Causalité : principes et méthodes

Organisation

- Les documents (slides des cours, codes R, etc) sont sur ma pageweb http://www.math-evry.cnrs.fr/members/aguilloux/welcome
- Pour me joindre agathe.guilloux@math.cnrs.fr
- Une référence (gratuire et en ligne) Hernan and J. Robins 2020

Introduction

- Confusion et causilité
- Outcomes potentiels et contrefactualité
- Interventions hypothétiques
- Effets causaux
- Hypothèses de l'inférence causale

Stratification et appariement (matching)

- Stratification
- TP 1 : comparer deux anti-diabétiques
- Etude observationnelle vs étude randomisée
- Appariement/matching
- Score de propension
- TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

- Intuition pour l'inverse de la probabilité de traitement (IPTW)
- Modèles marginaux structuraux
- TP 3 : données (rhc) right heart catheterization

Introduction

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Corrélation parasite (spurious correlation) / causalité I

Des variables non-liées peuvent paraître fortement corrélées

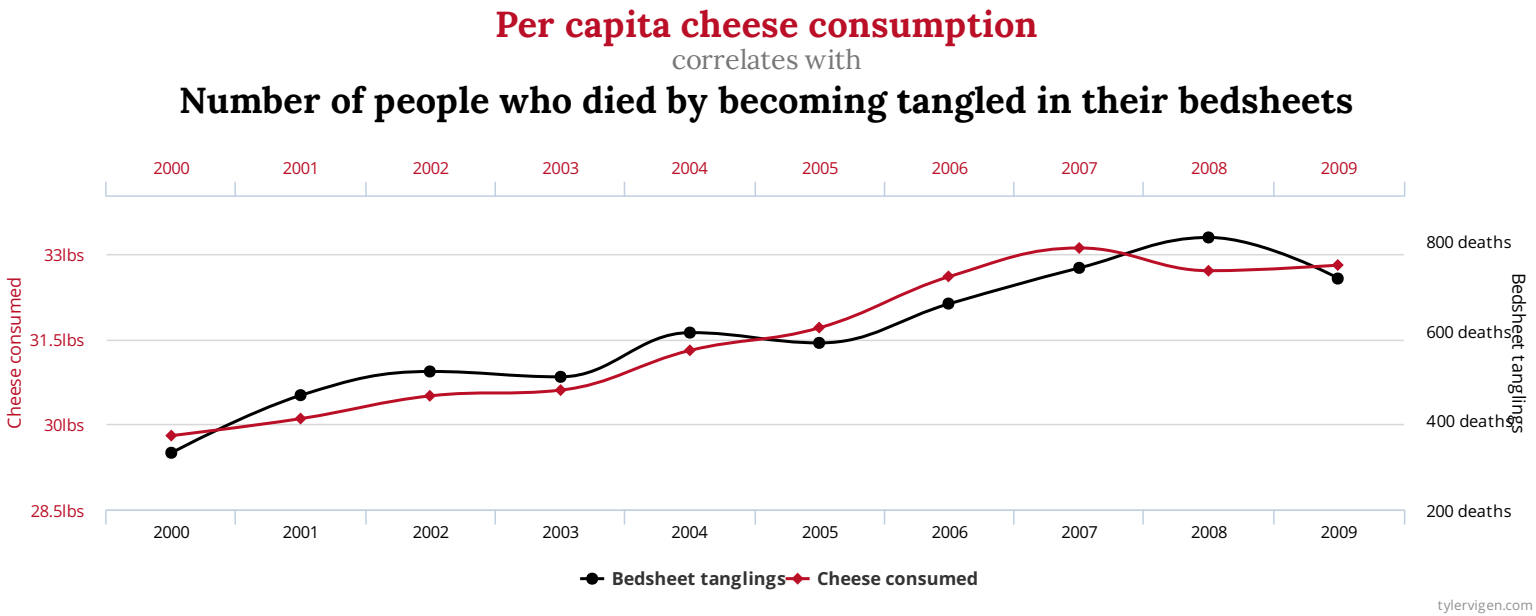


Figure 1: From https://www.tylervigen.com/spurious-correlations

Corrélation parasite (spurious correlation) / causalité II

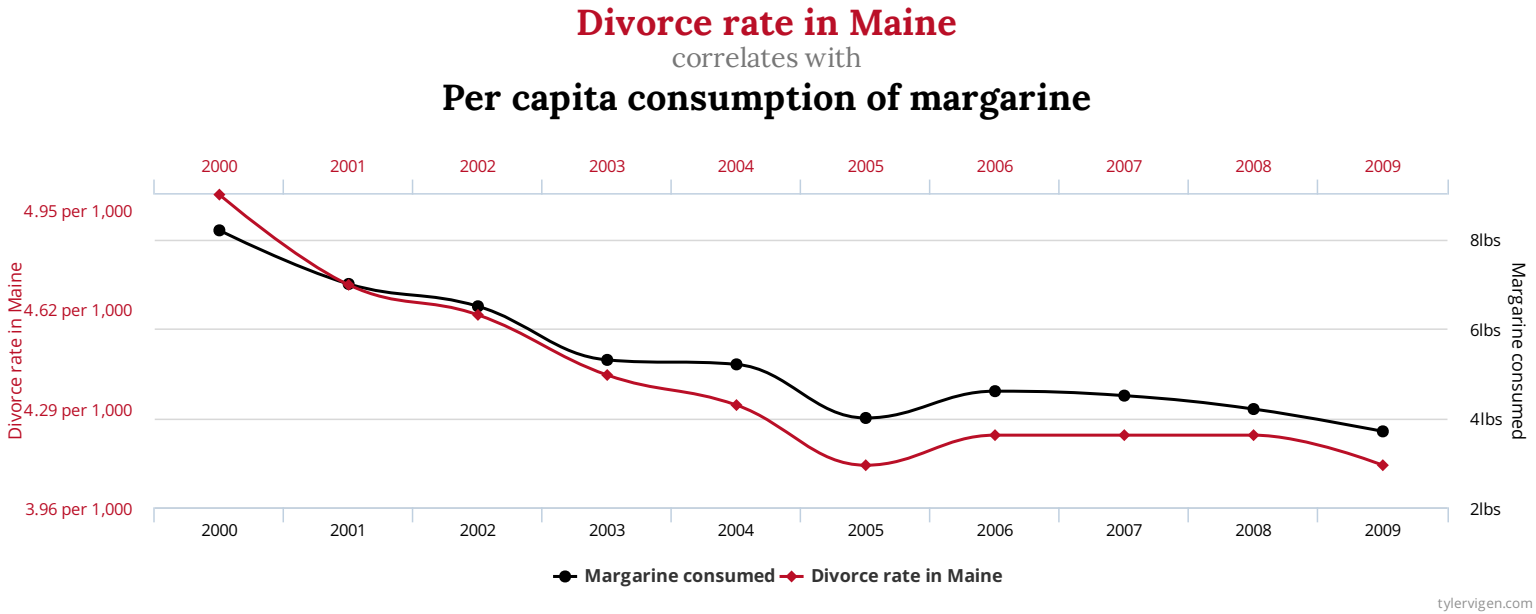


Figure 2: From https://www.tylervigen.com/spurious-correlations

sans qu'il y ait de lien causal...

