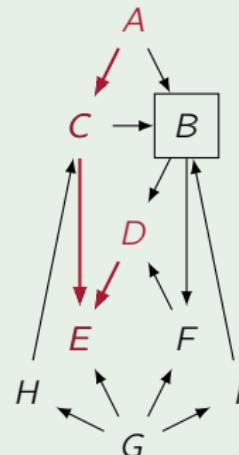


$\langle A, C, D \rangle$ n'est
pas bloqué

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



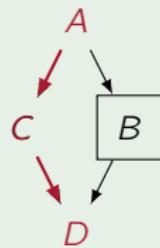
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

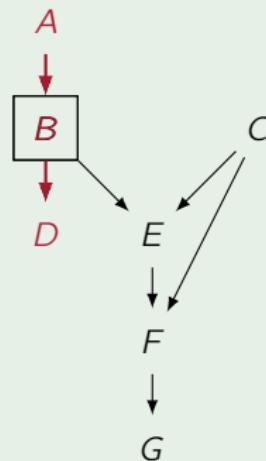
Example

$$A \perp\!\!\!\perp_P D | B$$

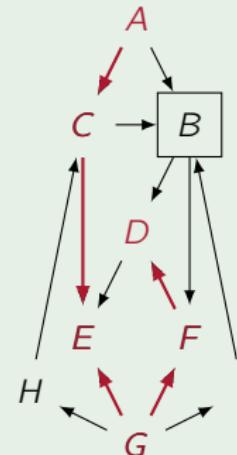


$\langle A, C, D \rangle$ n'est
pas bloqué

$$\Rightarrow A \perp\!\!\!\perp_P D | B$$



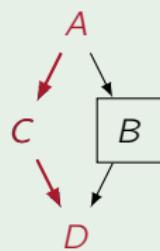
Tous les chemins
sont bloqués
 $\Rightarrow A \perp\!\!\!\perp_P D | B$



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

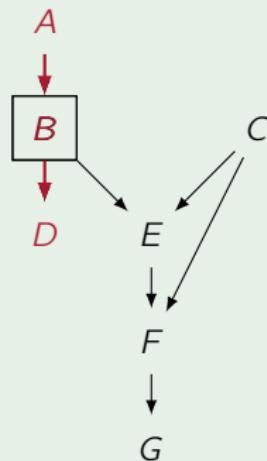
Example

$$A \perp\!\!\!\perp_P D | B$$

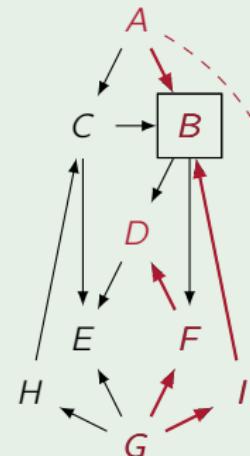


$\langle A, C, D \rangle$ n'est
pas bloqué

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



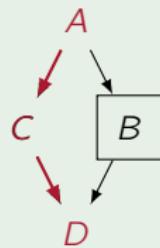
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

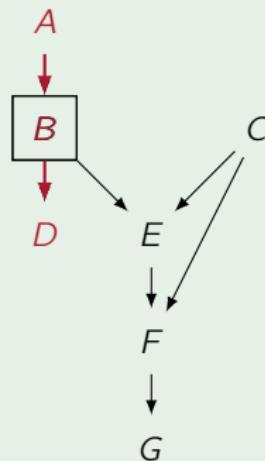
Example

$$A \perp\!\!\!\perp_P D | B$$

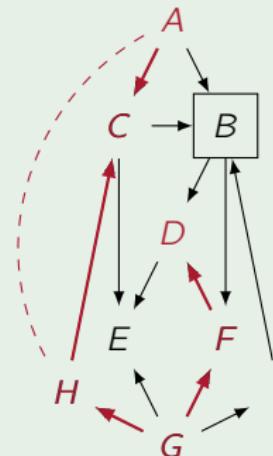


$\langle A, C, D \rangle$ n'est
pas bloqué

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



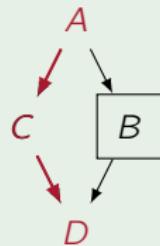
Tous les chemins
sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



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

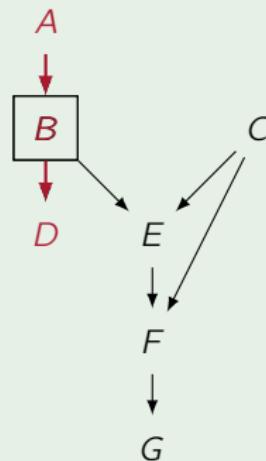
Example

$$A \perp\!\!\!\perp_P D | B$$

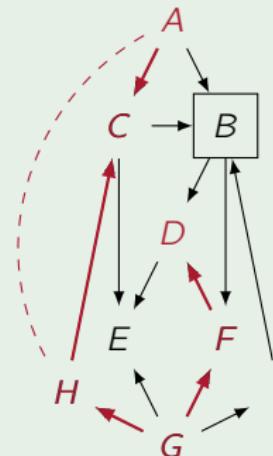


$\langle A, C, D \rangle$ n'est pas bloqué

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



Tous les chemins sont bloqués
 $\implies A \perp\!\!\!\perp_P D | B$



$\langle A, B, I, G, F, D \rangle$ n'est pas bloqué
 $\implies A \perp\!\!\!\perp_P D | B$

Code: d-seperation using DAGitty package

```
#Load the DAGitty package
library(dagitty)
#Define the DAG
dag <- dagitty("Z->A \u222a Z->Y")
#Are A and Y d-separated given Z?
res <- dseparated(dag, "A", "Y", "Z")
print(res)
```

- Tous les DAGs ne sont pas causaux.
- Un DAG causal doit être compatible avec la distribution observée ainsi qu'avec toutes les distributions interventionnelles.
- 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 créer et de visualiser des DAG, ainsi que de tester la d-séparation dans R.

4

Identification d'un effet causal

Identification

Direct causes adjustment

The back-door criterion

L'ensemble back-door optimal

4

Identification d'un effet causal

Identification

Direct causes adjustment

The back-door criterion

L'ensemble back-door optimal

Rappel:

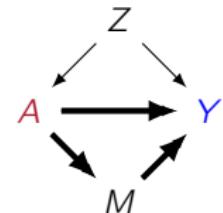
Un effet causal de \mathbb{A} sur \mathbb{Y}

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

Rappel:

Un effet causal de A sur Y

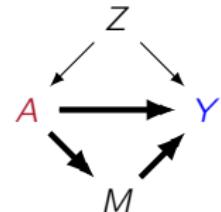
$$= \mathbb{E}(Y \mid do(A = a)) - \mathbb{E}(Y \mid do(A = a'))$$



Rappel:

Un effet causal de A sur Y

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



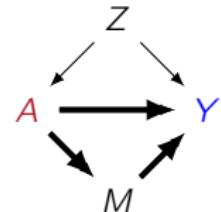
L'effet causal de A sur Y est dit **identifiable** à partir de P dans \mathcal{G} s'il peut être calculé de manière unique à partir de $P(Y)$.

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

Rappel:

Un effet causal de A sur Y

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

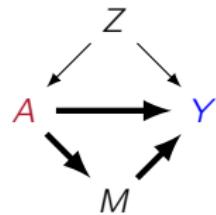


L'effet causal de A sur Y est dit **identifiable** à partir de P dans \mathcal{G} s'il peut être calculé de manière unique à partir de $P(Y)$.

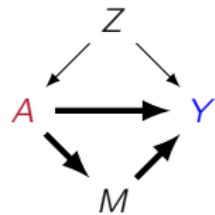
Autrement dit, si $P(Y = y \mid do(A = a))$ peut être calculé de manière unique à partir de $P(Y)$.

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

Biais liés à des facteurs de confusion:

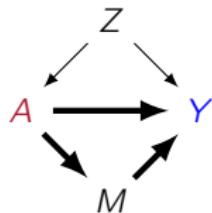


Biais liés à des facteurs de confusion:



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

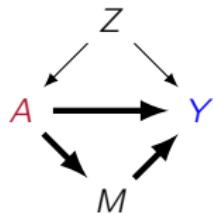
Biais liés à des facteurs de confusion:



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

Devons-nous toujours ajuster sur toutes les variables disponibles?

