Sciences Economiques & Sociales de la Santé
& Traitement de l'Information Médicale

www.sesstim-orspaca.org

Benoit LEPAGE

Service d'Epidémiologie, CHU de Toulouse
Université de Toulouse III Paul Sabatier
INSERM UMR 1027 (Equipe 5)

Estimation d'un effet causal et graphes acycliques dirigés

mars 2016

Cliquez ici pour voir l'intégralité des ressources associées à ce document

Estimation d'un effet causal et Graphes Acycliques Dirigés

Benoît Lepage (MCU-PH)
Service d'Epidémiologie, CHU Toulouse
Université Toulouse III Paul Sabatier,
18 mars 2016 - Webinars QuanTIM

Plan

I. Intro – rappels causalité en Epidémiologie
II. Un outil : les graphes acycliques dirigés – modèles causaux non paramétriques
    1. Représentation de distributions jointes
    2. Déterminer si un effet causal est identifiable
    3. Inférer des effets causaux à partir de données observationnelles (= oracle)

I. Causalité en Epidémiologie

- Définition simple de la causalité en Epidémiologie :

Une exposition *X* est une cause d'un critère de jugement *Y*

si la modification de *X* entraîne une modification de *Y*

I. Causalité en Epidémiologie

- Dans une étude épidémiologique étiologique,
  on cherche habituellement à évaluer des relations causales
  à partir d'associations statistiques

- Mais on ne peut pas déduire directement des relations de causalité à partir de simples associations

- **Comment déduire des relations de causalité à partir d'associations statistiques ?**

I. Causalité en Epidémiologie

- Déduire de la causalité à partir d'associations ?

  Critères de Sir Austin Bradford Hill (1965) pour discuter de la causalité dans les études non expérimentales :

  - force d'association
  - relation dose-effet

  } **Critères de nature statistique** (peuvent être évalués à partir de données observées)

  - régularité *(consistency)*
  - spécificité de l'association
  - temporalité
  - plausibilité biologique
  - cohérence avec littérature
  - expérimentation/parallélisme
  - analogie

  } **Critères reposant sur des hypothèses causales**
  - ne peuvent s'évaluer à partir des données observées uniquement (sur 1 seule étude)
  - reposent sur le schéma d'étude, les qualités intrinsèques de l'étude, la littérature, les modèles théoriques sous-jacents…

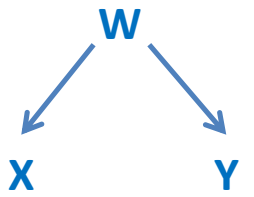
I. Causalité en Epidémiologie

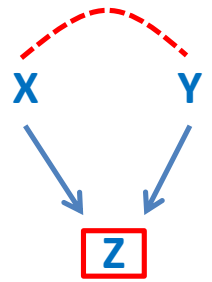
- Association ≠ causalité

  mais une association statistique (dépendance) entre X et Y peut refléter 5 situations dont 4 sont liées à une notion de causalité :

  1. Fluctuations aléatoire *(= pas de causalité)*
  2. X cause Y
  3. Y cause X
  4. X et Y ont une cause commune W
  5. Ajustement sur un effet commun de X et Y, alors que X et Y sont indépendantes marginalement

     Pr(Y|X) = Pr(Y) mais Pr(Y|X,Z) ≠ Pr(Y|Z)





- Ces différentes situations peuvent également se combiner…

I. Causalité en Epidémiologie

D'après Pearl, un frein à l'analyse causale par les statisticiens est l'absence de notations exprimant des hypothèses causales

$\Rightarrow$ Pour garder les idées claires dans une analyse « causale » :

1 Séparation entre concepts statistiques et causaux

Concepts statistiques = définis à partir de la distribution des variables	Concepts causaux = la distribution des variables ne suffit pas pour les définir
Corrélation	Randomisation
Régression	Influence, Effet
Indépendance	Confusion
Indépendance conditionnelle	« spurious correlation»
Association	« path coefficient »
Vraisemblance	Variable instrumentale
Collapsibility	Intervention
Risque relatif, OR	Explication
…	

I. Causalité en Epidémiologie

D'après Pearl, un frein à l'analyse causale par les statisticiens est l'absence de notations exprimant des hypothèses causales

$\Rightarrow$ Pour garder les idées claires dans une analyse « causale » :

2 **Proposer des modèles de causalité qui s'accompagnent de notations pour exprimer les hypothèses causales nécessaires**

- **Notation algébrique du résultat potentiel ou contrefactuel (*potential outcome*)**

  pour un individu *i*, deux résultats sont possibles pour une exposition *X* binaire :

  - $Y_{X=0}(i)$ est le résultat que l'on observerait si $X=0$,
  - $Y_{X=1}(i)$ est le résultat que l'on observerait si $X=1$,
  - Au mieux : un seul des résultats peut être observé

