

Introduction à l'inférence causale et aux diagrammes causaux

Dans le cadre du Club Math du 8 février 2023

Janie Coulombe
janie.coulombe@umontreal.ca

Professeure adjointe

Département de mathématiques et de statistique
Université de Montréal

Plan

Exemple (exagéré)

Confusion ou paradoxe de Simpson

Étude randomisée vs observationnelle

Définir, postuler, et estimer

Diagramme causal

Exemple (exagéré)

Situation

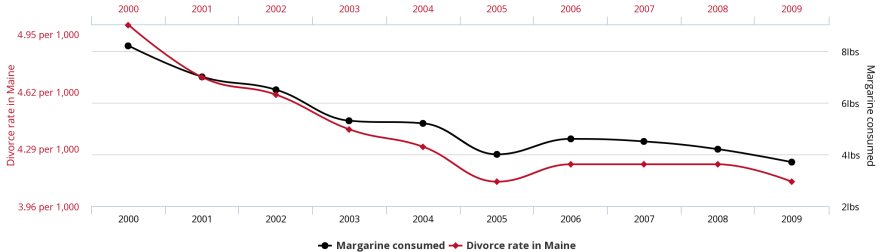
Un ami vous parle de son couple qui bat de l'aile.

Situation

Un ami vous parle de son couple qui bat de l'aile.

Après quelques minutes de recherche en ligne...

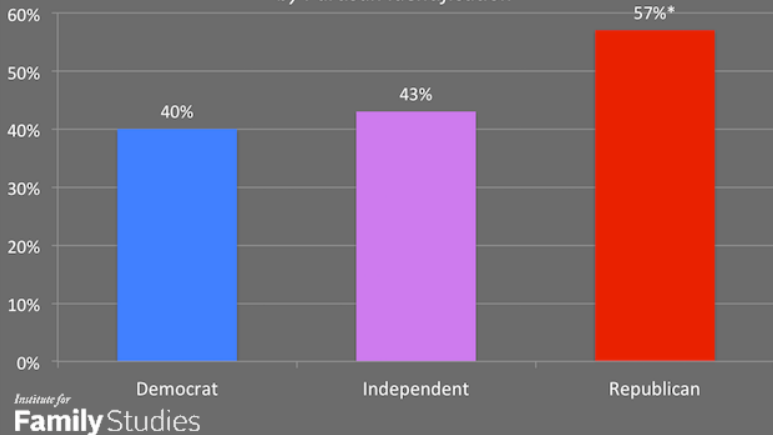
Divorce rate in Maine
correlates with
Per capita consumption of margarine



tylervigen.com

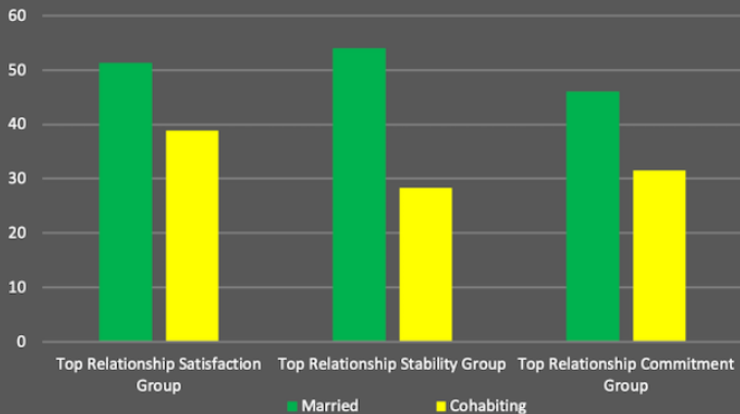
Source: <https://www.tylervigen.com/spurious-correlations>. Données de la National Vital Statistics Reports and U.S. Department of Agriculture

Figure A1: Percent of Americans Ages 20 to 60
Who Are Married
by Partisan Identification



Source: <https://ifstudies.org/blog/red-families-vs-blue-families-which-are-happier>.