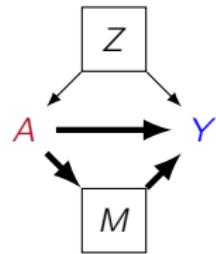
Biais liés à des facteurs de confusion:



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

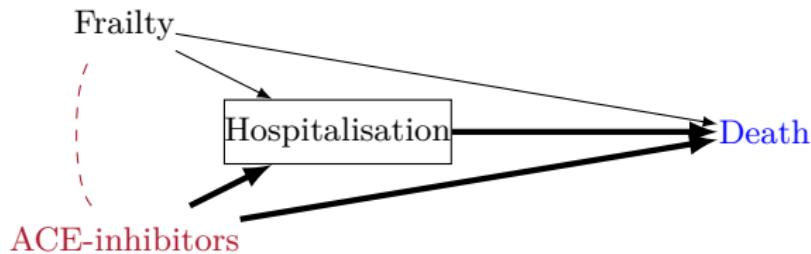
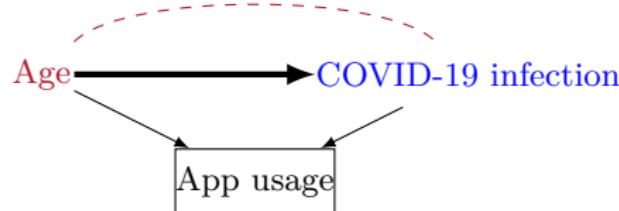
Devons-nous toujours ajuster sur toutes les variables disponibles? **Non!**

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

Identification

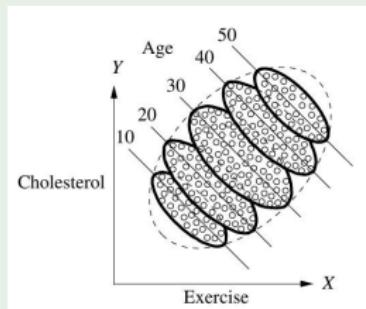
Direct causes adjustment

The back-door criterion

L'ensemble back-door optimal

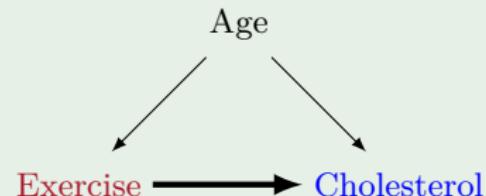
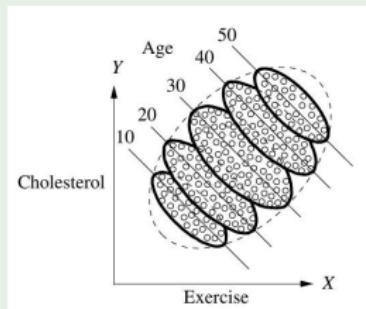
Example

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 $P(c | do(e))$?



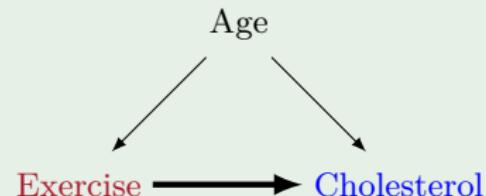
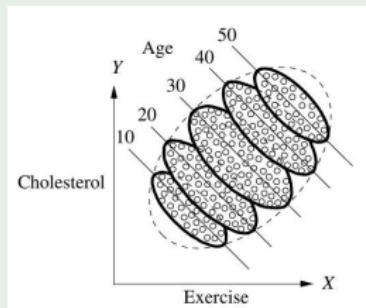
Example

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 $P(c | do(e))$?



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 $P(c | do(e))$?

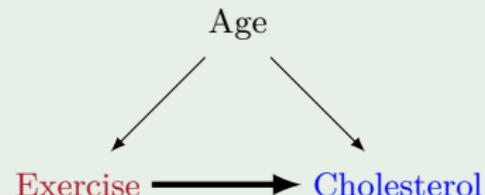
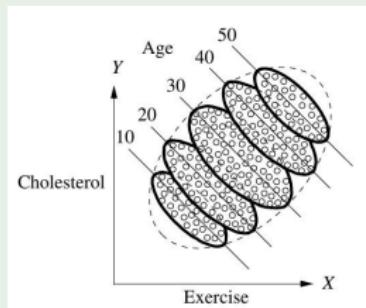


$$P(a, e, c) = P(a)P(e | a)P(c | a, e) \quad (\text{Compatibilité})$$

$$P(a, c | do(e)) = P(a)P(c | a, e) \quad (\text{Compatibilité})$$

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 $P(c | do(e))$?



$$P(a, e, c) = P(a)P(e | a)P(c | a, e) \quad (\text{Compatibilité})$$

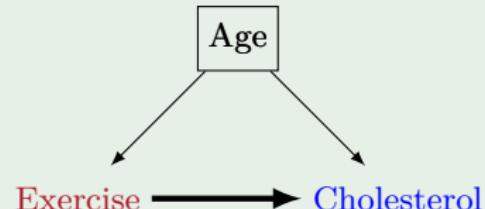
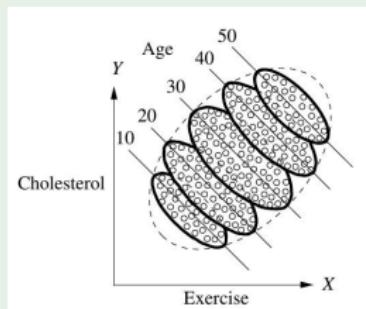
$$P(a, c | do(e)) = P(a)P(c | a, e) \quad (\text{Compatibilité})$$

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

Direct causes adjustment

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 $P(c | do(e))$?



$$P(a, e, c) = P(a)P(e | a)P(c | a, e) \quad (\text{Compatibilité})$$

$$P(a, c | do(e)) = P(a)P(c | a, e) \quad (\text{Compatibilité})$$

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

Direct causes adjustment

Theorem

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

$$P(y | \text{do}(a)) = \sum_{Z \in Pa(A)} P(y | a, z)P(z)$$

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

L'ensemble back-door optimal

Un ensemble de variables 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 Z n'est un descendant de A ; et
- Z bloque tous les chemins entre A et Y qui contiennent une flèche pointant vers A .

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 :

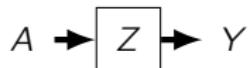
- Aucun nœud de \mathbb{Z} n'est un descendant de A ; et
- \mathbb{Z} bloque tous les chemins entre A et Y qui contiennent une flèche pointant vers A .

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

- 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 ?

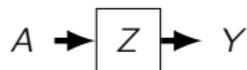
- Pourquoi aucun noeud de Z ne doit être un descendant de A ?



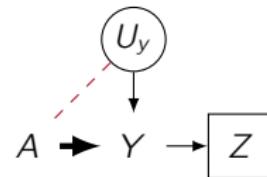
Pour éviter de bloquer sur des causes intermédiaires.

- Pourquoi Z doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A ?

- Pourquoi aucun noeud de Z ne doit être un descendant de A ?



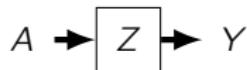
Pour éviter de bloquer sur des causes intermédiaires.



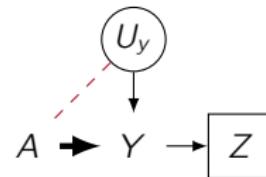
Pour éviter d'introduire un biais de confusion artificiel en conditionnant sur un collider.

- Pourquoi Z doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A ?

- Pourquoi aucun noeud de Z ne doit être un descendant de A ?



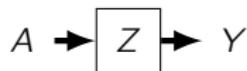
Pour éviter de bloquer sur des causes intermédiaires.



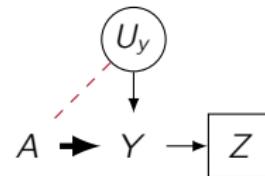
Pour éviter d'introduire un biais de confusion artificiel en conditionnant sur un collider.
Pour éviter les biais de sélection!

- Pourquoi Z doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A ?

- Pourquoi aucun noeud de Z ne doit être un descendant de A ?



Pour éviter de bloquer sur des causes intermédiaires.



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

- Pourquoi Z doit bloquer tous les chemins entre A et Y qui contiennent une flèche pointant vers A ?

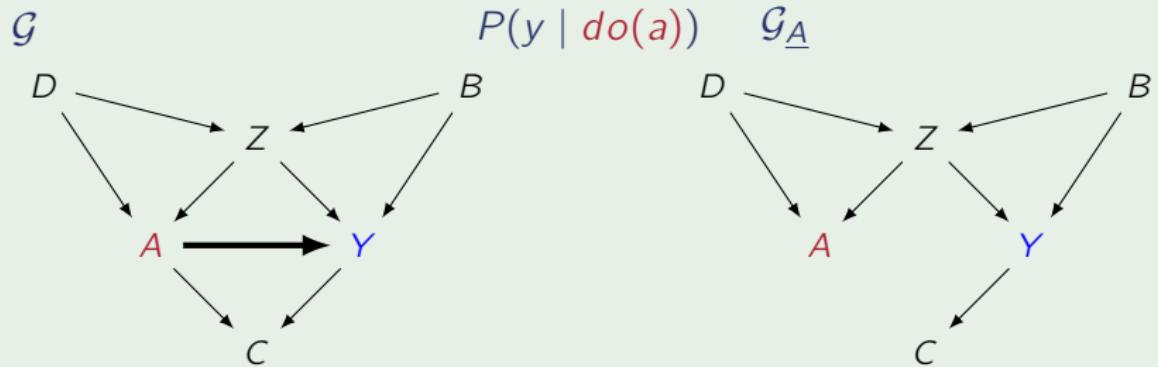
Pour éliminer les biais provenant des facteurs de confusions.