- **Notation graphique exprimant les relations de cause à effet (flèches)**

II. Graphes acycliques dirigés

- **Utilisation formelle des graphiques**

  1) Permet de <u>représenter des distributions jointes</u> de manière simple (par des graphes)

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)p(x_4|x_3, x_2, x_1)p(x_5|x_4, x_3, x_2, x_1)$$

  2) Permet de <u>déduire les associations statistiques qu'impliquent un ensemble d'hypothèses causales</u>

  $\Rightarrow$ en déduire si l'effet causal qui nous intéresse est « identifiable » ou non

  3) Permet <u>d'inférer des effets causaux à partir de données observationnelles</u>

1) Vocabulaire, définitions

Relation causale :
représentée par une simple flèche

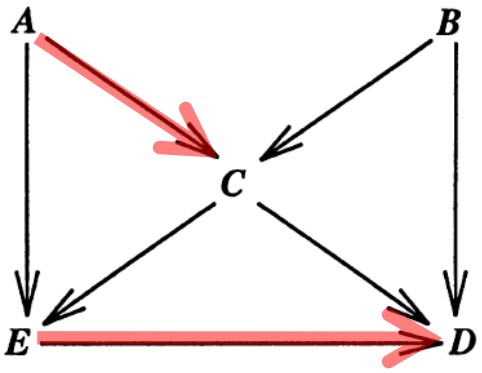


FIGURE 1.

Une flèche =
EFFET CAUSAL
STABLE ET AUTONOME

- absence de relation causale = aucune flèche ++++
- **ATTENTION,**

Ne pas tracer de flèche
est une hypothèse causale plus forte
que d'en tracer une !

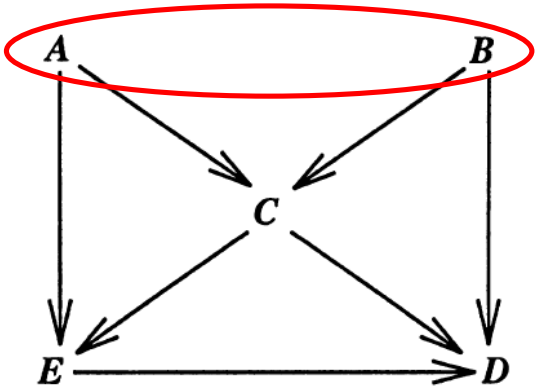


FIGURE 1.

1) Vocabulaire, définitions

Description des variables à l'aide de relations de parentés :

Relations directes ou indirectes :

- A, B et C sont les **ancêtres** (ou cause) de D et E

- E et D sont les **descendants** de A, B et C

Relations directes

- A **parent** de C, C est **l'enfant** de A

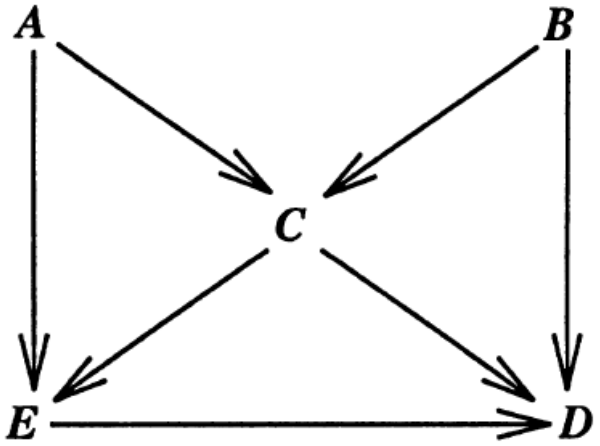


FIGURE 1.

1) Vocabulaire, définitions

- Une flèche bi-directionnelle (2 pointes) reliant deux variables indique la présence d'un ou plusieurs ancêtres communs

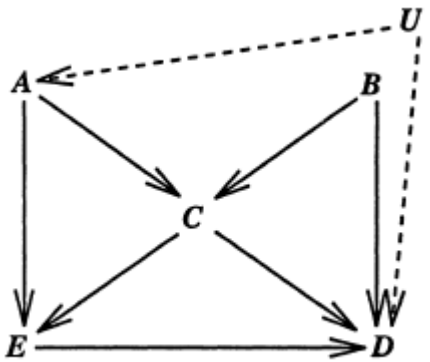


FIGURE 2.