Relationship Quality by Current Relationship Status



Source: National Marriage Project/Wheatley Institution Analysis of December 2018 YouGov "iFidelity Survey"

Intervention: Vous suggérez à votre ami de diminuer sa consommation de margarine, de voter plus conservateur et de se marier au plus vite.

Malheureusement, votre ami se sépare dans les semaines qui suivent.

Malheureusement, votre ami se sépare dans les semaines qui suivent.

“HA! Il faisait sûrement partie des exceptions? Ces graphiques ne nous montrent que des **effets** moyens”.

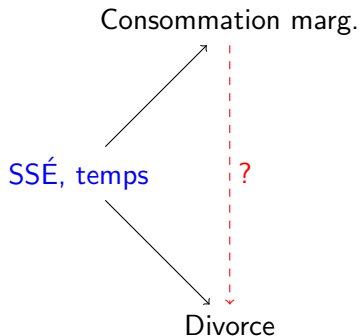
Malheureusement, votre ami se sépare dans les semaines qui suivent.

“HA! Il faisait sûrement partie des exceptions? Ces graphiques ne nous montrent que des **effets** moyens”.

“L'intervention suggérée n'était pas efficace puisque les “effets” observés ne sont pas causaux. Ils s'agiraient plutôt d'**associations**.”

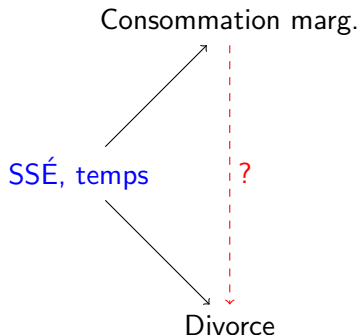
Confusion ou paradoxe de Simpson

Dans le cas de la margarine, on peut supposer:



Le statut socioéconomique (SSÉ), le temps, ou encore les deux sont peut-être des **facteurs confondants**.

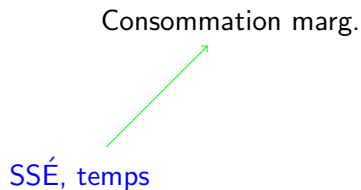
Dans le cas de la margarine, on peut supposer:



Le statut socioéconomique (SSÉ), le temps, ou encore les deux sont peut-être des **facteurs confondants**.

Il n'existe pas nécessairement **d'effet causal** ici.

La confusion

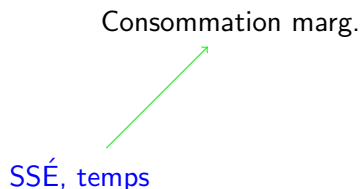


Divorce

Un facteur confondant

- Est associé avec l'exposition (traitement, intervention, etc.)

La confusion

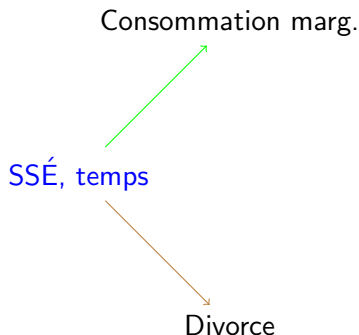


Divorce

Un facteur confondant

- ▶ Est **associé avec l'exposition** (traitement, intervention, etc.)
- ▶ Vient avant le traitement (peut être un parent ou un ancêtre dans le diagramme causal)

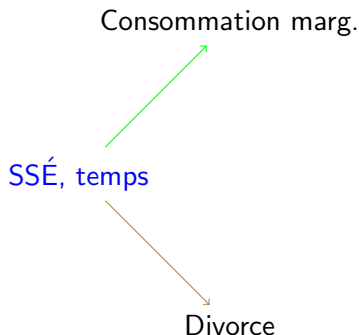
La confusion



Un facteur confondant

- ▶ Est **associé avec l'exposition** (traitement, intervention, etc.)
- ▶ Vient avant le traitement (peut être un parent ou un ancêtre dans le diagramme causal)
- ▶ **Affecte le résultat.**

La confusion



Un facteur confondant

- ▶ Est **associé avec l'exposition** (traitement, intervention, etc.)
- ▶ Vient avant le traitement (peut être un parent ou un ancêtre dans le diagramme causal)
- ▶ **Affecte le résultat.**
- ▶ Cela crée ce qu'on appelle un trajet de porte arrière.