Causalité et chronologie

Exemple

Plus d'espaces verts en ville $\longleftrightarrow$ plus de courses à pied

- plus d'espaces verts en ville $\rightarrow$ plus de courses à pied
- OU plus d'espaces verts en ville $\leftarrow$ plus de courses à pied ?

Causalité

Il faut

- toujours faire attention à la chronologie !
- introduire des définitions formelles et
- donner les hypothèses nécessaires à l'identification d'un effet causal à partir de données

Quelques mots d'histoire

- Wright 1921
- Neyman 1923
- cela devient un domaine de recherche en statistique dans les années 70
- "outcome" potentiel : Rubin 1974
- diagrammes causaux : Greenland and James M Robins 1986; Pearl 2000
- score de propension : Rosenbaum and Rubin 1983
- ...
- targeted learning : Van der Laan and Rose 2011

A garder en tête : "0bservational studies are an interesting and challenging field which demands a good deal of humility, since we can claim only to be groping toward the truth." (Cochran 1972)

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Traitement et outcomes

- On appelera "traitement" toute exposition à un environnement, une molécule, etc. Il sera noté A
- L'outcome sera noté Y

Exemples

- $A = 1$ si le patient a été vacciné contre la grippe, 0 sinon
- $A = 1$ si le patient a été traité par statines, 0 sinon
- $Y = 1$ si le patient a développé une maladie cardio-vaculaire dans les 2 ans, 0 sinon
- Y temps jusqu'à la rechute

Outcomes potentiels

Définition

Les **outcomes potentiels** sont les outcomes qui sont/seraient observés sous chaque traitement possible.
Notation : Y^a est l'outcome qui a/aurait été observé si le traitement est/avait été $A = a$
Chaque personne a donc 2 outcomes potentiels : Y^0 et Y^1 (dans les cas étudiés dans ce cours)

Exemple

- traitement $A = 1$ si le patient a été vacciné contre la grippe, 0 sinon
- observation : $Y = 1$ si le patient a développé une grippe dans les 6 mois, 0 sinon
- potentiels :
  - $Y^1 = 1$ si le patient a développé une grippe dans les 6 mois après avoir été vacciné/s'il avait été vacciné, 0 sinon
  - $Y^0 = 1$ si le patient a développé une grippe dans les 6 mois sans avoir reçu le vaccin/s'il n'avait pas été vacciné, 0 sinon
  - Si le patient a été vacciné $Y = Y^1$
  - Si le patient n'a pas été vacciné $Y = Y^0$

Contrefactualité

Les outcomes contrefactuels sont ce qu'il aurait été observé si le traitement avait été différent.

- Si mon traitement a été $A = 1$, mon outcome contrefactuel est Y^0
- Si mon traitement a été $A = 0$, mon outcome contrefactuel est Y^1

Exemple

Si j'ai reçu le vaccin anti-grippe et je ne suis pas tombé malade

- ma vraie exposition est $A = 1$
- mon outcome observé est $Y = Y^1 = 0$
- mon outcome contrefactuel est Y^0 qui est ce qu'il se serait passé si je n'avais pas reçu le vaccin (contrairement à la réalité).
- Avant la décision sur le traitement chaque outcome Y^0 et Y^1 est potentiel
- Après l'étude, il y a un outcome observé $Y = Y^A$
- et un outcome contrefactuel Y^{1-A}
- les outcomes contrefactuels sont supposés être égaux aux outcomes potentiels

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Intervention

- L'intervention est certainement la manière la plus propre de penser les effets causaux.
- On pourra donc étudier les effets causaux de variables qui peuvent être "manipulées"

Exemple

On veut étudier l'effet de l'IMC sur un outcome. On ne peut pas directement "manipuler" l'IMC, on peut intervenir sur d'autres variables (arrêt du tabac, reprise du sport) qui auront une incidence sur l'IMC puis sur la santé, donc l'outcome d'intéret.
On ne pourra donc pas directement étudier l'effet causal de la baisse d'IMC sur la santé mais plutôt celui des interventions qui peuvent avoir une influence sur l'IMC.

Variables immuables

Il est difficile de définir l'effet causal de variables sur lesquelles on ne peut pas agir : les variables immuables.

Non manipulable	Manipulable
race	nom sur un cv
obésité	bypass gastrique
CSP	don d'argent

Nous allons nous concentrer sur les effets causaux d'un traitement ou d'une exposition sur lesquels on peut imaginer agir (même si cela ne serait pas éthique).

Effets causaux

En général, on dira qu'il y a un effet causal si Y^1 est différent de Y^0.

Exemple

- $A = 1$ si j'ai pris de l'ibuprofène, 0 sinon
- $Y = 1$ si je n'ai plus mal à la tête au bout d'une heure

"J'ai pris de l'ibuprofène et je n'ai plus mal à la tête, donc le traitement est efficace" n'est pas un raisonnement causal correct.
Il faut raisonner comme suit

- mon outcome $Y = 1$
- que se serait-il passé si je n'avais pas pris d'ibuprofène $Y^0 =$?
- il y a un effet causal si $Y^1 \neq Y^0$ (en distribution)

Problème fondamental de l'inférence causale

On ne peut pas observer deux outcomes par individu !

Cependant, avec certaines hypothèses, on pourra estimer un effet causal moyen sur une population (mais pas au niveau individuel donc).

Exemple

Dans l'exemple sur l'ibuprofène, on définira donc l'effet causal comme la différence entre le taux moyen d'individus sans mal de tête si tout le monde avait reçu de l'ibuprofène et celui dans le cas où personne n'aurait pris d'ibuprofène.