Important +++
Si deux variables d'un DAG partagent une cause commune, **cette cause commune doit apparaître sur le DAG, qu'elle soit mesurée ou non**
- conventions : U=unknown,
pointillés = assocations non mesurées

- Un graphe est **dirigé** si tous les arcs sont des flèches (simples ou doubles)
- Un graphe est **acyclique** si aucun chemin dirigé ne forme de boucle fermée
- *DAG = directed acyclic graph*

  = *graphe acyclique dirigé* ou *graphe acyclique orienté*

1) Vocabulaire, définitions

- **Backdoor path** (en français ? Chemin dérobé ?)

  le chemin de X vers Y est un *backdoor path* s'il y a une flèche qui pointe vers X

Exemples de « back door path » de E vers D

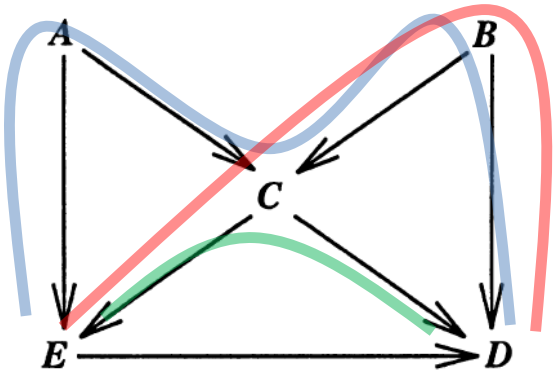


FIGURE 1.

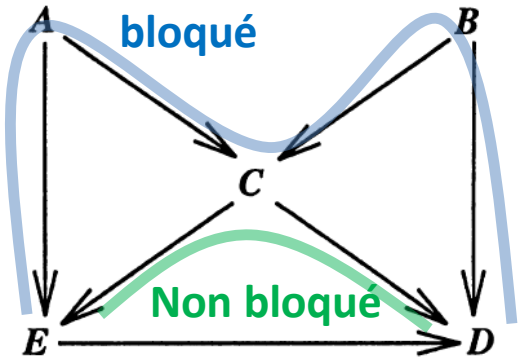


FIGURE 1.

- La variable C est appelée une **collision (*collider*)** **sur le chemin E←A→C ← B → D**
- **La variable C n'est pas une collision sur le chemin E ← C ← B → D, ni sur le chemin E ← C → D**
- Un chemin est **bloqué** quand il comporte au moins une collision, sinon il est **non bloqué (*unblocked*)** ou **ouvert (*open*)**

1) Vocabulaire, définitions

Parfois, on voit deux variables reliées par un trait pointillé :

C'est la représentation de la présence d'une association non causale entre A et B (spurious correlation)

Généralement utilisé pour interpréter un graphe acyclique dirigé

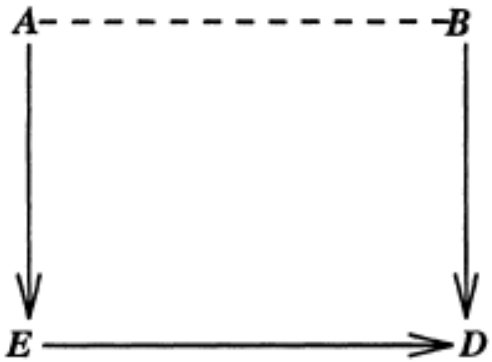
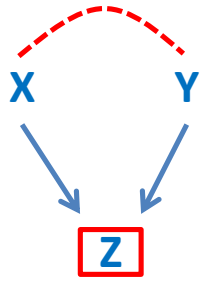


FIGURE 3.

Exemple, si on ajuste sur une collision, on crée une association non causale.



2) Principes théoriques

Utilisation des DAG pour représenter des lois jointes ?

Par exemple, représenter la loi jointe de 5 variables $\{X_1, X_2, X_3, X_4, X_5\}$

Rappel : $p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)}$ donc $p(x_1, x_2) = p(x_1)p(x_2|x_1)$

Généralisation (théorème de la multiplication, *chain rule formula*) :

La probabilité d'observer l'évènement combiné $\{X_1=x_1, X_2=x_2, X_3=x_3, X_4=x_4, X_5=x_5\}$ est égale à :

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)p(x_4|x_3, x_2, x_1)p(x_5|x_4, x_3, x_2, x_1)$$

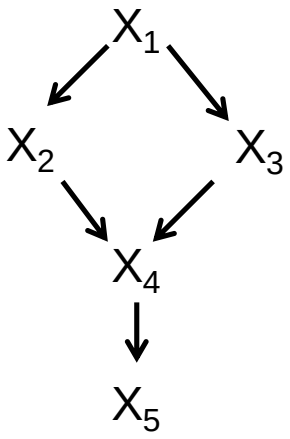
Utilisation des DAG pour représenter des lois jointes ?