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 :

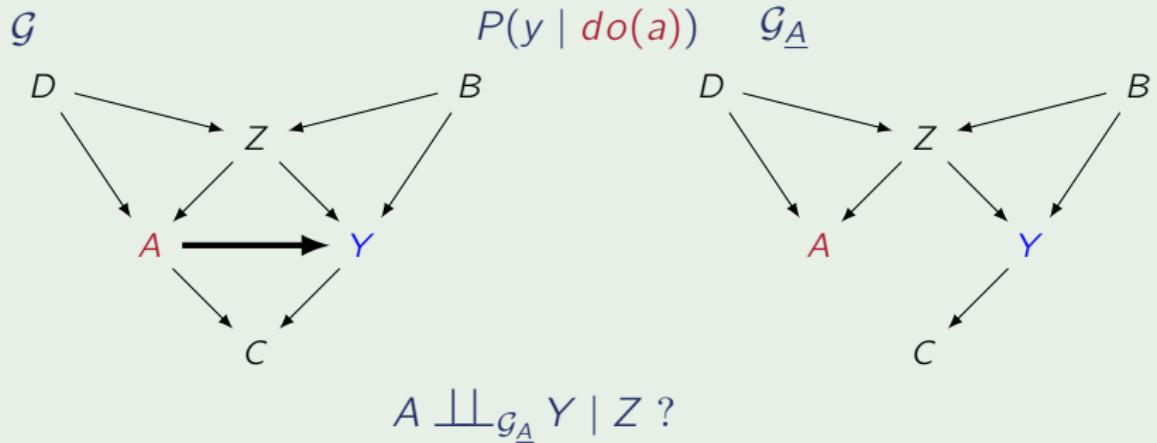
- $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid \mathbb{Z}$,

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

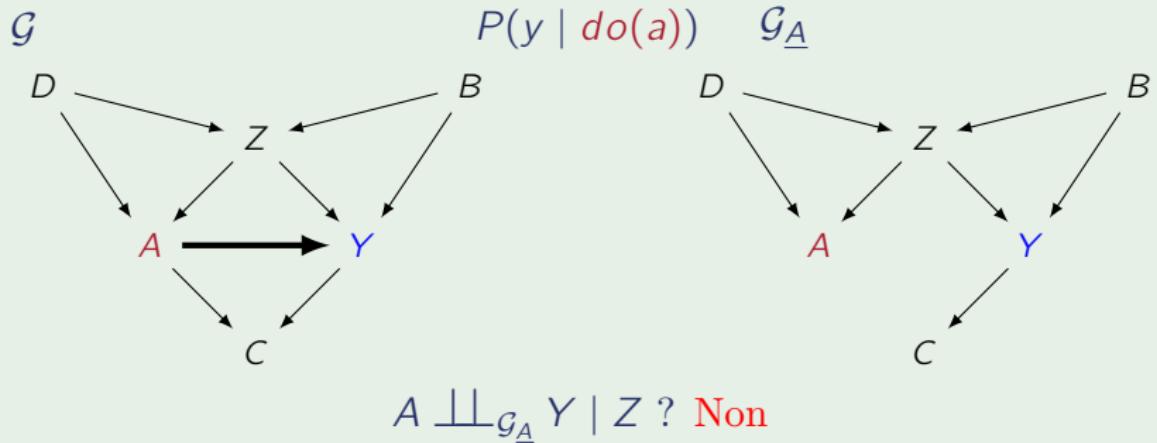
Example



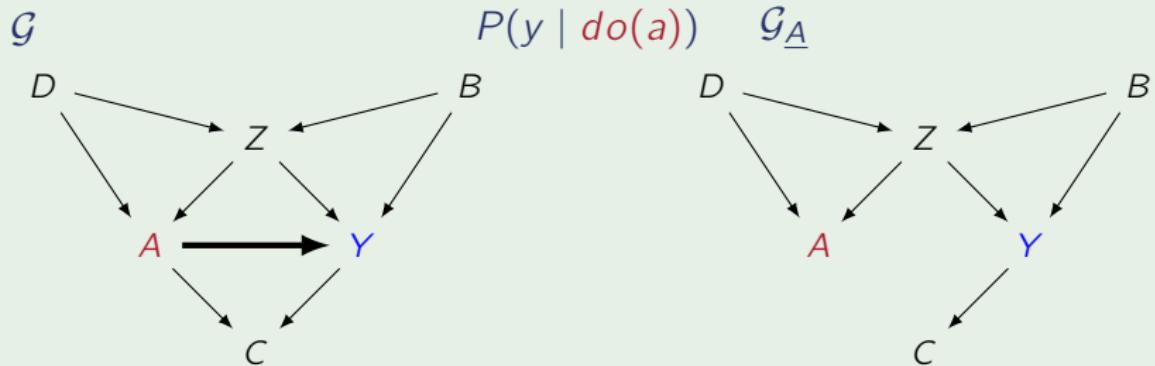
Example



Example

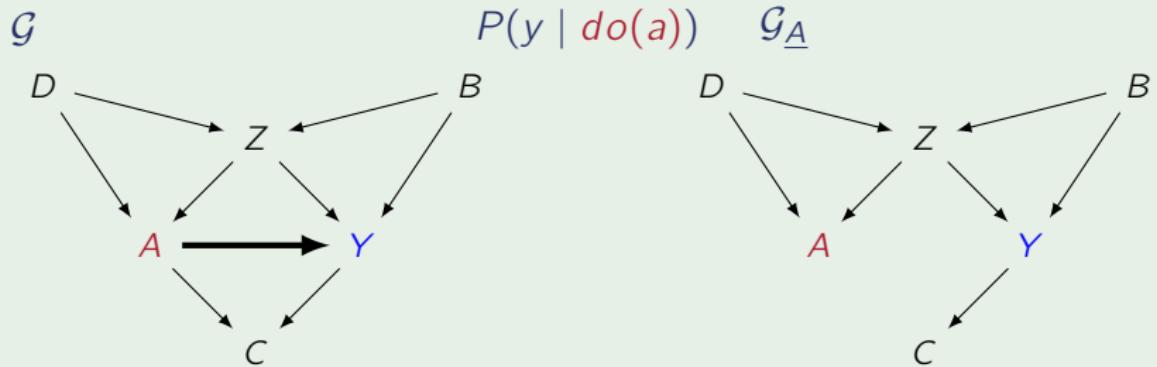


Example



$A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid Z?$ Non
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid D?$

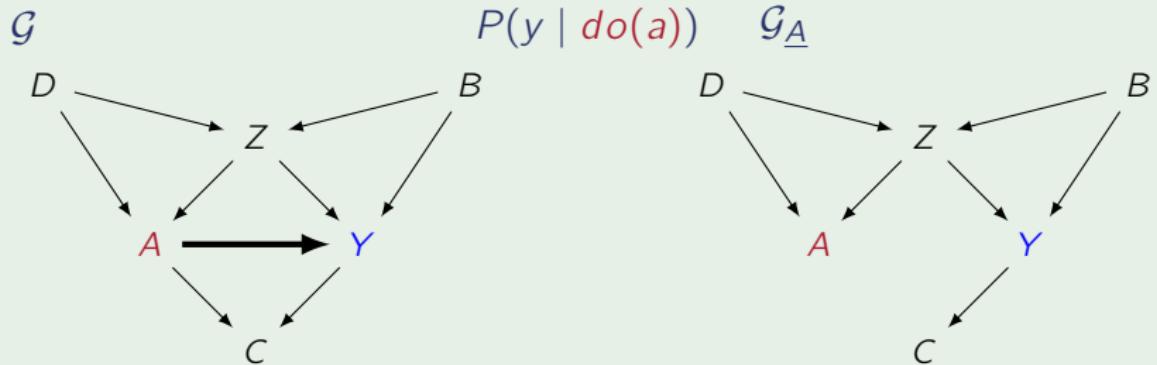
Example



$A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid Z?$ Non
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid D?$ Non