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Causalité et association

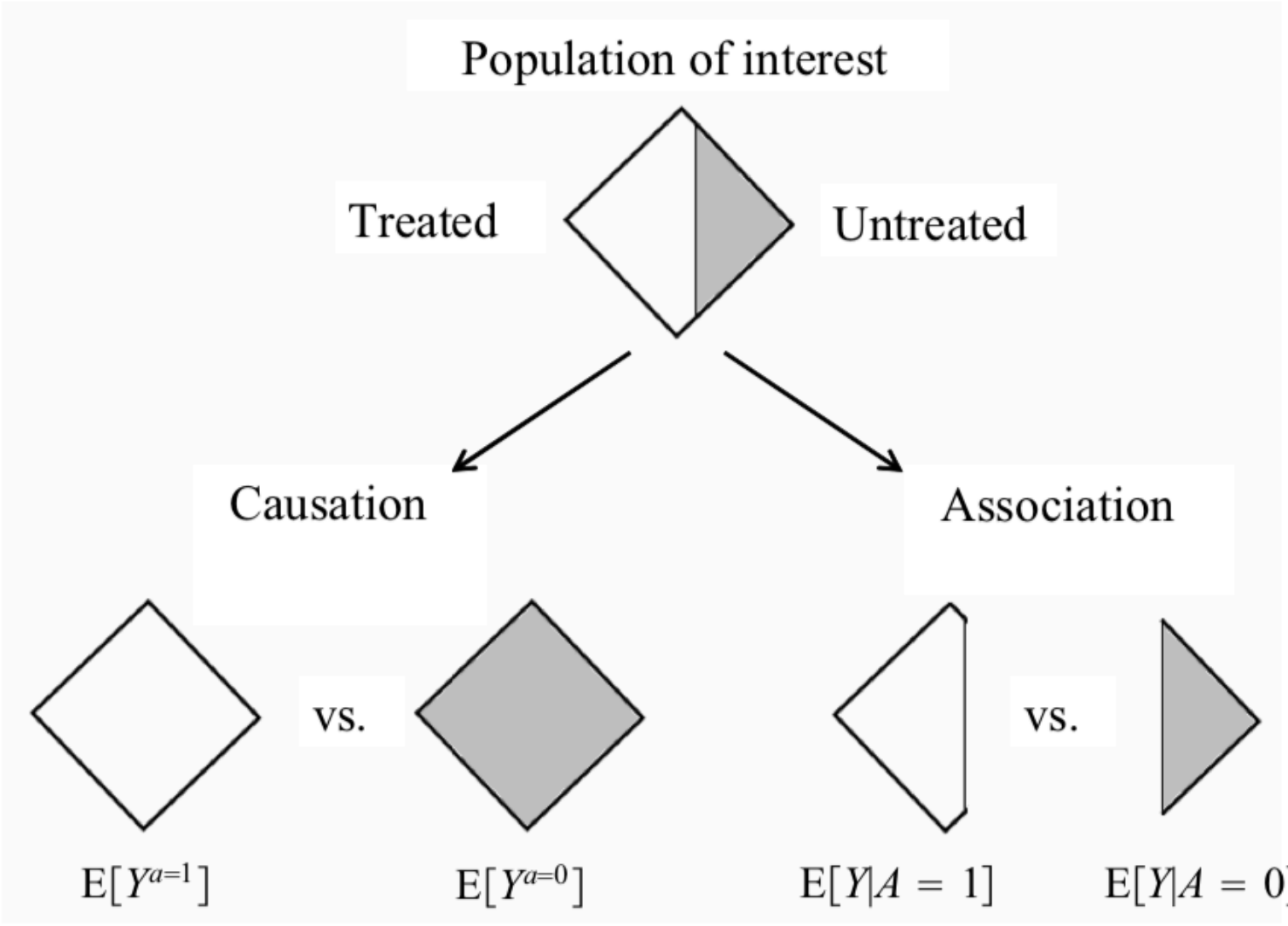


Figure 3: Hernan and J. Robins 2020

Effet causal moyen

Effet causal moyen - Average causal effect (ACE)

$$\text{ACE} = \mathbb{E}(Y^1) - \mathbb{E}(Y^0)$$

avec

- $\mathbb{E}(Y^1)$ moyenne (espérance) de Y si tout le monde avait été traité
- $\mathbb{E}(Y^0)$ moyenne (espérance) de Y si personne n'avait été traité.

Remarque fondamentale

En général

$$\text{ACE} = \mathbb{E}(Y^1) - \mathbb{E}(Y^0) \neq \mathbb{E}(Y|A=1) - \mathbb{E}(Y|A=0)$$

Pourquoi ? Parce que $\mathbb{E}(Y|A=1)$ est la moyenne de Y parmi les gens qui ont reçu le traitement, c'est une moyenne restreinte à la sous-population des traités, qui peut différer de la population générale

Exemple

Dans l'exemple du vaccin contre la grippe, une personne à risque pour la grippe (ex. personnel soignant) peut d'avantage vouloir être vaccinée.

Autres mesures d'effets causaux

Suivant les cas, on peut vouloir également calculer

- le risque relatif causal $\mathbb{E}(Y^1)/\mathbb{E}(Y^0)$
- l'odds ratio causal

$$\frac{\frac{\mathbb{E}(Y^1)}{1-\mathbb{E}(Y^1)}}{\frac{\mathbb{E}(Y^0)}{1-\mathbb{E}(Y^0)}}$$

- $\mathbb{E}(Y^1|A=1)/\mathbb{E}(Y^0|A=1)$ l'effet causal du traitement chez les traités
- $\mathbb{E}(Y^1|V=v)/\mathbb{E}(Y^0|V=v)$ l'effet causal du traitement dans la sous-population avec $V=v$

Pourquoi c'est compliqué ?

Problème fondamental de l'inférence causale

On n'observe qu'un traitement et un outcome pour chaque individu !

Dans la suite, on verra

- comment utiliser les données disponibles pour lier les outcomes observés aux outcomes potentiels (contrefactuels)
- quelles hypothèses sont nécessaires pour estimer un effet causal à partir des données.

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Conditions d'identifiabilité

Conditions nécessaires pour identifier un effet causal

- **consistance** : $Y^a = Y$ si $A = a$ pour tout a, l'outcome potentiel Y^a vaut Y si le traitement est $A = a$ (voir p31 et suivante de Hernan and J. Robins 2020)
- **échangeabilité** : $Y^a \perp\!\!\!\perp A|X = x$ pour tout a, l'outcome contrefactuel est indépendant du traitement. Ou bien si deux personnes parfaitement semblables ont reçu le traitement pour l'une et rien pour l'autre, l'outcome observé de l'une est le contrefactuel de l'autre (en distribution)
- **positivité** : $\mathbb{P}(A = a|X = x) > 0$ pour tout a et tout x, le traitement peut prendre les valeurs 0 et 1 avec probabilités > 0 pour tout x.

Pourquoi cela va marcher ?

Si on veut estimer

$$\mathbb{E}(Y^1|X=x) - \mathbb{E}(Y^0|X=x).$$

Par positivité, on peut définir le conditionnement par rapport à A, puis on pourra toujours estimer

$$\mathbb{E}(Y|A=a, X=x)$$

à partir des données. Et on remarque, qu'avec les hypothèses

$$\mathbb{E}(Y|A=a, X=x) \underset{consistance}{=} \mathbb{E}(Y^a|A=a, X=x) \underset{ignorabilite}{=} \mathbb{E}(Y^a|X=x).$$

Stratification et appariement (matching)

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Stratification

On vient de montrer que, sous les bonnes hypothèses :

$$\mathbb{E}(Y|A=a, X=x) = \mathbb{E}(Y^a|X=x)$$

si on veut estimer l'effet causal marginal $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$ on peut moyenner sur la loi de X.

Si X est une covariable discrète

$$\mathbb{E}(Y^a) = \sum_x \mathbb{E}(Y^a|X=x)\mathbb{P}(X=x) = \sum_x \mathbb{E}(Y|A=a, X=x)\mathbb{P}(X=x)$$

ça marche car on peut estimer $\mathbb{E}(Y|A=a, X=x)$ à partir des données !

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

TP 1 : comparer deux anti-diabétiques

On veut comparer deux anti-diabétiques (metformine et sitagliptine) et leurs effets sur l'outcome MACE (major adverse cardiac event).

Le problème est que la metformine est plus récente que la sitagliptine. Les individus sous metformine ont donc plus de chances d'avoir déjà eu un traitement par OAD. Les patients qui ont déjà suivi un traitement par OAD sont sans doute plus à risque pour le MACE.

Les données du TP ont été simulées.

Cela ne marchera pas en général s'il y a plusieurs covariables mesurées et/ou si elles sont continues, car les effectifs dans chaque case sera petit (voir nul)...