En santé...

Utiliserez-vous la même stratégie pour choisir **quel médicament prescrire** à un patient?

En santé...

Utiliserez-vous la même stratégie pour choisir **quel médicament prescrire** à un patient?

En santé, l'inférence sur les effets causaux de diverses interventions est très importante.

En santé...

Utiliseriez-vous la même stratégie pour choisir **quel médicament prescrire** à un patient?

En santé, l'inférence sur les effets causaux de diverses interventions est très importante.

Pourquoi?

Étude randomisée vs observationnelle

La norme de référence

La meilleure façon d'estimer un effet causal est de planifier un **essai randomisé contrôlé** conçu pour votre question de recherche causale.

La norme de référence

La meilleure façon d'estimer un effet causal est de planifier un **essai randomisé contrôlé** conçu pour votre question de recherche causale.

Dans ce cas, vous contrôlez plusieurs des paramètres de l'étude et, par le fait même, **vous évitez des biais comme le biais de confusion**, le biais de sélection, etc.

Exemple:

Nous concevons une étude où chaque patient sera randomisé à une consommation de 0 cuillère à thé de margarine par semaine ou de 15 cuillères à thé de margarine par semaine.

Exemple:

Nous concevons une étude où chaque patient sera randomisé à une consommation de 0 cuillère à thé de margarine par semaine ou de 15 cuillères à thé de margarine par semaine.

Grâce à la randomisation, les deux groupes seront similaires dans leurs caractéristiques en début d'étude.

Exemple:

Nous concevons une étude où chaque patient sera randomisé à une consommation de 0 cuillère à thé de margarine par semaine ou de 15 cuillères à thé de margarine par semaine.

Grâce à la randomisation, les deux groupes seront similaires dans leurs caractéristiques en début d'étude.

Ensuite, nous observons les patients pendant 15 ans et notons les divorces dans le temps.

Exemple:

Nous concevons une étude où chaque patient sera randomisé à une consommation de 0 cuillère à thé de margarine par semaine ou de 15 cuillères à thé de margarine par semaine.

Grâce à la randomisation, les deux groupes seront similaires dans leurs caractéristiques en début d'étude.

Ensuite, nous observons les patients pendant 15 ans et notons les divorces dans le temps.

Pourquoi pas?

Étude observationnelle

“Une étude observationnelle, à la différence d'une étude randomisée, est généralement entreprise lorsqu'il est impossible, peu pratique ou non éthique d'avoir un groupe témoin.”

Étude observationnelle

“Une étude observationnelle, à la différence d'une étude randomisée, est généralement entreprise lorsqu'il est impossible, peu pratique ou non éthique d'avoir un groupe témoin.”

“Ce genre d'étude est utile pour générer des hypothèses qui peuvent être vérifiées rigoureusement dans un essai contrôlé randomisé. Son principal inconvénient : il est impossible de supposer que les participants sont représentatifs de l'ensemble des personnes atteintes de la maladie étudiée.”

Source: IRSC (<https://cihr-irsc.gc.ca/>).

Exemple de confusion

- ▶ Étude utilisant les dossiers médicaux électroniques pour évaluer l'effet causal des antidépresseurs sur les risques de dépression majeure (DM)

Exemple de confusion

- ▶ Étude utilisant les dossiers médicaux électroniques pour évaluer l'effet causal des antidépresseurs sur les risques de dépression majeure (DM)
- ▶ Les patients qui ont reçu des antidépresseurs sont différents de ceux qui n'en recevaient pas (ex., proportion d'hommes, âges moyens, SSÉ, etc.)