Le critère du back-door: exemple

Example



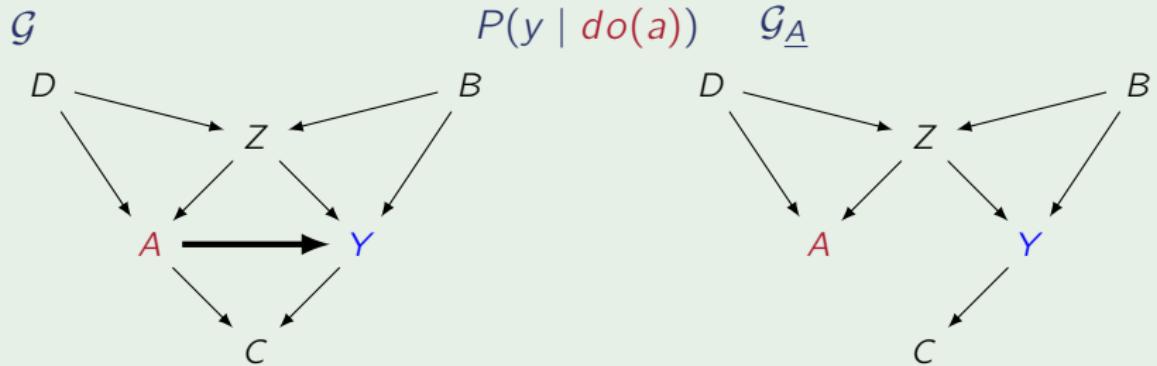
$A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid Z?$ Non

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

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

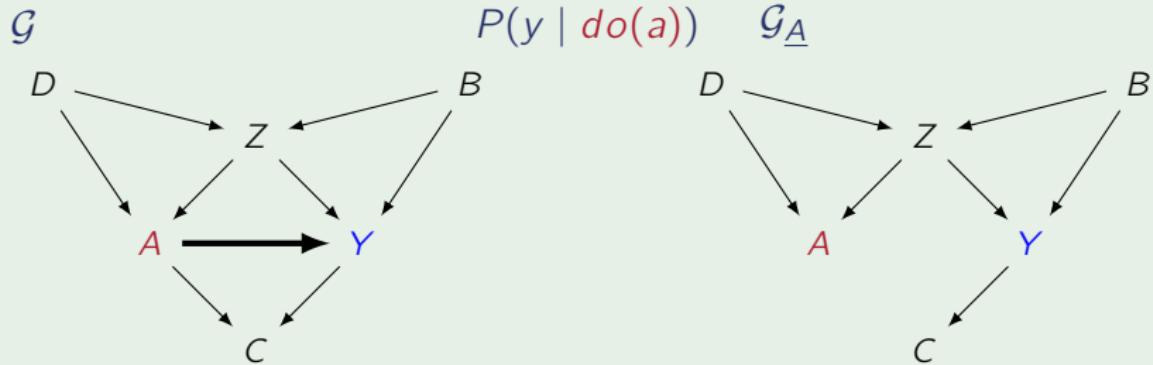
Le critère du back-door: exemple

Example



$A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid Z?$ Non
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid D?$ Non
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid B?$ Non

Example

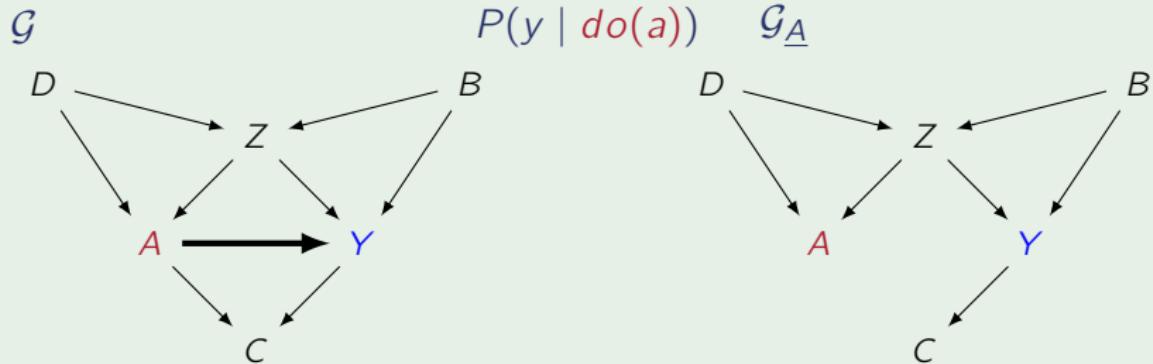

 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid Z?$ Non

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

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

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

Example


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

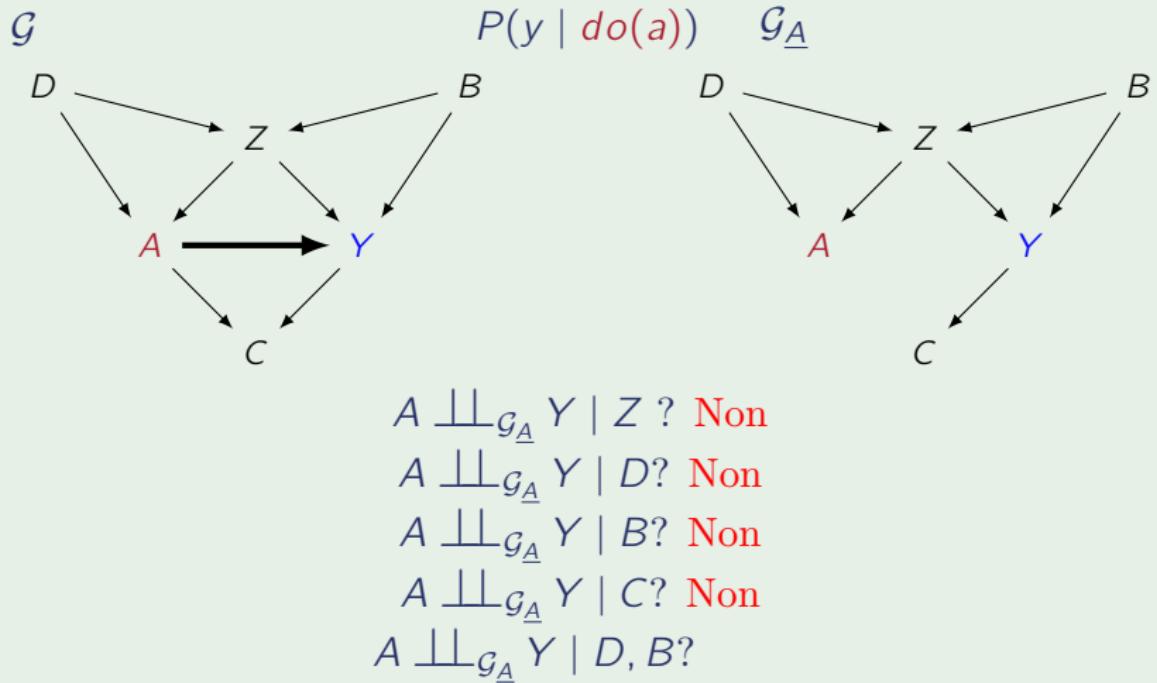
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid D?$ Non

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

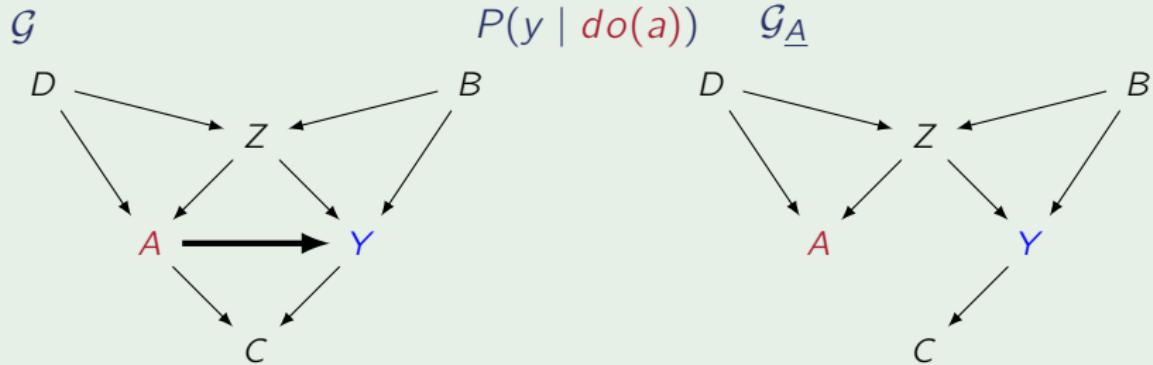
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid C?$ Non