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Etude observationnelle vs étude randomisée I

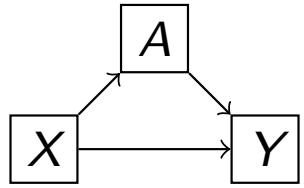


Figure 4: Etude observationnelle

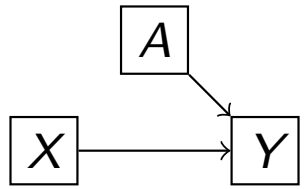


Figure 5: Étude randomisée

La différence est

- dans une étude randomisée, comme le traitement est randomisé, les distributions des facteurs de confusion (autres covariables) sont donc les mêmes dans les deux bras de traitement
- dans une étude observationnelle, le traitement dépend des facteurs de confusion, leur distribution peuvent différer dans les deux bras de traitement.

Etude observationnelle vs étude randomisée II

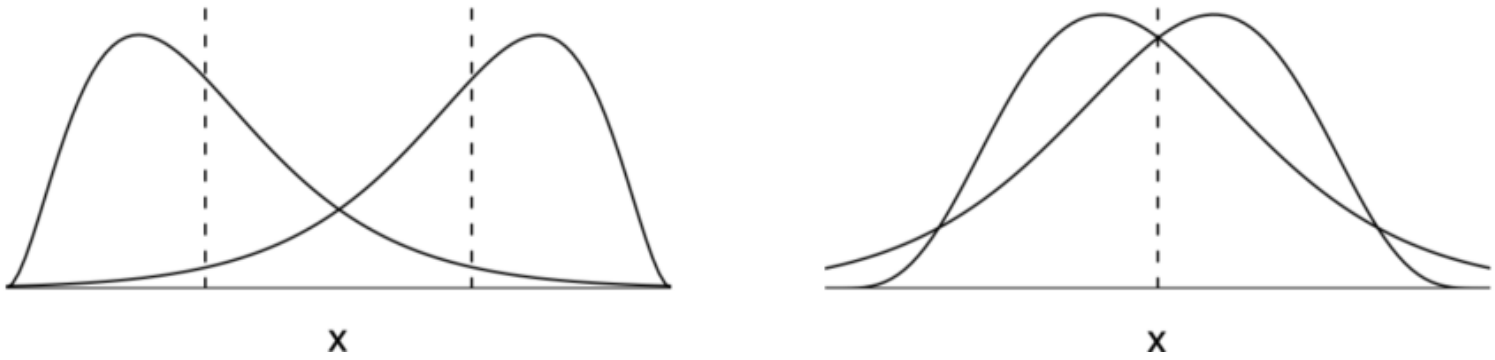


Figure 6: Dans *Matching (and discarding non-matches)* n.d.

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Appariement/matching I

C'est une méthode qui cherche à rendre une étude observationnelle "plus comme" une étude randomisée.

Idée générale

On va apparier/matcher chaque individu du groupe traité à un individu non-traité à partir de leurs covariables.

Figure 7: D'après Rubin 2006

Appariement/matching II

Cela va donc permettre

- un contrôle des facteurs de confusion sans regarder l'outcome
- une fois les données appariées d'étudier l'outcome comme dans un essai randomisé
- de vérifier l'hypothèse de positivité

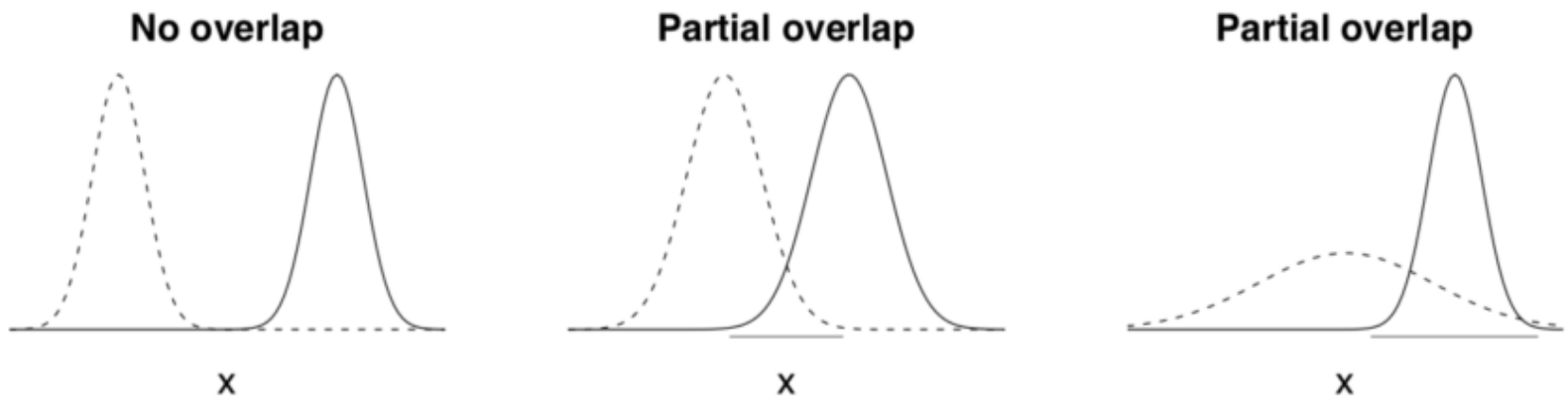


Figure 8: Dans *Matching (and discarding non-matches)* n.d.

Remarques

- Comme on apparie chaque individu du groupe traité avec un individu du groupe contrôle, la distribution des covariables/facteurs de confusion dans la population contrôle sera la même que dans la population traitéE, on ne pourra donc calculer qu'un **effet moyen du traitement chez les traités**.
- On peut apparier un contrôle à tout traité, ou plus $K \geq 1$.
- S'il y a beaucoup de covariables, on ne pourra pas faire d'appariement parfait, on cherchera alors un **équilibre stochastique** : c'est-à-dire une même distribution des covariables dans les deux groupes.

Exemple

Dans le cas suivant, on a l'équilibre stochastique

- Traité M 40 ans $\longleftrightarrow$ Contrôle F 45 ans
- Traité F 45 ans $\longleftrightarrow$ Contrôle M 40 ans

même si aucun appariement n'est parfait.

Appariement glouton / greedy matching (1)

Pour chaque individu, on a un vecteur de covariables/facteurs de confusion

- $X_{T,k}$ pour k traité
- $X_{C,l}$ pour l contrôle.

Pour chaque individu traité k, on calcule les distances

$$d(X_{T,k}, X_{C,l}) \text{ pour tout } l \text{ contrôle}$$

entre $X_{T,k}$ et $X_{C,l}$ pour tous les individus contrôle l.