Exemple de confusion

- ▶ Étude utilisant les dossiers médicaux électroniques pour évaluer l'effet causal des antidépresseurs sur les risques de dépression majeure (DM)
- ▶ Les patients qui ont reçu des antidépresseurs sont différents de ceux qui n'en recevaient pas (ex., proportion d'hommes, âges moyens, SSÉ, etc.)
- ▶ Il n'y a jamais eu de randomisation

Exemple de confusion

- ▶ Étude utilisant les dossiers médicaux électroniques pour évaluer l'effet causal des antidépresseurs sur les risques de dépression majeure (DM)
- ▶ Les patients qui ont reçu des antidépresseurs sont différents de ceux qui n'en recevaient pas (ex., proportion d'hommes, âges moyens, SSÉ, etc.)
- ▶ Il n'y a jamais eu de randomisation
- ▶ On ne peut pas comparer simplement les taux de DM sans tenir compte des facteurs confondants
- ▶ On a besoin de “standardiser” selon les caractéristiques des deux groupes

Autre exemple d'actualité - Le paradoxe de Simpson

Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects

Julius von Kügelgen*, Luigi Gresele*, Bernhard Schölkopf

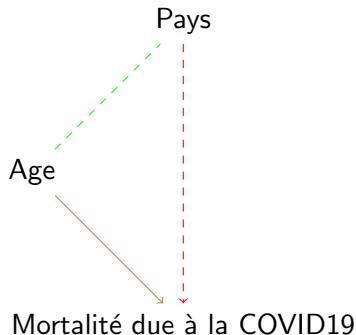
Abstract—We point out an instantiation of Simpson's paradox in Covid-19 case fatality rates (CFRs): comparing a large-scale study from China (17 Feb) with early reports from Italy (9 Mar), we find that CFRs are lower in Italy for every age group, but higher overall. This phenomenon is explained by a stark difference in case demographic between the two countries. Using this as a motivating example, we introduce basic concepts from mediation analysis and show how these can be used to quantify different direct and indirect effects when assuming a coarse-grained causal graph involving country, age, and case fatality. We curate an age-stratified CFR dataset with >750k cases and conduct a case study, investigating total, direct, and indirect (age-mediated) causal effects between different countries and at different points in time. This allows us to separate age-related effects from others unrelated to age and facilitates a more transparent comparison of CFRs across countries at different stages of the Covid-19 pandemic. Using longitudinal data from Italy, we discover a sign reversal of the direct causal effect in mid-March which temporally aligns with the reported collapse of the healthcare system in parts of the country. Moreover, we find that direct and indirect effects across 132 pairs of countries are only weakly correlated, suggesting that a country's policy and case demographic may be largely unrelated. We point out limitations and extensions for future work, and, finally, discuss the role of causal reasoning in the broader context of using AI to combat the Covid-19 pandemic.

Comparaison Italie-Chine

Mesure: mortalité due
à la COVID19.

Résultat: Mortalité
plus faible en Italie
pour chaque tranche
d'âge (séparément)
mais, globalement,
l'Italie a des taux plus
élevés de mortalité
que la Chine.

Effet causal de la stratégie pour combattre le virus



Il n'est pas clair que l'Italie avait une moins bonne stratégie pour ralentir le virus mais les gens y étaient plus âgés. La comparaison par tranche d'âge est possiblement plus juste (standardisation).

Définir, postuler, et estimer

Le paramètre: Un effet causal

On a besoin d'un cadre pour l'estimation d'un effet causal.

Le paramètre: Un effet causal

On a besoin d'un cadre pour l'estimation d'un effet causal.

Comment concevoir un effet causal et comment l'estimer?

Le paramètre: Un effet causal

On a besoin d'un cadre pour l'estimation d'un effet causal.

Comment concevoir un effet causal et comment l'estimer?

On supposera un choix entre deux traitements ($A = 1$ ou $A = 0$).





Essentiellement, que se passe-t-il lorsque la seule chose qui change entre les deux mondes est le traitement?

Le paramètre: Un effet causal (2)

Résultats contrefactuels: Y_{1i} et Y_{0i} pour le patient i (Neyman 1923, Rubin 1974).