Le critère du back-door: exemple

Example



Example


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

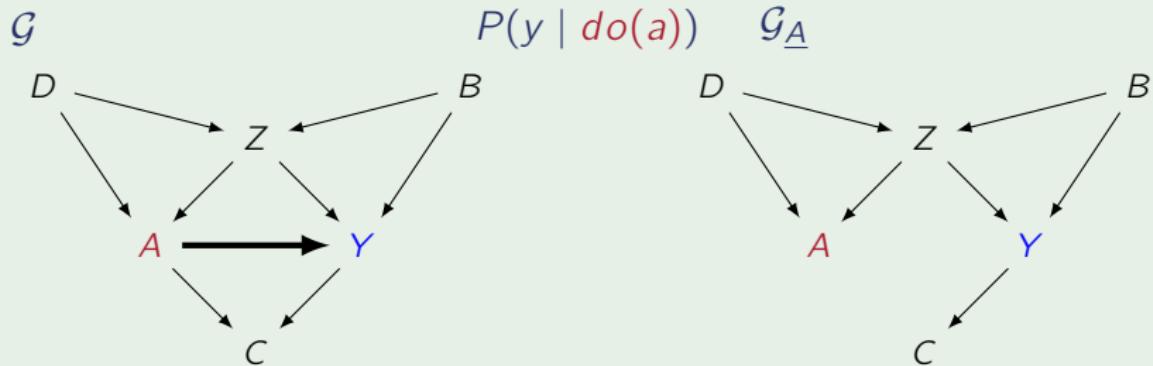
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid D?$ Non

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

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

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

Example


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

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

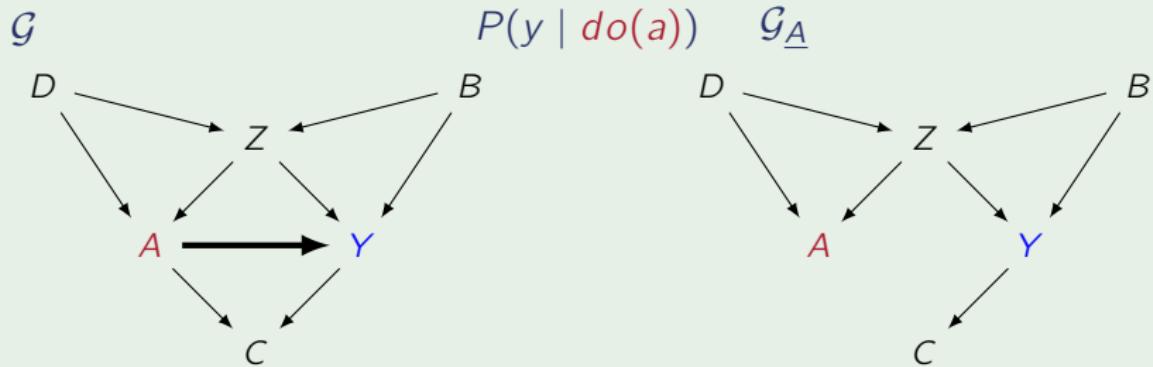
 $A \perp\!\!\!\perp_{\mathcal{G}_A} Y \mid B?$ Non

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

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

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

Example


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

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

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

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

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

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

Theorem

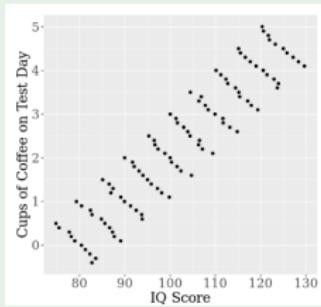
Si Z satisfait le critère du back-door par rapport à (A, Y) et si $P(a, z) > 0$, alors l'effet causal de A sur Y est identifiable et est donné par

$$P(y \mid \textcolor{red}{do}(a)) = \sum_z P(y \mid a, z)P(z).$$

Preuve dans (Pearl, 2000)

Example

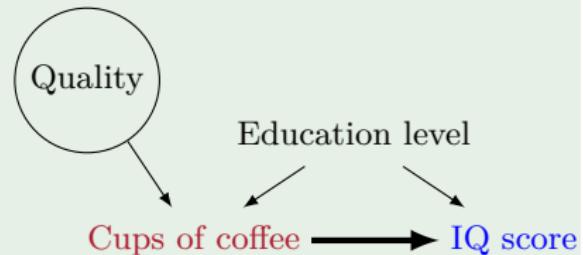
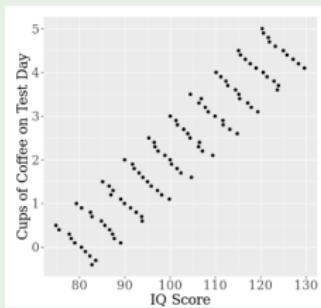
Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.



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

Example

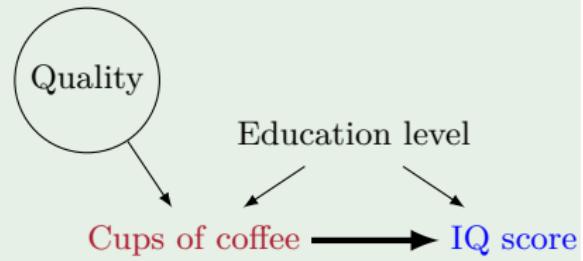
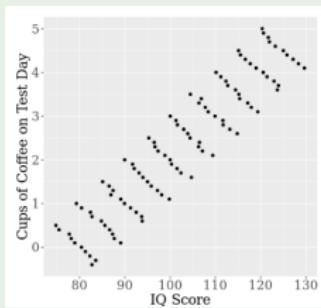
Dans une étude, nous mesurons la consommation de café et le score de QI pour un échantillon de la population ayant différents niveaux d'éducation.



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

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.

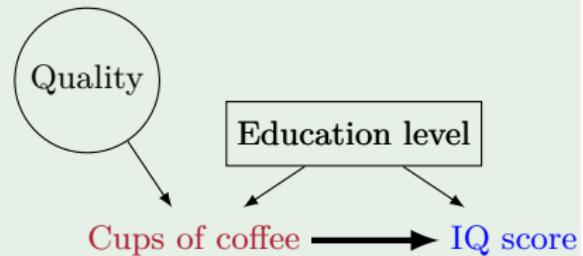
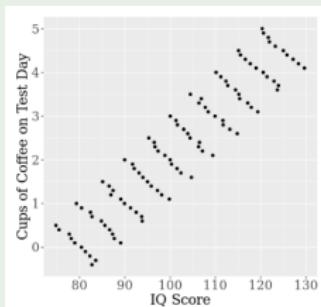


Quel est l'effet du nombre de tasses de café sur le score de QI,
 $P(i | do(c))$?

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

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,
 $P(i | do(c))$?

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

Code: critère du back-door avec DAGitty

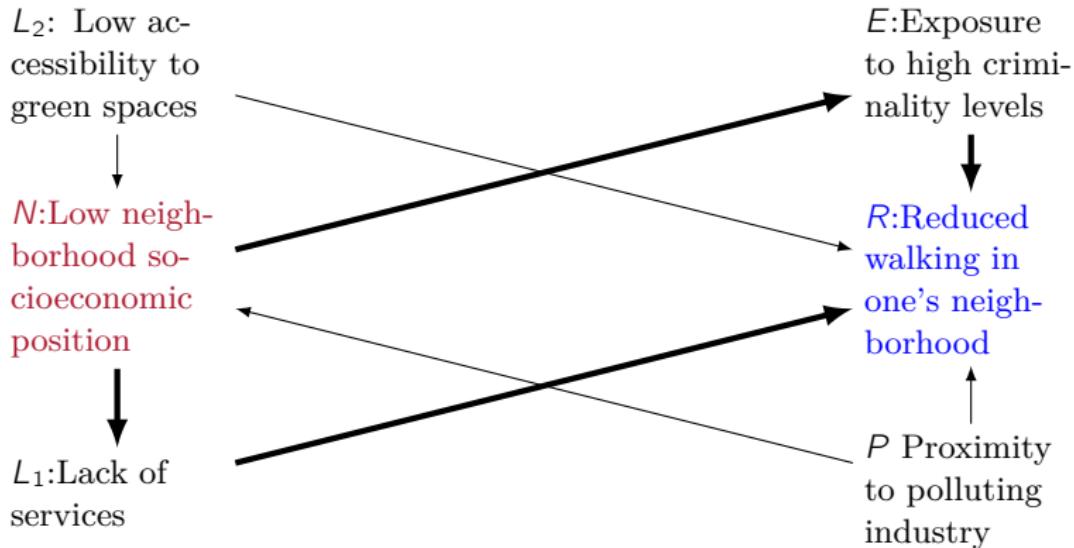
```
#Load the DAGitty package
library(dagitty)
#Define the DAG
dag <- dagitty("Z->A \cup Z->Y \cup A->Y")
#Sets that satisfy the back-door criterion for (A, Y)?
res <- adjustmentSets(dag, exposure="A", outcome="Y")
print(res)
```

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, $P(r | do(n))$?