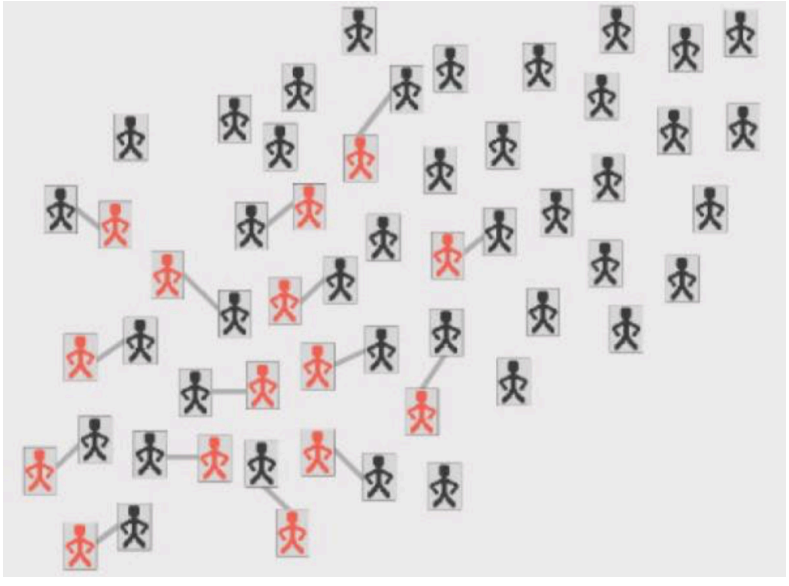


Figure 9: Dans Rubin 2006

Appariement glouton / greedy matching (2)

Algorithme d'appariement

- Step 1 : on permute aléatoirement la liste des traités
- Step 2 : pour le (nouveau) 1er traité, on cherche le contrôle l qui est le plus proche ($\min_l d(X_{T,k}, X_{C,l})$)
- Step 3 : on enlève le contrôle choisi à l'étape précédent de la liste des contrôles disponibles
- Step 4 : on prend le traité suivant, on l'apparie avec le contrôle le plus proche
- Step 5 : on répète les étapes 3 et 4 jusqu'à ce que tous les traités aient été appariés.

Remarques :

- Tout dépend de l'ordre aléatoire de l'étape 1 (il est conseillé de recommencer plusieurs fois)
- On peut choisir d'apparier plusieurs contrôles à chaque fois.

En pratique : la distance

Il faut choisir la distance, la plus courante est la distance de Mahalanobis

Distance de Mahalanobis

Pour $X_{T,k}$ covariables k traité et $X_{C,l}$ covariables de l contrôle,

$$d(X_{T,k}, X_{C,l}) = \sqrt{(X_{T,k} - X_{C,l})^\top \Sigma^{-1} (X_{T,k} - X_{C,l})}$$

avec Σ matrice de variance-covariance des covariables.

En pratique : le caliper

- On peut choisir de ne pas apparier s'il n'existe pas de contrôle dont les covariables sont à distance raisonnable de celles du traité.
- On se donne une valeur de caliper ϵ, si on ne trouve pas pour un traité k de contrôle l tel que $d(X_{T,k}, X_{C,l}) < \epsilon$, on enlève ce traité de l'étude.
- Cela a un rapport avec l'hypothèse de positivité : s'il y a aucun appariement possible, c'est ce que l'hypothèse de positivité est fausse (et qu'on ne pourra rien faire...)

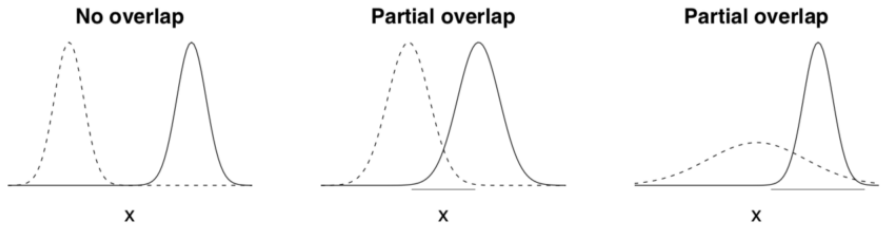


Figure 10: Dans *Matching (and discarding non-matches)* n.d.

En pratique : la "Table 1"

Une fois l'appariement fait, il faut vérifier que les distributions des covariables sont les mêmes dans les deux populations traités/contrôles.

Différence des moyennes standardisée

On calcule alors pour chaque covariable j

$$\text{smd}_j = \frac{\text{moy}(X_T^j) - \text{moy}(X_C^j)}{\sqrt{0.5\text{var}(X_T^j) + 0.5\text{var}(X_C^j)}}$$

Avec la règle empirique

- $\text{smd}_j < 0.1$ la balance est atteinte
- $0.1 \leq \text{smd}_j < 0.2$ pas trop alarmant
- $\text{smd}_j > 0.2$ la balance n'a pas été atteinte.

voir Austin 2011; Normand et al. 2001.

	0	1	SMD
n	3551	2184	
CHF (mean (SD))	0.07 (0.25)	0.10 (0.29)	0.095
Cirr (mean (SD))	0.11 (0.32)	0.03 (0.16)	0.342
colcan (mean (SD))	0.05 (0.22)	0.02 (0.15)	0.145

En pratique : l'analyse après appariement

Une fois l'appariement fait, on peut alors estimer (avec intervalle de confiance) ou tester pour un effet causal du traitement

- avec un t-test apparié (`t.test(x, y, paired=TRUE)`) si l'outcome est numérique (continu)
- avec un test de McNeymar (`mcnemar.test`) si l'outcome est binaire

Exemple de test de McNemar

	0	1
0	303	395
1	509	977

$$K = \frac{(509 - 395)^2}{509 + 395}$$

comparer avec un quantile de la loi du χ^2.

Appariement optimal

- L'algorithme d'appariement glouton n'assure pas que la somme des distances dans chaque paire appariée soit minimale.
- Il existe des algorithmes qui cherchent à minimiser cette distance totale, attention c'est un problème très lourd quand il y a beaucoup d'individus
  - 2T/3C : 6 appariements possibles
  - 10T/40C : 3×10^15 appariements possibles

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Score de propension

C'est, pour un individu i avec des covariables X_i, la probabilité de recevoir le traitement

$$\pi_i = \mathbb{P}(A = 1|X_i)$$

Exemple

Supposons que les personnes âgées sont plus susceptibles de recevoir le traitement alors

$$\mathbb{P}(A = 1|\text{âge}_k = 60) > \mathbb{P}(A = 1|\text{âge}_l = 30) \Longleftrightarrow \pi_k > \pi_l.$$

Un score d'équilibre

- Même si deux individus k et l ont des covariables différentes $X_k \neq X_l$, ils peuvent avoir la même probabilité d'être traité $\pi_k = \pi_l$
- Cela signifie que les deux individus ont la même probabilité d'être dans le groupe de traitement et donc dans le groupe contrôle
- Donc si on fait l'appariement sur le score de propension, on va réussir à obtenir deux groupes traités/contrôle bien équilibrés
- On va donc apparier à partir du score de propension et, pour cela, il faut l'estimer !

Estimation du score de propension

- Il nous faut donc estimer $\pi_i = \mathbb{P}(A = 1|X_i)$ pour chaque individu i
- Comme, dans nos exemples, A est binaire, on peut utiliser n'importe quelle méthode de classification
- Le plus courant est de faire une régression logistique
- On obtient alors pour chaque individu i

$$\hat{\pi}_i = \hat{\mathbb{P}}(A = 1|X_i) \in [0, 1]$$

et le matching est bien plus simple !