Le paramètre: Un effet causal (2)

Résultats contrefactuels: Y_{1i} et Y_{0i} pour le patient i (Neyman 1923, Rubin 1974).

Effet causal individuel défini par $Y_{1i} - Y_{0i}$.

Le paramètre: Un effet causal (2)

Résultats contrefactuels: Y_{1i} et Y_{0i} pour le patient i (Neyman 1923, Rubin 1974).

Effet causal individuel défini par $Y_{1i} - Y_{0i}$.

En réalité, pour ce patient, on observera seulement l'un des deux résultats. L'effet ci-haut n'est pas identifiable.

Le paramètre: Un effet causal (2)

Résultats contrefactuels: Y_{1i} et Y_{0i} pour le patient i (Neyman 1923, Rubin 1974).

Effet causal individuel défini par $Y_{1i} - Y_{0i}$.

En réalité, pour ce patient, on observera seulement l'un des deux résultats. L'effet ci-haut n'est pas identifiable.

On se tourne souvent vers l'effet marginal (moyen) causal:

$$\mathbb{E}[Y_{1i} - Y_{0i}].$$

Données disponibles

Données observées/disponibles:

A_i	X_{1i}	X_{2i}	Y_i	Y_{0i}	Y_{1i}
1	53	124	145		145
1	19	108	218		218
0	32	120	190	190	
1	56	122	138		138
0	65	135	234	234	

Données disponibles

Données observées/disponibles:

A_i	X_{1i}	X_{2i}	Y_i	Y_{0i}	Y_{1i}
1	53	124	145		145
1	19	108	218		218
0	32	120	190	190	
1	56	122	138		138
0	65	135	234	234	

Quel estimateur pour $\mathbb{E}[Y_{1i} - Y_{0i}]$?

Hypothèses causales

1. Interchangeabilité conditionnelle:

$$\{Y_1, Y_0\} \perp A | \mathbf{X}$$

pour A le traitement et \mathbf{X} des facteurs confondants.

Hypothèses causales

1. Interchangeabilité conditionnelle:

$$\{Y_1, Y_0\} \perp A | \mathbf{X}$$

pour A le traitement et \mathbf{X} des facteurs confondants.

Interprétation: Une fois qu'on contrôle pour \mathbf{X} , un “traité” ($A = 1$) ou un “non-traité” ($A = 0$) sont interchangeables.

Hypothèses causales

1. Interchangeabilité conditionnelle:

$$\{Y_1, Y_0\} \perp A | \mathbf{X}$$

pour A le traitement et \mathbf{X} des facteurs confondants.

Interprétation: Une fois qu'on contrôle pour \mathbf{X} , un “traité” ($A = 1$) ou un “non-traité” ($A = 0$) sont interchangeables.

En gros, \mathbf{X} contient tous les facteurs confondants qui créent des trajets de porte arrière problématiques.

2. Positivité du traitement:

$$0 < P(A = a|\mathbf{X}) < 1$$

pour tout a .

Interprétation: Tous les patients utilisés dans le calcul de l'effet causal marginal doivent avoir une chance de recevoir tous les niveaux de traitement possibles.

2. Positivité du traitement:

$$0 < P(A = a|\mathbf{X}) < 1$$

pour tout a .

Interprétation: Tous les patients utilisés dans le calcul de l'effet causal marginal doivent avoir une chance de recevoir tous les niveaux de traitement possibles.

Rappel: On essaie d'évaluer la différence moyenne du résultat si tout le monde recevait $A = 1$ vs si tout le monde recevait $A = 0$.

3. Cohérence du résultat Y , aussi appelé *consistency*, signifie

$$Y_i = A_i Y_{1i} + (1 - A_i) Y_{0i}.$$

Interprétation: Le traitement est bien défini de sorte que le résultat observé correspond bel et bien à un des deux résultats potentiels.

3. Cohérence du résultat Y , aussi appelé *consistency*, signifie

$$Y_i = A_i Y_{1i} + (1 - A_i) Y_{0i}.$$

Interprétation: Le traitement est bien défini de sorte que le résultat observé correspond bel et bien à un des deux résultats potentiels.