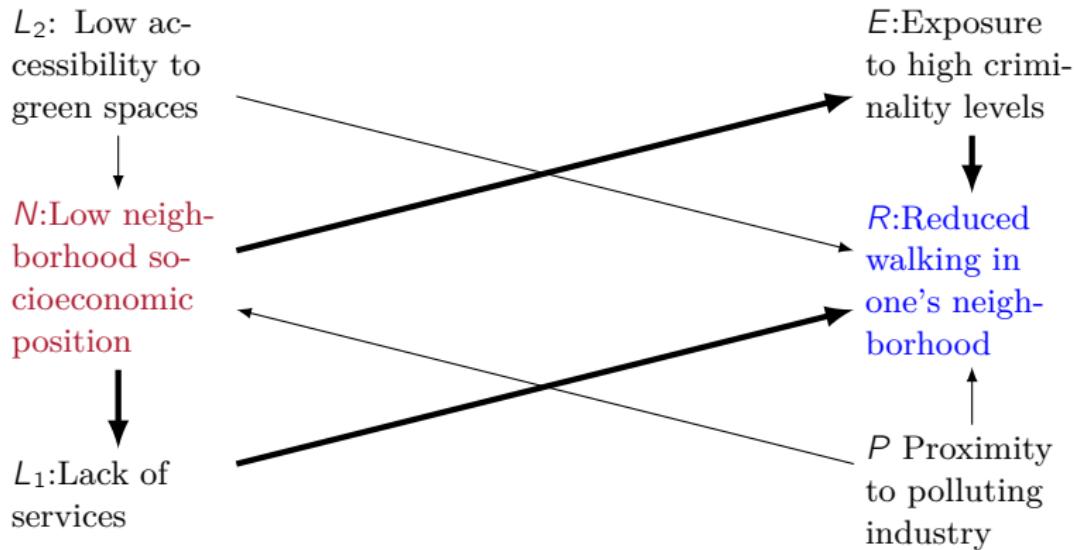
Une application réelle où le critère du back-door est appliqué

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, $P(r | do(n))$?



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, $P(r | do(n))$?



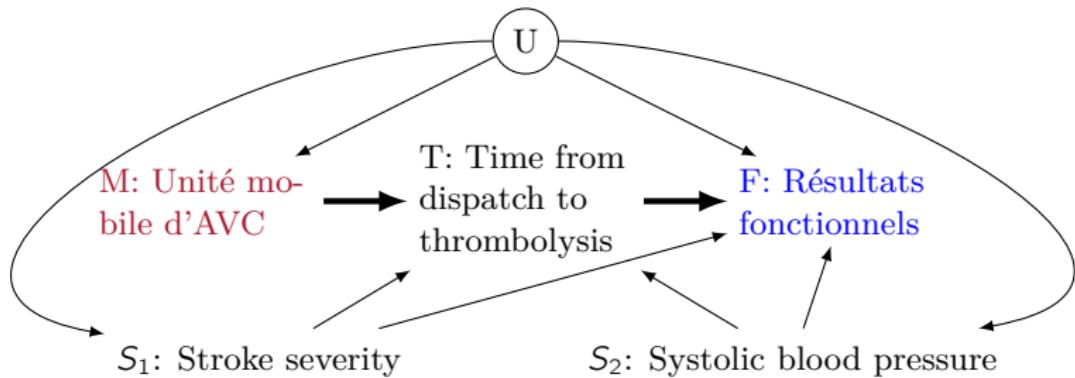
$$P(r | do(n)) = \sum_{l_2, p} P(r|n, l_2, p)P(l_2, p)$$

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, $P(f \mid \textcolor{red}{do}(m))$?

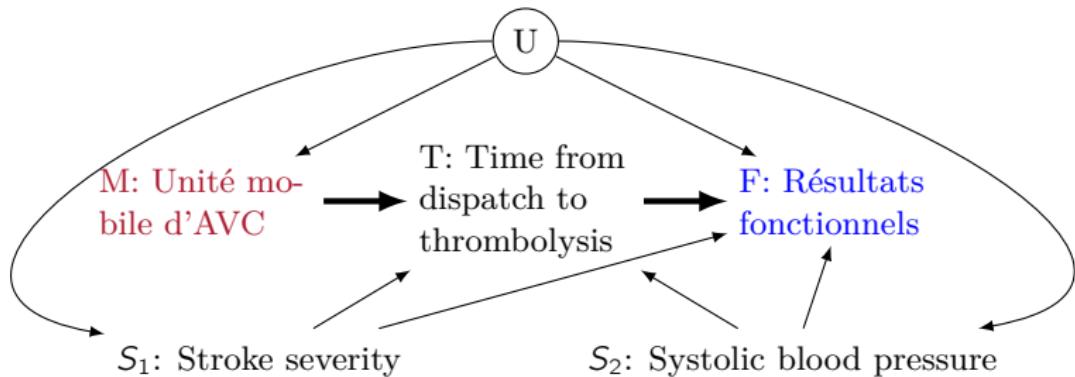
Une application réelle où le critère du back-door ne peut pas être appliqué

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, $P(f | do(m))$?



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

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 $P(y | do(a))$, cela ne signifie pas nécessairement que $P(y | do(a))$ 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 :

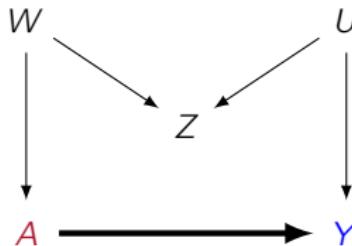
- Z est associé à A ; et
- Z est associé à Y , conditionnellement à A .

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

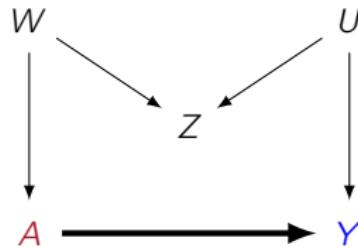


Critère d'association 1 pour identifier les facteurs de confusion (incorrect)

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



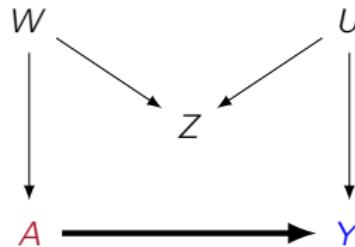
Z est associé à A et Z est associé à Y , conditionnellement à A

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



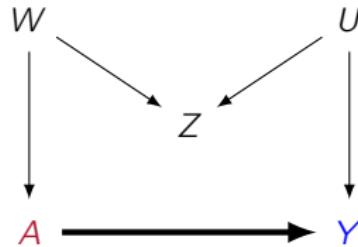
Z est associé à A et Z est associé à Y , conditionnellement à A
Faut-il ajuster sur Z ?

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



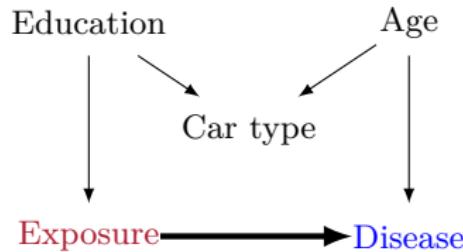
Z est associé à A et Z est associé à Y , conditionnellement à A
Faut-il ajuster sur Z ? **Non!**

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 :

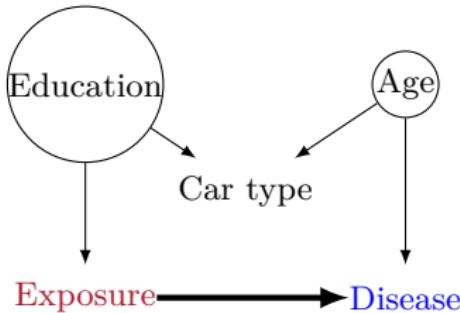
- \mathbb{Z}_1 est associé à A ; et
- \mathbb{Z}_2 est associé à Y , conditionnellement à A et \mathbb{Z}_1 .