On peut représenter une simplification de la distribution jointe
Si certaines indépendances conditionelles sont vraies

$$p(x_1,x_2,x_3,x_4,x_5) = p(x_1)p(x_2|x_1)p(x_3|x_2,x_1)p(x_4|x_3,x_2,x_1)p(x_5|x_4,x_3,x_2,x_1)$$

$$p(x_1,x_2,x_3,x_4,x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2,x_3)p(x_5|x_4)$$

À l'aide du graphique suivant :



$$P(x_1,x_2,x_3,x_4,x_5) = \prod_{j=1}^{5} P(x_j|pa_j)$$

Ce modèle est un modèle **probabiliste** « réseau bayésien »

la notion de causalité n'est pas encore précisée +++

D-séparation

- **D-séparation = critère graphique** qui permet de déduire des relations d'indépendances conditionnelles à partir d'un graphe acyclique dirigé (… à condition que le DAG soit correct)

- Mathématiquement, il y a une correspondance exacte entre la forme simplifiée de la probabilité jointe

$$P(x_1, x_2, \ldots, x_n) = \prod_j P\left(x_j \middle| pa_j\right)$$

- et les ensembles X_1, X_2, …, X_n, qui vérifient les critères de d-séparation dans un DAG

D-séparation

Définition formelle

- Si les hypothèses indiquées sur un DAG sont correctes, **deux variables sont statistiquement indépendantes, conditionnellement à un ensemble de covariables, si chaque chemin entre les deux variables est bloqué**

- Un chemin est bloqué, conditionnellement à un ensemble de variables *Z (« Z bloque le chemin »),* s'il y a une variable *w* sur le chemin répondant à au moins une des 2 possibilités :
  - *w* est une collision et ni *w* ni aucun de ses descendants n'appartient à *Z*
  - *w* n'est pas une collision et est compris dans *Z*
- *Z* « d-sépare » *X* de *Y*, ssi *Z* bloque chaque chemin entre *X* et *Y*

$$\left(X \perp\!\!\!\perp Y | Z\right)_G$$

⇒ **En clair, deux variables X et Y sont indépendantes si les seuls chemins qui les relient sont bloqués**

⇒ **Un chemin est bloqué si :**

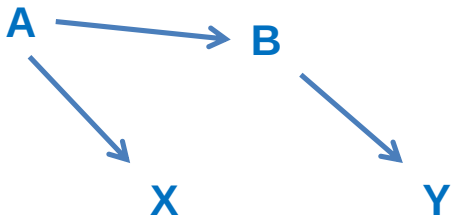
- on ajuste sur une des variables du chemin qui n'est pas une collision

- il contient une collision, et on n'ajuste ni sur la collision, ni sur un descendant de la collision

Note : le terme « conditionnellement » correspond à :
« ajuster », « stratifier », « apparier », « standardiser » , « sélectionner un sous-groupe » ... dans une analyse statistique

Exemple 1 :

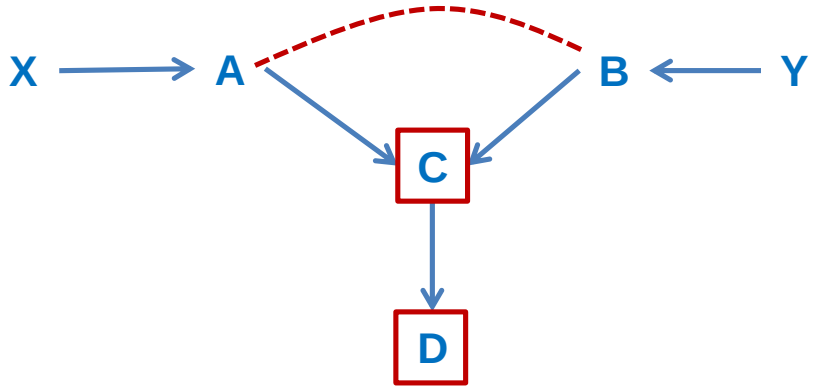
Exemple 2 :



On ne peut pas dire que X et Y sont indépendants, car il existe un chemin non bloqué reliant X à Y

Par contre, en ajustant sur A et/ou sur B, le seul chemin reliant X à Y est bloqué : X est indépendant de Y conditionnellement à (A,B)

Exemple 3 :



Sans aucun ajustement, on peut dire que X et Y sont indépendants,
car le seul chemin qui les relie est bloqué par une collision C

Par contre, en ajustant sur C ou D, on crée une association non causale et un chemin non bloqué :
Conditionnellement à C ou D, on ne peut pas dire que X est indépendant de Y

Application de la d-séparation :

Peut-on identifier un effet causal dans une étude observationnelle ?
Va-t-on être gêné par des **biais de confusion** ? +++

- On peut identifier à partir d'un DAG un ensemble suffisant de covariables pour estimer ou tester l'effet de X sur Y, par la règle du *backdoor criterion*