Manque de cohérence: J'observe Y le divorce (oui/non) sous 14 cuillères de margarine $\neq Y_1$ le résultat potentiel sous 15 cuillères à thé.

Estimation

Grâce aux hypothèses causales...

$$\begin{aligned}\mathbb{E}[Y_{i1} - Y_{i0}] &= \mathbb{E}[Y_{i1}] - \mathbb{E}[Y_{i0}] \\ &= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_{i1}|\mathbf{X}_i]] - \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_{i0}|\mathbf{X}_i]] \quad \text{espérance itérée} \\ &= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_{i1}|A_i = 1, \mathbf{X}_i]] - \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_{i0}|A_i = 0, \mathbf{X}_i]] \\ &\quad \text{(Interchangeabilité cond. et la positivité)} \\ &= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]] - \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 0, \mathbf{X}_i]] \quad \text{(cohérence)}.\end{aligned}$$

Estimation (2)

$$\begin{aligned}\mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]] &= \sum_{\mathbf{X}} \mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]P(\mathbf{X} = \mathbf{X}_i) \\&= \sum_{\mathbf{X}} \frac{P(A_i = 1|\mathbf{X}_i)}{P(A_i = 1|\mathbf{X}_i)} \mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]P(\mathbf{X} = \mathbf{X}_i) \\&= \sum_{\mathbf{X}} \frac{1}{P(A_i = 1|\mathbf{X}_i)} \mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]P(A_i = 1|\mathbf{X}_i)P(\mathbf{X} = \mathbf{X}_i) \\&= \mathbb{E}\left[\frac{\mathbb{I}(A_i = 1)Y_i}{P(A_i = 1|\mathbf{X}_i)}\right]\end{aligned}$$

(similairement $\mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 0, \mathbf{X}_i]] = \mathbb{E}\left[\frac{\mathbb{I}(A_i=0)Y_i}{P(A_i=0|\mathbf{X}_i)}\right]$).

Estimation (3)

Cela implique deux méthodes pour estimer l'effet causal marginal:

- Estimer

$$\mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 1, \mathbf{X}_i]] - \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i|A_i = 0, \mathbf{X}_i]]$$

- ou estimer

$$\mathbb{E} \left[\frac{\mathbb{I}(A_i = 1) Y_i}{P(A_i = 1|\mathbf{X}_i)} \right] - \mathbb{E} \left[\frac{\mathbb{I}(A_i = 0) Y_i}{P(A_i = 0|\mathbf{X}_i)} \right].$$

La première méthode s'appelle le *calcul-G* (Robins, 1986) et la deuxième méthode s'appelle la *pondération par poids inverses à la probabilité de traitement* (Horvitz et Thompson 1952; Rosenbaum et Rubin 1983; Rosenbaum, 1987).

Le calcul-G peut être vu comme une méthode de standardisation.

Diagramme causal

Les diagrammes causaux

- ▶ Représentent les relations (assumées) entre les variables.

Les diagrammes causaux

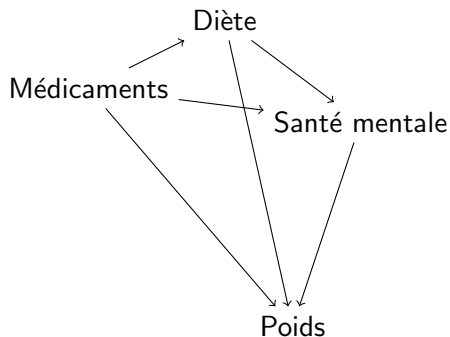
- ▶ Représentent les relations (assumées) entre les variables.
- ▶ Se basent sur les connaissances de fond sur le sujet (mécanisme de prescription, mécanismes biologiques) et non pas sur des connaissances statistiques.