Appariement à partir du score de propension

- En pratique, on utilise le log-odds du score de propension

$$\frac{\pi_i}{1-\pi_i} = X_i\hat{\beta} \text{ (dans le cas de la régression logistique)}$$

- Attention à bien vérifier qu'il y a recouvrement entre les scores entre les groupes traités et contrôles

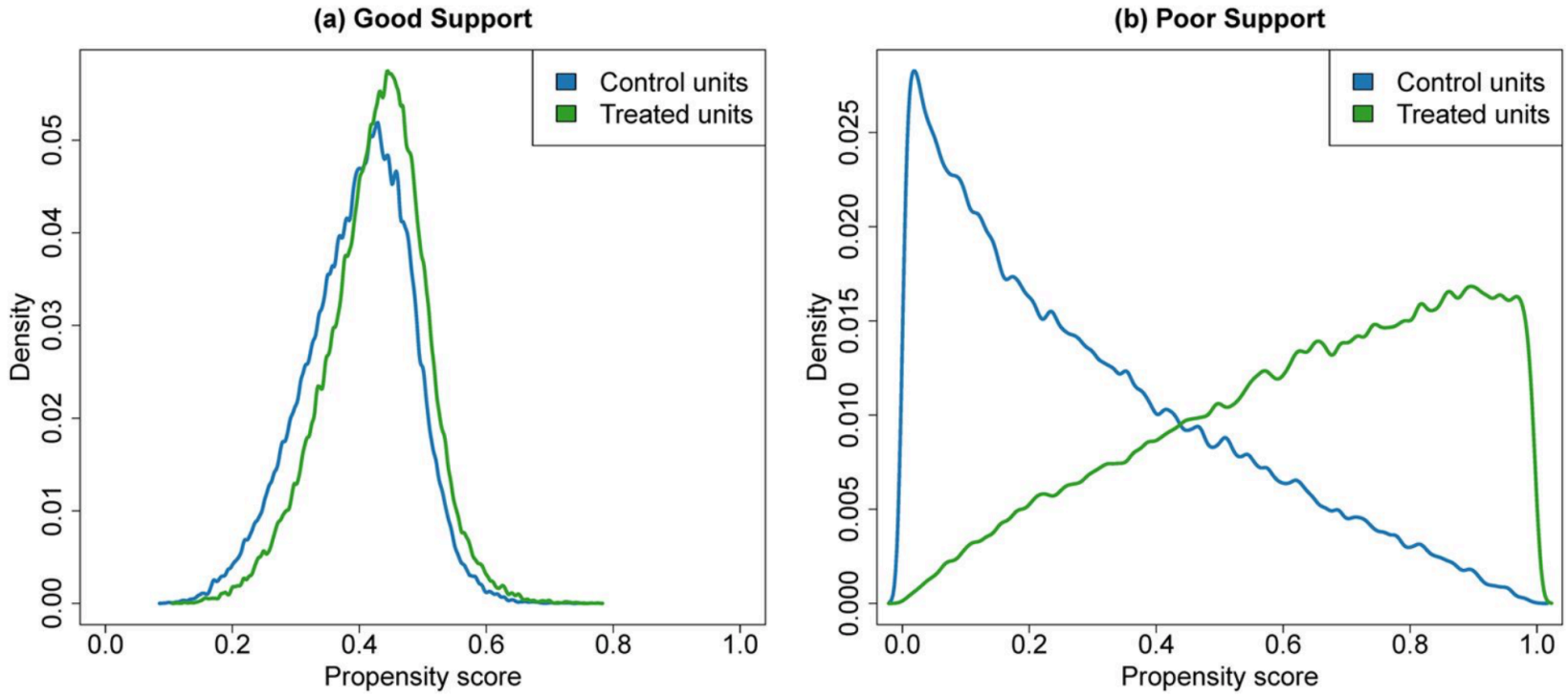


Figure 11: Dans Colson et al. 2016

Algorithme et définition du caliper

- Step 1 : estimer les scores de propension
- Step 2 : calculer les log-odds
- Step 3 : calculer l'écart-type $\hat{\sigma}$ des log-odds
- Step 4 : définir le caliper comme $\epsilon = 0.2 \times \hat{\sigma}$
- Step 5 : apparier

Introduction

- Confusion et causilité
- Outcomes potentiels et contrefactualité
- Interventions hypothétiques
- Effets causaux
- Hypothèses de l'inférence causale

Stratification et appariement (matching)

- Stratification
- TP 1 : comparer deux anti-diabétiques
- Etude observationnelle vs étude randomisée
- Appariement/matching
- Score de propension
- TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

- Intuition pour l'inverse de la probabilité de traitement (IPTW)
- Modèles marginaux structuraux
- TP 3 : données (rhc) right heart catheterization

TP 2 : données (rhc) right heart catheterization

Une description complète est disponible à cette adresse http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html.

Il s'agit de données sur des patients (2184 traités et 3551 contrôles) admis aux urgences dans 5 hôpitaux, la variable de traitement est `swang1` (right heat catheterization vs. non) et l'outcome est `death` (yes or no). Nous allons considérer les variables de confusion suivantes :

- `cat1` : Primary disease category
- `age`
- `sex`
- `meanbp1` : mean blood pressure

Estimer l'effet causal du traitement via l'appariement glouton et l'appariement par score de propension

Pondération par IPTW et MSM

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Intuition pour l'IPTW

Exemple d'après Schulte and Mascha 2018

Dans Komatsu et al. 2015, est étudié le lien entre

- l'outcome Y arythmie auriculaire et
- le traitement A par analgésie épidurale thoracique (thoracic epidural analgesia : TEA) vs versus anesthésie générale

le facteur de confusion considéré X est le type de chirurgie (thoracotomie ou autre).

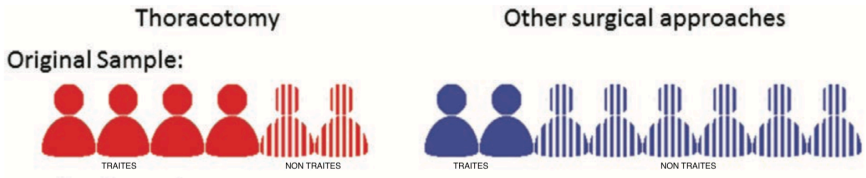


Figure 12: Dans Schulte and Mascha 2018

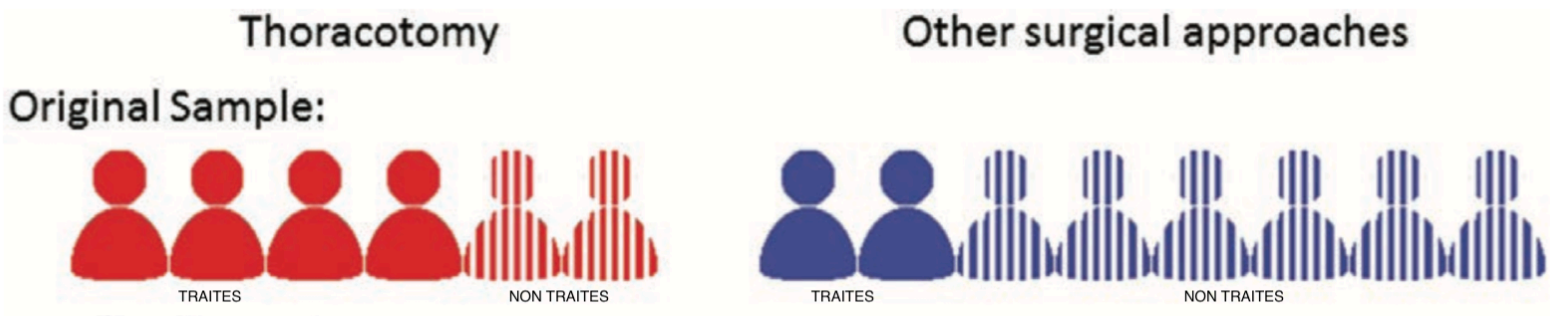
Si on fait un appariement, dans le groupe thoracotomie$=0$, un traité par TEA compte pour 3 contrôles.