***Z* est un ensemble suffisant de covariables pour estimer et tester l'effet causal de *X* sur *Y* si :**

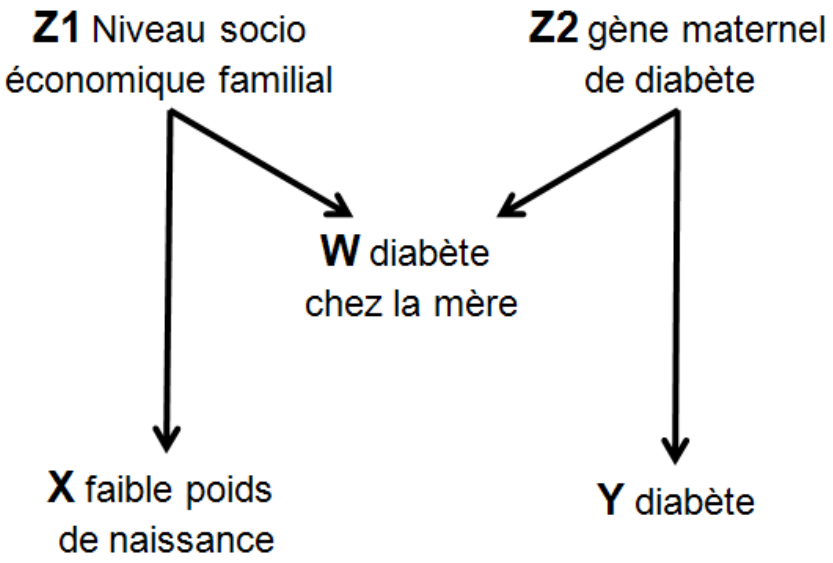
1) **Aucune variable de *Z* n'est un descendant de *X***

ET

2) **Chaque *backdoor path* entre *X* et *Y* est bloqué**

Application de la d-séparation - biais de confusion

- On souhaite évaluer l'effet causal d'un faible poids de naissance (X) sur la survenue d'un diabète à l'âge adulte (Y) (note : Z1 et Z2 sont souvent non mesurées)



Si on regarde les **critères statistiques habituels** :

Si les hypothèses de ce graphe sont exactes :

- X et W sont associées
- Y et W sont associées
- W n'est pas influencé par X ni par Y

Conclusion selon nos critères statistiques habituels ⇒ Il faudrait ajuster sur W

Ce graphe représente l'hypothèse nulle « il n'y a pas de relation causale entre X et Y »

Application de la d-séparation - biais de confusion

Analyse graphique du même problème :

Les règles du backdoor criterion ont abouti à un algorithme permettant de vérifier si un ensemble de variable *S* est suffisant pour évaluer l'effet causal entre *X* et *Y* de manière non confondue *(backdoor test for sufficiency) :*

Pour un ensemble de variables {S_1, …, S_n} qui ne contient aucun descendant de X ni de Y :

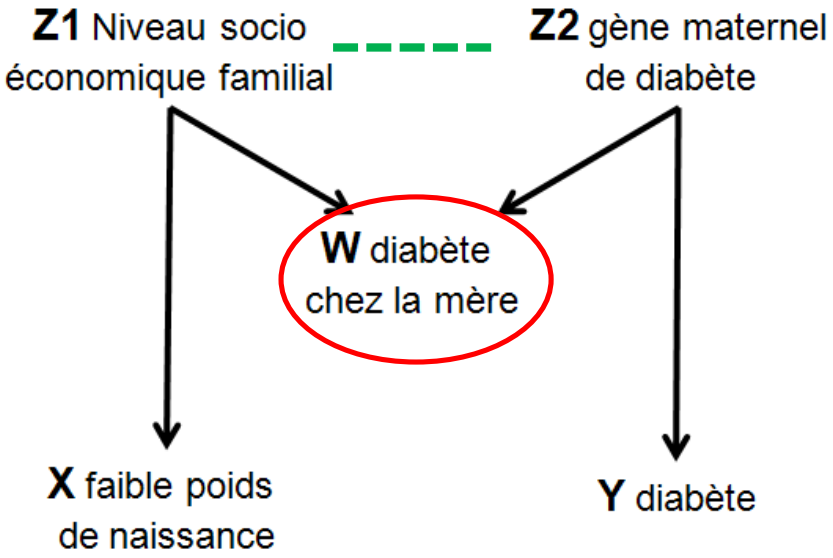
M1 : effacer toutes les flèches qui partent de X

M2 : tracer les segments non dirigés entre chaque paire de variable ayant un enfant ou un descendant commun dans S

M3 : dans le nouveau graphe, chercher s'il y a un chemin non bloqué entre X et Y qui ne passe pas par S.
S'il n'y a pas de chemin non bloqué, S est « suffisant »

Application de la d-séparation - biais de confusion

- Retour à l'exemple : on ajuste sur S = { W }



M1 : effacer toutes les flèches qui partent de X

-> aucune flèche ne part de X

M2 : tracer les segments non dirigés entre chaque paire de variable ayant un enfant ou un descendant commun dans S