Les diagrammes causaux

- ▶ Représentent les relations (assumées) entre les variables.
- ▶ Se basent sur les connaissances de fond sur le sujet (mécanisme de prescription, mécanismes biologiques) et non pas sur des connaissances statistiques.
- ▶ Nous permettent de vérifier l'interchangeabilité conditionnelle (trouver un ensemble \mathbf{X} suffisant) pour ensuite estimer l'effet causal de A sur Y .

Les diagrammes causaux (2)

Exemple simple de DAG (**D**irected, **A**cyclic, **G**raph):



Les diagrammes causaux (3)

- Un trajet sur lequel toutes les flèches pointent dans la même direction est appelé un trajet causal.

Causal $A \longrightarrow B \longrightarrow C \longrightarrow D$

Pas causal $A \longrightarrow B \longleftarrow C \longrightarrow D$

Pas causal $A \longleftarrow B \longrightarrow C \longrightarrow D$

Les diagrammes causaux (3)

- Un trajet sur lequel toutes les flèches pointent dans la même direction est appelé un trajet causal.

Causal $A \longrightarrow B \longrightarrow C \longrightarrow D$

Pas causal $A \longrightarrow B \longleftarrow C \longrightarrow D$

Pas causal $A \longleftarrow B \longrightarrow C \longrightarrow D$

- Un trajet bloqué par un facteur de collision vs un trajet ouvert (dépendance). Exemples:

Ouvert entre A et D : $A \longrightarrow B \longrightarrow C \longrightarrow D$

Ouvert entre A et D : $A \longrightarrow D$

Ouvert entre A et D : $A \longleftarrow C \longrightarrow D$

Fermé entre A et D : $A \longrightarrow C \longleftarrow D$

Fermé entre A et D : $A \longrightarrow C \longleftarrow B \longrightarrow E \longrightarrow D$

Les diagrammes causaux (3)

- Un trajet sur lequel toutes les flèches pointent dans la même direction est appelé un trajet causal.

Causal $A \longrightarrow B \longrightarrow C \longrightarrow D$

Pas causal $A \longrightarrow B \longleftarrow C \longrightarrow D$

Pas causal $A \longleftarrow B \longrightarrow C \longrightarrow D$

- Un trajet bloqué par un facteur de collision vs un trajet ouvert (dépendance). Exemples:

Ouvert entre A et D : $A \longrightarrow B \longrightarrow C \longrightarrow D$

Ouvert entre A et D : $A \longrightarrow D$

Ouvert entre A et D : $A \longleftarrow C \longrightarrow D$

Fermé entre A et D : $A \longrightarrow C \longleftarrow D$

Fermé entre A et D : $A \longrightarrow C \longleftarrow B \longrightarrow E \longrightarrow D$

- Conditionner sur un facteur de collision débloquent un trajet alors que conditionner sur un facteur de confusion bloque son trajet.

Les diagrammes causaux (4)

Une distribution \mathcal{P} est *Markov-compatible* avec le diagramme causal si elle admet une factorisation utilisant la règle de dérivation en chaîne de type *parents-Markov* $P(\mathbf{U}) = \prod_i P(\mathbf{U}_i | \mathbf{PA}_i)$ relative au diagramme.

Les diagrammes causaux (4)

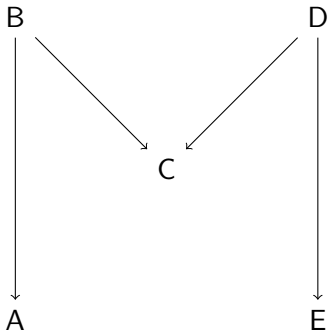
Une distribution \mathcal{P} est *Markov-compatible* avec le diagramme causal si elle admet une factorisation utilisant la règle de dérivation en chaîne de type *parents-Markov* $P(\mathbf{U}) = \prod_i P(\mathbf{U}_i | \mathbf{PA}_i)$ relative au diagramme.

Les propriétés des densités conjointes peuvent être utilisées. Il existe aussi des algorithmes pour déterminer si on a l'interchangeabilité conditionnellement à \mathbf{X} en se basant directement sur le diagramme.