Pondération (1)

On peut donc imaginer pondérer les individus de chacun des 4 sous-groupes (TEA ou non / thoracotomie ou autre).

On définit les pondérations comme suit

- chez les traités, on pondèrera par l'inverse de la probabilité d'être traité (score de propension) $\pi_i = \mathbb{P}(A = 1|X_i)$
- chez les contrôles, on pondèrera par l'inverse de la probabilité d'être non-traité $1 - \pi_i = \mathbb{P}(A = 0|X_i)$



Pondération (2)

On obtient alors une **pseudo-population**

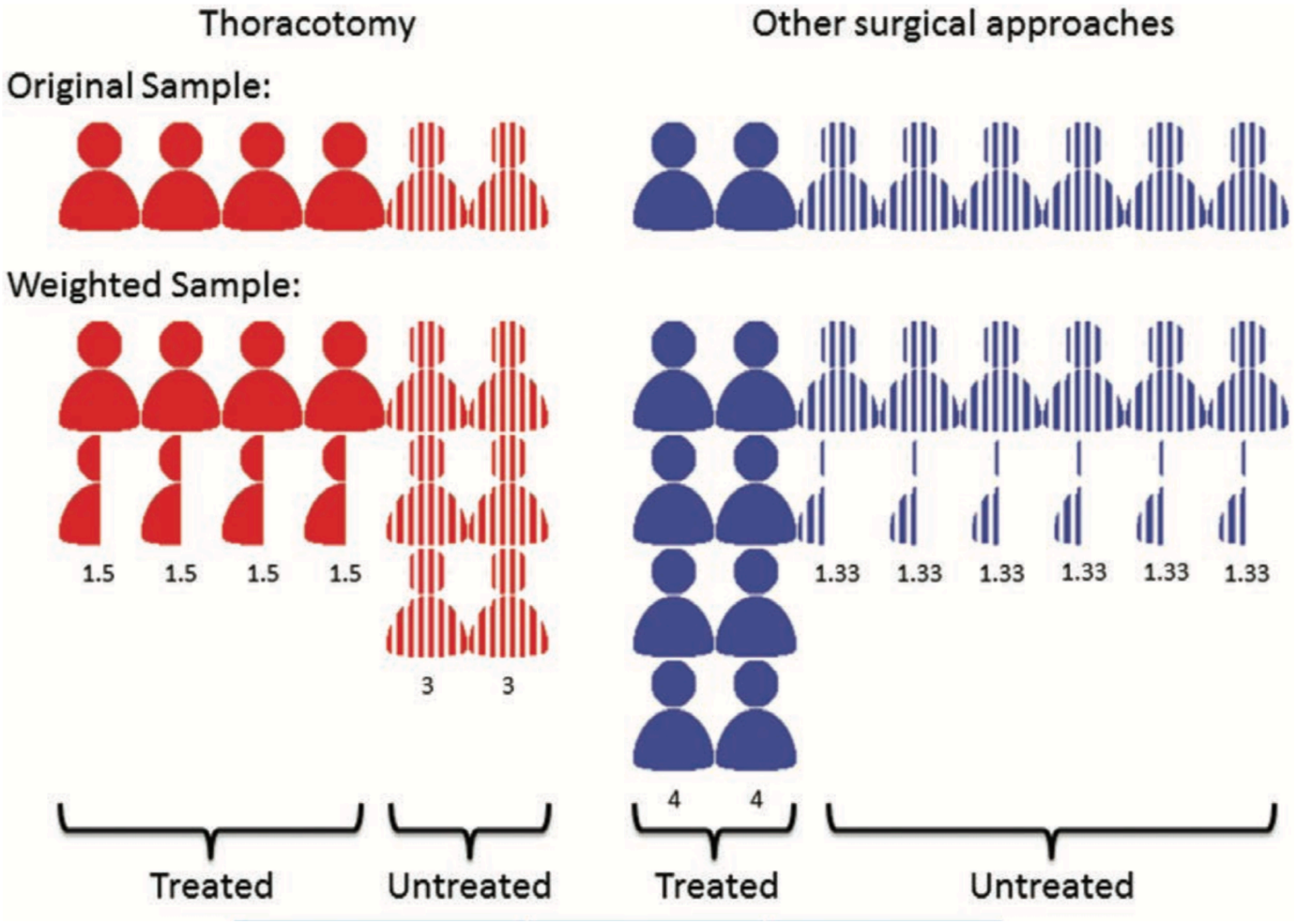


Figure 13: Dans Schulte and Mascha 2018

Estimation

Pour estimer $\mathbb{E}(Y^1)$, par exemple, on fait alors une moyenne pondérée

$$\widehat{\mathbb{E}(Y^1)} = \frac{\sum_{i=1,\ldots,n:A_i=1} \frac{Y_i}{\pi_i}}{\sum_{i=1,\ldots,n:A_i=1} \frac{1}{\pi_i}}$$

Remarques :

- c'est exactement la moyenne dans la pseudo-population
- cela n'est défini que sous l'hypothèse de positivité (il faut que chaque $\pi_i > 0$)
- c'est une version de l'estimateur d'Horvitz–Thompson
- c'est différent de la moyenne chez les traités

$$\widehat{\mathbb{E}(Y|A=1)} = \frac{\sum_{i=1,\ldots,n:A_i=1} Y_i}{\sum_{i=1,\ldots,n:A_i=1} 1}.$$

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

Modèles marginaux structuraux - marginal structural models (MSM) (1)

Modèle marginal structural

C'est un modèle liant Y^a à a. On utilise la plupart du temps un GLM dans lesquels on écrira

$$g(\mathbb{E}(Y^a)) = \psi_0 + \psi_1 a.$$

La valeur estimée de ψ_1 sera alors l'effet causal estimé.

Ils sont dits :

- **marginaux** car on ne conditionne pas par rapport aux covariables (facteurs de confusion) X (ils n'apparaisent pas dans le modèle)
- **structuraux** car ce sont des modèles pour les outcomes potentiels (pas les observés)

Modèles marginaux structuraux (2)

MSM linéaire et logistique

Quand l'outcome Y est continu, il est courant de considérer un MSM linéaire

$$\mathbb{E}(Y^a) = \psi_0 + \psi_1 a$$

dans ce cas $\psi_1 = \mathbb{E}(Y^1) - \mathbb{E}(Y^0)$ est l'effet causal moyen. Quand l'outcome Y est binaire, il est courant de considérer un MSM logistique

$$\text{logit}(\mathbb{E}(Y^a)) = \frac{\mathbb{E}(Y^a)}{1 - \mathbb{E}(Y^a)} = \psi_0 + \psi_1 a$$

et $\exp(\psi_1)$ est l'odds ratio causal.