-> segments entre Z1 et Z2

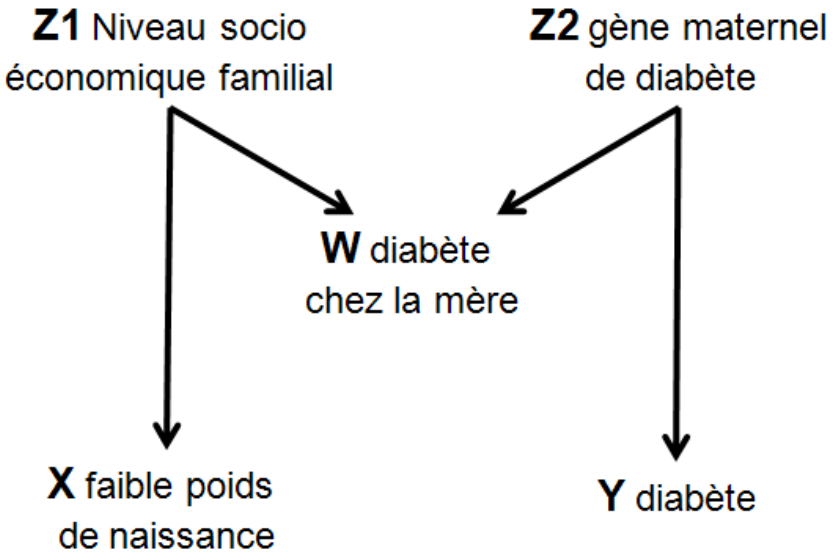
-> il a chemin non bloqué X-Z1-Z2-Y qui ne passe pas par S

=> S n'est pas suffisant pour tester l'effet causal de X sur Y

M3 : dans le nouveau graphe, chercher s'il y a un chemin non bloqué entre X et Y qui ne passe pas par S

Application de la d-séparation - biais de confusion

- Retour à l'exemple : on n'ajuste sur rien, S = { ∅ }



-> il n'y a aucun chemin non bloqué
=> S est suffisant pour tester l'effet causal de X sur Y

M1 : effacer toutes les flèches qui partent de X

-> aucune flèche ne part de X

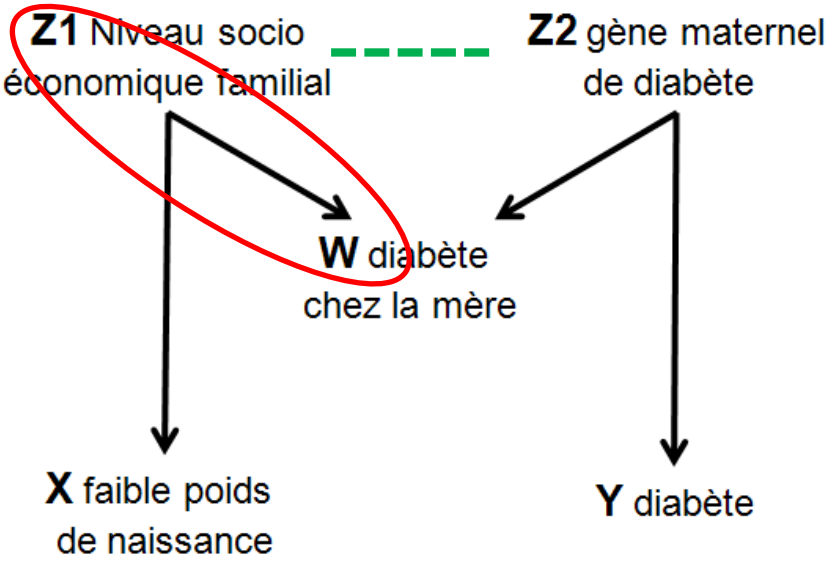
M2 : tracer les segments non dirigés entre chaque paire de variable ayant un enfant ou un descendant commun dans S

-> S ne contient aucune variable, on ne rajoute pas de segment

M3 : dans le nouveau graphe, chercher s'il y a un chemin non bloqué entre X et Y qui ne passe pas par S

Application de la d-séparation - biais de confusion

- Retour à l'exemple : S = { Z1, W }



M1 : effacer toutes les flèches qui partent de X

-> aucune flèche ne part de X

M2 : tracer les segments non dirigés entre chaque paire de variable ayant un enfant ou un descendant commun dans S

-> segments entre Z1 et Z2

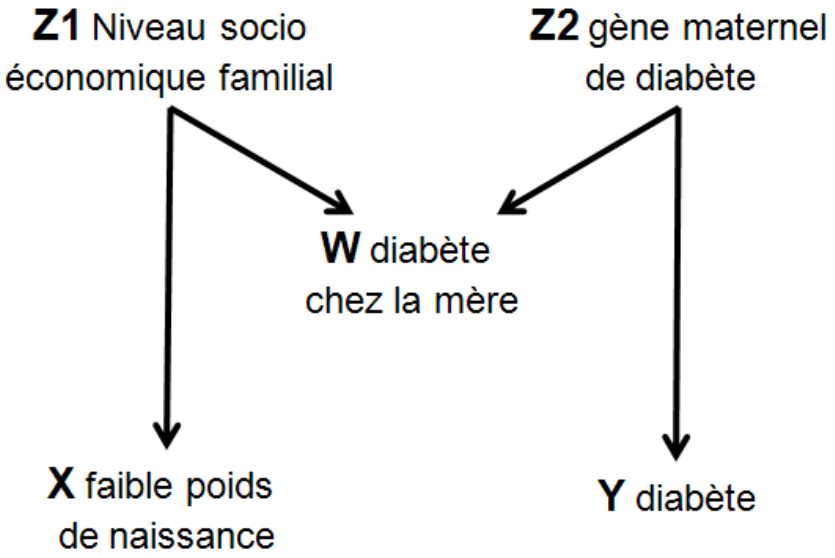
-> le seul chemin non bloqué X-Z1-Z2-Y passe par S