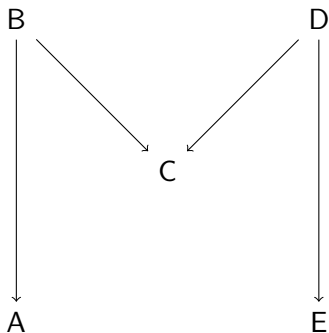
Voir par exemple Shrier et Platt, 2008 qui se basent sur Pearl, 2000.

M-Bias

Un exemple particulier où on combine trajet de porte arrière et collision...



Un exemple particulier où on combine trajet de porte arrière et collision...



Ici, n'ajuster pour aucune variable nous permet d'évaluer l'effet causal de A sur E . Par contre, ajuster pour C seulement en pensant que c'est le seul facteur confondant ($A \leftarrow C \rightarrow E$) serait problématique.

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme
- ▶ Cependant, les données observationnelles contiennent des informations qui nous permettent de mieux comprendre le monde

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme
- ▶ Cependant, les données observationnelles contiennent des informations qui nous permettent de mieux comprendre le monde
- ▶ Hypothèses après hypothèses

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme
- ▶ Cependant, les données observationnelles contiennent des informations qui nous permettent de mieux comprendre le monde
- ▶ Hypothèses après hypothèses
- ▶ Il est important de développer des estimateurs causaux pour les contextes observationnels

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme
- ▶ Cependant, les données observationnelles contiennent des informations qui nous permettent de mieux comprendre le monde
- ▶ Hypothèses après hypothèses
- ▶ Il est important de développer des estimateurs causaux pour les contextes observationnels
- ▶ Je n'ai même pas parlé d'incertitude!

Conclusion

- ▶ On estime souvent des associations en statistique (mais...)
- ▶ L'essai randomisé demeure la norme
- ▶ Cependant, les données observationnelles contiennent des informations qui nous permettent de mieux comprendre le monde
- ▶ Hypothèses après hypothèses
- ▶ Il est important de développer des estimateurs causaux pour les contextes observationnels
- ▶ Je n'ai même pas parlé d'incertitude!
- ▶ Ma recherche?

Références I



Dumas, O., Siroux, V., Le Moual, N., and Varraso, R. (2014).
Approches d'analyse causale en épidémiologie.
Revue d'épidémiologie et de santé publique, 62(1):53–63.



Hill, A. B. (1965).
The environment and disease: association or causation?
Journal of the Royal Society of Medicine, 58(5).



Horvitz, D. G. and Thompson, D. J. (1952).
A generalization of sampling without replacement from a finite universe.
Journal of the American Statistical Association, 47(260):663–685.



Neyman, J. S. (1923).
On the application of probability theory to agricultural experiments. Essay on principles.
Section 9 (translation published in 1990).
Statistical Science, 5(4):472–480.



Pearl, J. (2000).
Models, reasoning and inference.
Cambridge, UK: Cambridge University Press, 19(2).

Références II



Robins, J. M. (1986).

A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect.

[Mathematical Modelling, 7\(9-12\):1393–1512.](#)



Rosenbaum, P. R. (1987).

Model-based direct adjustment.

[Journal of the American Statistical Association, 82\(398\):387–394.](#)



Rosenbaum, P. R. and Rubin, D. B. (1983).

The central role of the propensity score in observational studies for causal effects.

[Biometrika, 70\(1\):41–55.](#)



Rubin, D. B. (1974).

Estimating causal effects of treatments in randomized and nonrandomized studies.

[Journal of Educational Psychology, 66\(5\):688–701.](#)

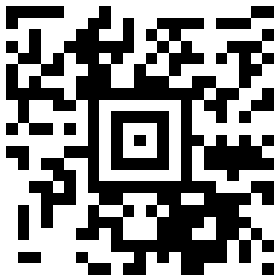


Shrier, I. and Platt, R. W. (2008).

Reducing bias through directed acyclic graphs.

[BMC Medical Research Methodology, 8\(1\):1–15.](#)

Diapositives et références disponibles au



<https://janiecoulombestat.github.io>

Merci!