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

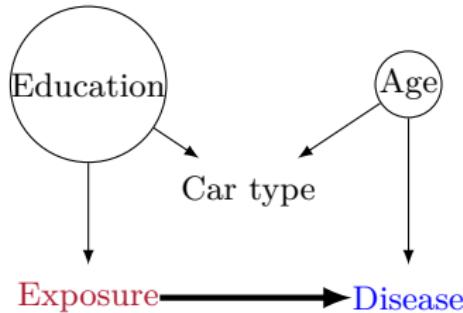


Critère d'association 2 pour identifier les facteurs de confusion (incorrect)

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



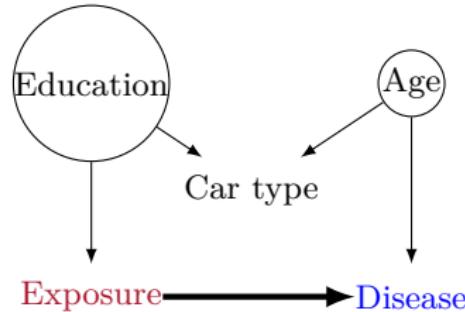
Car type est associé à Exposure

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



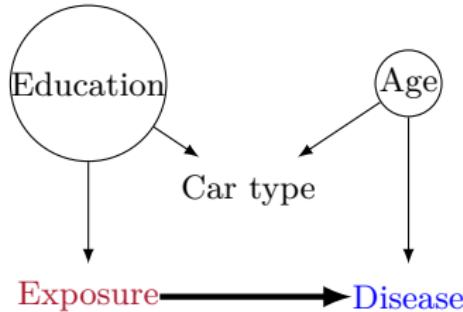
Car type est associé à Exposure
Faut-il ajuster sur Car type?

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 ajuster sur Car type? Non!

4

Identification d'un effet causal

Identification

Direct causes adjustment

The back-door criterion

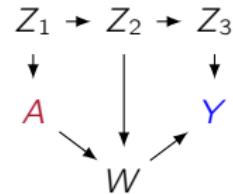
L'ensemble back-door optimal

Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.

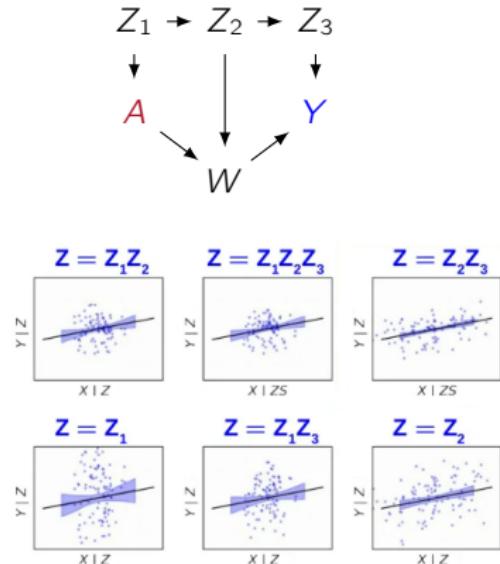
Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.



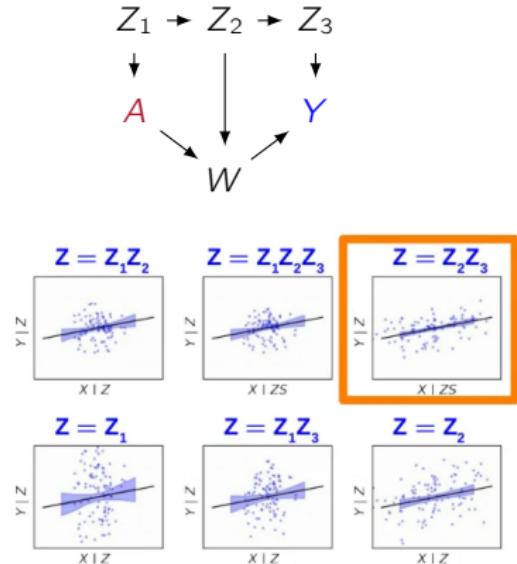
Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.



Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.



Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.
- Peut-on trouver l'ensemble **optimal**, celui qui minimise la variance?

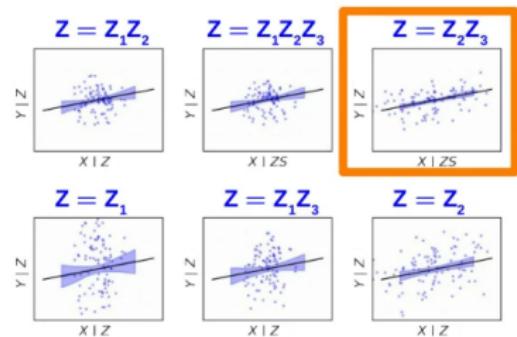
$$Z_1 \rightarrow Z_2 \rightarrow Z_3$$

$$\downarrow \qquad \qquad \qquad \downarrow \qquad \qquad \qquad \downarrow$$

$$A \qquad \qquad \qquad Y$$

$$\searrow \qquad \qquad \qquad \swarrow$$

$$W$$

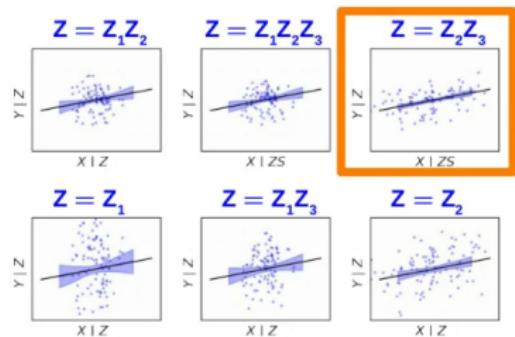
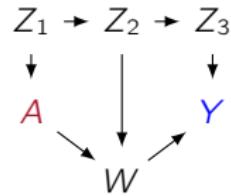


Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.
- Peut-on trouver l'ensemble **optimal**, celui qui minimise la variance?

Un ensemble Z satisfaisant le critère du back-door relativement à (A, Y) est dit **optimal** s'il vérifie :

- Z contient tous les parents de Y qui ne sont pas des médiateurs entre A et Y ; et

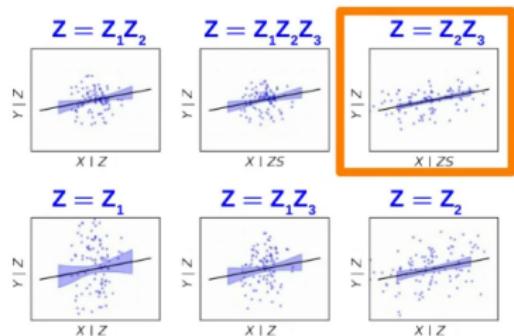
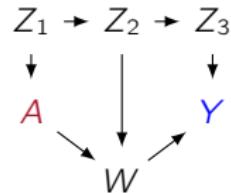


Plusieurs ensembles peuvent satisfaire le critère du back-door :

- Tous ces ensembles permettent d'éliminer le biais confusion.
- Cependant, ils ne conduisent pas tous à la même variance.
- Peut-on trouver l'ensemble **optimal**, celui qui minimise la variance?

Un ensemble \mathbb{Z} satisfaisant le critère du back-door relativement à (A, Y) est dit **optimal** s'il vérifie :

- \mathbb{Z} contient tous les parents de Y qui ne sont pas des médiateurs entre A et Y ; et
- \mathbb{Z} contient tous les parents de tout médiateur entre A et Y , qui ne sont pas eux-mêmes des médiateurs entre A et Y .

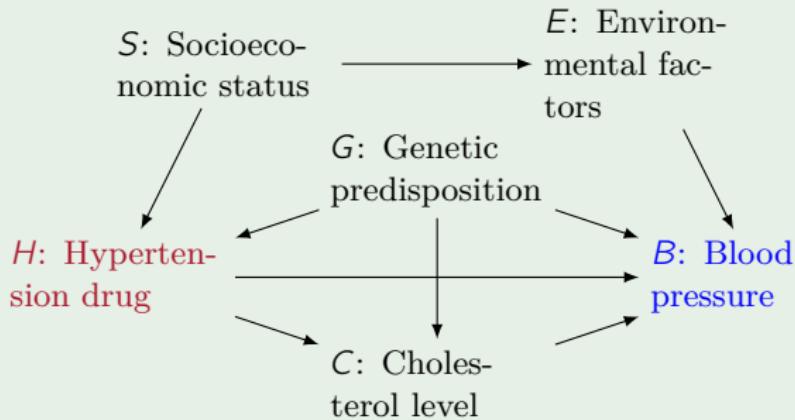


Example

Comment peut-on estimer de façon optimale l'effet causal d'un nouveau traitement antihypertenseur (H) sur la pression artérielle (B) $P(b | do(h))$?

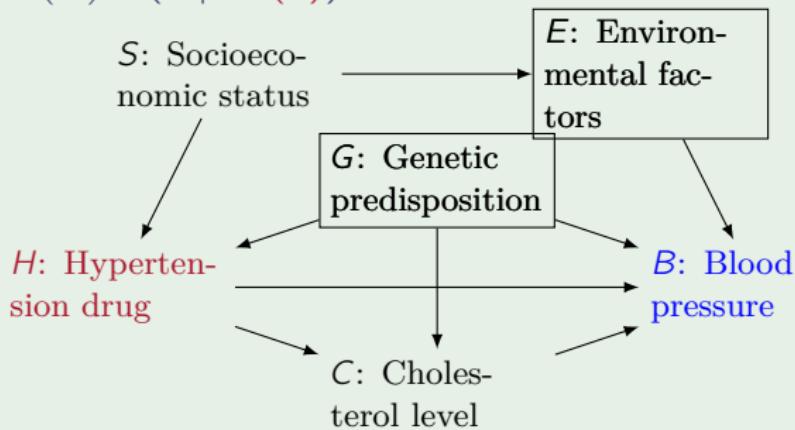
Example

Comment peut-on estimer de façon optimale l'effet causal d'un nouveau traitement antihypertenseur (H) sur la pression artérielle (B) $P(b \mid do(h))$?