=> S est suffisant pour tester l'effet causal de X sur Y

M3 : dans le nouveau graphe, chercher s'il y a un chemin non bloqué entre X et Y qui ne passe pas par S

Application de la d-séparation - biais de confusion

- Retour à l'exemple :



De la même manière,

S = { W, Z2 }

ou S = { Z1, W, Z2 }

sont également des ensembles suffisants pour tester l'effet causal de X sur Y

=> Il y a plusieurs solutions correctes pour contrôler la confusion

Suffisance minimale = un ensemble S de variables a la suffisance minimale pour l'ajustement si S est suffisant, sans qu'aucun sous ensemble de S ne soit suffisant

Ici, S = { ∅ } a la suffisance minimale pour contrôle la confusion entre X et Y

Application de la d-séparation - biais de confusion

Remarque sur l'étude des biais de confusion à l'aide de DAG :

⇒ il peut donc y avoir différents ensembles suffisants de co-variables pour contrôler une confusion
toutes aussi valides les unes que les autres

⇒ on ne parle plus de « facteur de confusion », mais d'associations confondues ou non…

⇒ On peut définir un ensemble minimal suffisant pour contrôler une confusion

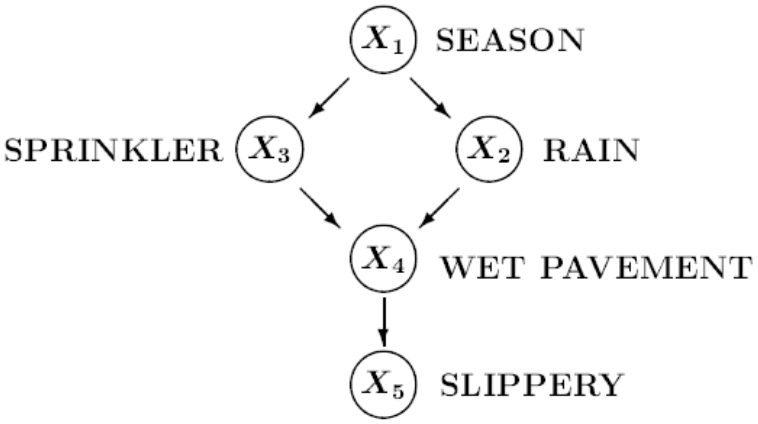
⇒ Un ajustement peut être inutile

⇒ Un ajustement peut être dangereux (ajustement sur collisions ou descendant de collision)

3) Inférer des effets causaux à partir de données observationnelles (=oracle)

Exemple :

on part de l'hypothèse représentée graphiquement par :



qui correspond à la probabilité jointe

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_4)$$

Pour passer d'un modèle probabiliste à un modèle causal :

on ajoute l'hypothèse suivante :
les relations parents-enfants sont des mécanismes…

- **stables** = la relation fonctionnelle est constante dans le temps et selon les circonstances

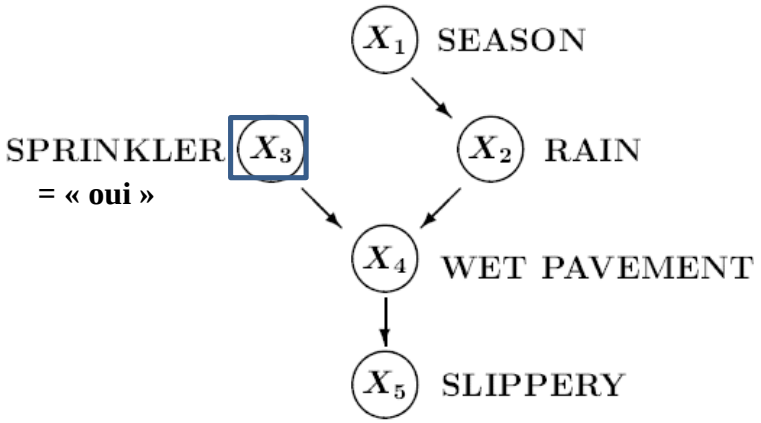
- **et autonomes** = une modification locale d'une partie du graphe n'influence pas le reste de la structure

3) Inférer des effets causaux à partir de données observationnelles (=oracle)