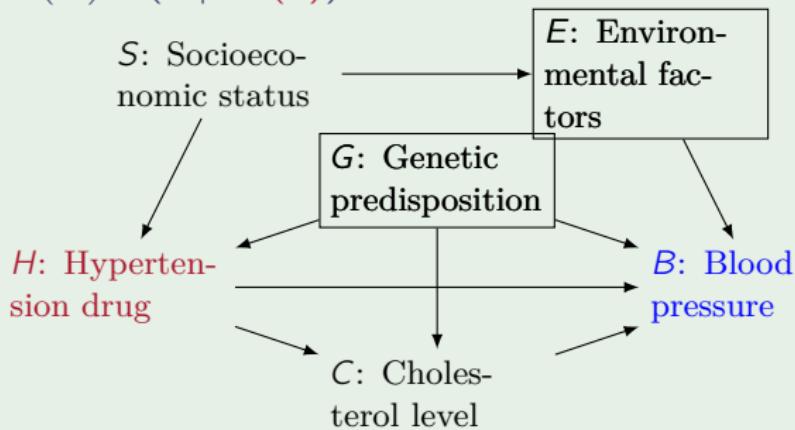
Example

Comment peut-on estimer de façon optimale l'effet causal d'un nouveau traitement antihypertenseur (H) sur la pression artérielle (B) $P(b \mid do(h))$?



Example

Comment peut-on estimer de façon optimale l'effet causal d'un nouveau traitement antihypertenseur (H) sur la pression artérielle (B) $P(b | do(h))$?



$$P(b | do(h)) = \sum_{g,e} P(b|h,g,e)P(g,e)$$

- 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 !
- Tous les ensembles satisfaisant le critère du back-door éliminent le biais de confusion, mais ils ne minimisent pas tous la variance de l'estimation.
- 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.

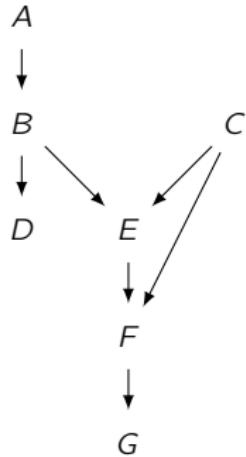
1. Factoriser la probabilité jointe $P(a, b, c, d)$
2. Factoriser la probabilité conditionnelle $P(b, c, d | a)$
3. Factoriser la probabilité interventionnelle $P(b, c, d | \text{do}(a))$

Supposons que le DAG suivant est causal.

1. Factoriser la probabilité jointe $P(a, b, c, d)$
2. Factoriser la probabilité conditionnelle $P(b, c, d | a)$
3. Factoriser la probabilité interventionnelle $P(b, c, d | \text{do}(a))$

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

Exercice 3



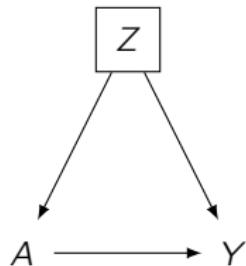
- $B \perp\!\!\!\perp_P G | F?$
- $A \perp\!\!\!\perp_P F | E?$
- $A \perp\!\!\!\perp_P F | C, E?$
- $B \perp\!\!\!\perp_P E | F?$
- $\{A, B, D\} \perp\!\!\!\perp_P C | G?$
- $\{A, B, D\} \perp\!\!\!\perp_P \{C, F, G\}?$
- $\{A, B, D\} \perp\!\!\!\perp_P \{C, F, G\} | E?$

Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?

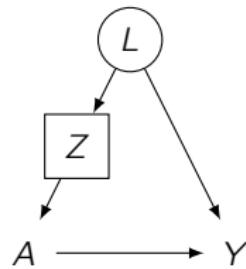
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



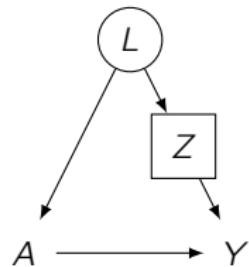
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



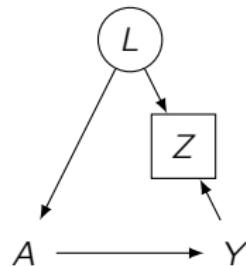
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



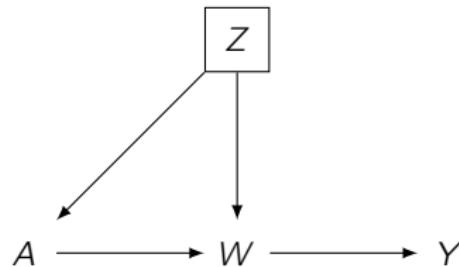
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



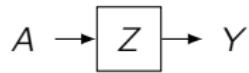
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



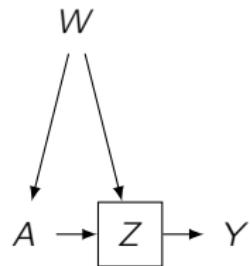
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



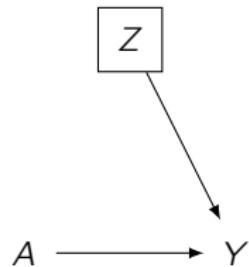
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



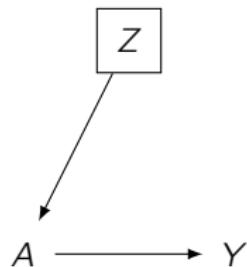
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



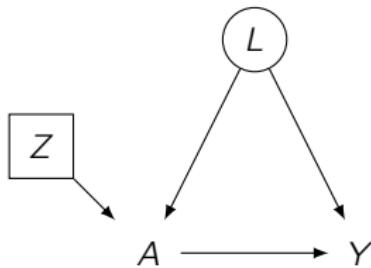
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



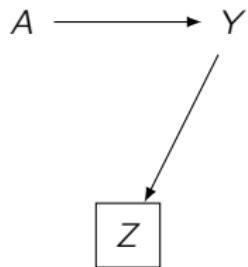
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



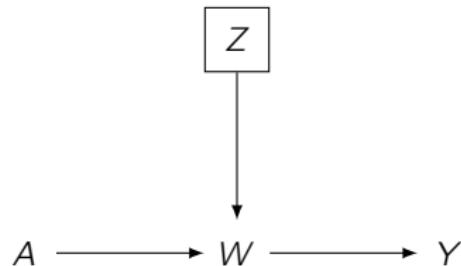
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



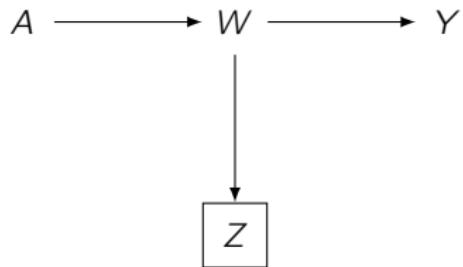
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



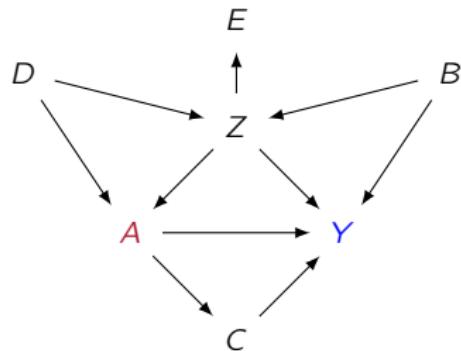
Exercise 4

Est ce que $\{Z\}$ satisfait le critère du back-door pour (A, Y) ?
Est ce qu'il est minimal? Est ce qu'il est optimal?



Exercice 5

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



À partir du papier disponible ici :

<https://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? Est ce qu'il est optimal?
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. Causality: Models, Reasoning and Inference. Cambridge University Press. 2009.
- Leonard Henckel, Emilia Perković, Marloes H. Maathuis. Graphical Criteria for Efficient Total Effect Estimation Via Adjustment in Causal Linear Models. *Journal of the Royal Statistical Society Series B*. 2022.
- 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 Confounding in Epidemiologic Studies Unavoidable Challenges, Uncertain Solutions. *Epidemiology*. 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