On peut modéliser une intervention factice directement sur le graphique et les équations structurelles :

Exemple, **l'arrosage est toujours allumé** : X1 n'a plus d'effet sur X3

Graphiquement, on supprime simplement la flèche $X_1 \rightarrow X_2$:



Avant l'intervention, on avait :

$$p(x_1, x_2, x_3, x_4, x_5)$$
$$= p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_4)$$

Après intervention, les équations structurelles deviennent :

$$p(x_1, x_2, x_3, x_4, x_5 | \text{do}(X_3 = \text{"oui"}))$$
$$= p(x_1)p(x_2|x_1)\underbrace{p(X_3 = \text{"oui"})}(x_4|x_2, X_3 = \text{"oui"})p(x_5|x_4)$$

= 1 (arrosage allumé dans 100% des cas)

$$= p(x_1)p(x_2|x_1)(x_4|x_2, X_3 = \text{"oui"})p(x_5|x_4)$$

3) Inférer des effets causaux à partir de données observationnelles (= oracle)

On peut ainsi estimer de 2 potential outcomes (contrefaits non observés directement dans nos données) et les comparer pour calculer un effet causal :

{ Probabilité que la chaussée soit glissante si l'arrosage est allumé }

versus

{ Probabilité que la chaussée soit glissante si l'arrosage est fermé }

$$p\left(x_5 = 1 \middle| \text{do}(X_3 = \text{"oui"})\right) - p\left(x_5 = 1 \middle| \text{do}(X_3 = \text{"non"})\right)$$

$$= \sum_{x_1, x_2, x_4} p(x_1) p\left(x_2 \middle| x_1\right) p\left(x_4 \middle| x_2, X_3 = 1\right) p\left(x_5 = 1 \middle| x_4\right)$$

$$- \sum_{x_1, x_2, x_4} p(x_1) p\left(x_2 \middle| x_1\right) p\left(x_4 \middle| x_2, X_3 = 0\right) p\left(x_5 = 1 \middle| x_4\right)$$

Cette estimation d'un potential outcome correspond à la « G-computation » ou « G-formula » ou « Truncated factorization »

Conclusions

Intérêts de cette approche graphique :

1) Exposer de manière explicite les hypothèses de causalité sous-jacentes à une interprétation causale de résultats dans une étude observationnelle

On peut également mener des analyses de sensibilité en proposant différentes structures causales possibles

Recentre la question autour d'une exposition d'intérêt et d'un critère de jugement

Conclusions

Intérêts de cette approche graphique :

2) Cette représentation correspond à une représentation formelles de lois de probabilités jointes

En appliquant les règles d'utilisation, comme la D-séparation, on peut :

- En déduire des relations d'indépendance conditionnellement à un ensemble de variable d'ajustement ou de sélection…
- Déterminer si un effet causal peut être identifié (et donc pourrait être estimé sans biais) conditionnellement à un ensemble de variable d'ajustement ou de sélection…

Conclusions

Intérêts de cette approche graphique :

3) Sous l'hypothèse de relations stables et autonomes, on peut faire des estimations d'effets causaux complexes à partir de données observationnelles (non randomisées)

Ce modèle causal (potential outcomes, DAG, modèle non paramétrique) est directement associé à certaines méthodes d'estimation développées ces dernières années :

- Scores de propension
- Analyses de la médiation
- Analyse d'une exposition qui varie au cours du temps
- Utilisation des variables instrumentales

Références générales :

Pour l'inférence causale en statistique :
Livre de J Pearl: Causality, 2de edition (2009), Models, reasoning and inference.

Article qui résume le livre :
Pearl J. An introduction to causal inference. The international journal of biostatistics 2010;6(2):article 7.

Greenland S, et al. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37-48.

Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco: Jossey-Bass;2006. p.393-428.

Pearl J. Causal Diagrams for Empirical Research. Biometrika 1995;82(4):669-688.

Livre sur les méthodes d'estimations
Van Der Laan MJ, Rose S. Targeted Learning : Causal Inference for Observational and Experimental Data. New York : Springer Publishing Company ; 2011

Utilisation des graphes acycliques : Biais en Epidémiologie

Biais de sélection

• Hernán MA, et al. A structural approach to selection bias. *Epidemiology* 2004*; 15:615–625.*

Biais de mesure

• Hernán MA. Invited commentary: causal diagrams and measurement bias. Am J Epidemiol 2009;170:959-962.

• Shahar E. Causal diagrams for encoding and evaluation of information bias. J Eval Clin Pract 2009;15(3):436-40.

Evolution avant-après

• Glymour MM, et al. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol 2005;162(3):267-78.

• Lepage B et al. Estimating the causal effect of an exposure on change from baseline using DAG and path analysis. Epidemiology 2015.

Données manquantes

• Daniel R et al. Using causal diagrams to guide analysis in missing data problems. Stat methods med res 2012;21(3):243-56.