Rappels sur l'estimation dans un GLM linéaire

Notations : pour $i = 1, \ldots, n$

- Z_i est la réponse (outcome)
- U_i sont les covariables.

On suppose que $Z_i \sim \mathcal{N}(U_i\beta, \sigma^2)$, qui a alors la densité

$$f(z_i) = exp\Big(\frac{z_i U_i\beta - (U_i\beta)^2}{\sigma^2} + c(y_i, \sigma^2)\Big).$$

A partir de l'échantillon (Z_i, U_i), on forme alors la log-vraisemblance

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \log(f(Z_i)) = \sum_{i=1}^{n} \Big\{ \frac{Z_i U_i\beta - (U_i\beta)^2}{\sigma^2} + c(Z_i, \sigma^2) \Big\},$$

et on définit l'estimateur de β au maximum de vraisemblance

$$\nabla \log \mathcal{L}(\hat{\beta}) = 0 \ \text{(équation d'estimation)}.$$

MSM linéaire

On revient maintenant à notre MSM linéaire

$$\mathbb{E}(Y^a) = \psi_0 + \psi_1 a \quad \text{ou} \quad Y^a \sim \mathcal{N}(\psi_0 + \psi_1 a, \sigma^2)$$

on se place dans la pseudo-population est donc défini donc la log-vraisemblance comme

$$\log \mathcal{L}_{\text{weighted}}(\psi_0, \psi_1) = \sum_{i=1}^{n} W_i \Big\{ \frac{Y_i(\psi_0 + \psi_1 A_i) - (\psi_0 + \psi_1 A_i)^2}{\sigma^2} + c(y_i, \sigma^2) \Big\}$$

avec les poids

$$W_i = \frac{1}{\hat{\pi}_i} \quad \text{pour les traités et} \quad W_i = \frac{1}{1 - \hat{\pi}_i} \quad \text{pour les contrôles.}$$

On définit l'estimateur de ψ_1 au maximum de vraisemblance pondéré par l'équation

$$\nabla \log \mathcal{L}_{\text{weighted}}(\hat{\psi}_0, \hat{\psi}_1) = 0 \ \text{(équation structurelle)}.$$

Estimateur “sandwich” de la variance

On peut relire la log-vraisemblance pondérée

$$\log \mathcal{L}_{\text{weighted}}(\psi_0, \psi_1) = \sum_{i=1}^{n} W_i \Big\{ \frac{Y_i(\psi_0 + \psi_1 A_i) - (\psi_0 + \psi_1 A_i)^2}{\sigma^2} + c(y_i, \sigma^2) \Big\}$$

comme la log-vraisemblance dans un cas hétéro-scédastique

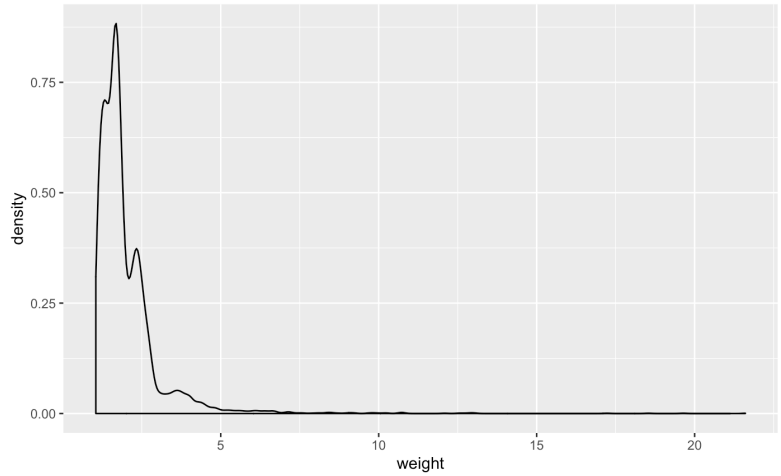
$$\widetilde{Y}_i \sim \mathcal{N}\Big(\psi_0 + \psi_1 A_i, \frac{\sigma^2}{W_i}\Big).$$

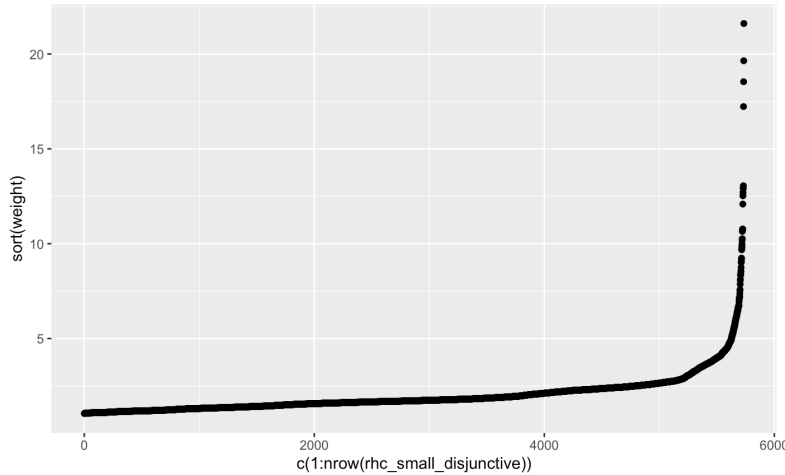
Dans un tel cas, pour estimer les variances, on utilise un estimateur sandwich.

Dans le cas d’une régression logistique, on utilisera la famille quasi-binomiale et l’estimateur sandwich.

En pratique

- la "Table 1" : une fois l'appariement fait, il faut vérifier que les distributions des covariables sont les mêmes dans les deux populations traités/contrôles, cela est fait sur la pseudo-population.
- Il faut également vérifier la distribution des poids





Si certains sont trop grands, on les tronque à une valeur fixe ou à un percentile (e.g. 99%).

Introduction

Confusion et causilité

Outcomes potentiels et contrefactualité

Interventions hypothétiques

Effets causaux

Hypothèses de l'inférence causale

Stratification et appariement (matching)

Stratification

TP 1 : comparer deux anti-diabétiques

Etude observationnelle vs étude randomisée

Appariement/matching

Score de propension

TP 2 : données (rhc) right heart catheterization

Pondération par IPTW et MSM

Intuition pour l'inverse de la probabilité de traitement (IPTW)

Modèles marginaux structuraux

TP 3 : données (rhc) right heart catheterization

TP 3 : données (rhc) right heart catheterization

Une description complète est disponible à cette adresse http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html.

Il s'agit de données sur des patients (2184 traités et 3551 contrôles) admis aux urgences dans 5 hôpitaux, la variable de traitement est `swang1` (right heat catheterization vs. non) et l'outcome est `death` (yes or no). Nous allons considérer les variables de confusion suivantes :

- `cat1` : Primary disease category
- `age`
- `sex`
- `meanbp1` : mean blood pressure

Estimer l'effet causal du traitement via l'IPTW et un MSM.

Références I

Matching (and discarding non-matches). https://statmodeling.stat.columbia.edu/. Accessed: 2019-11-08.

Peter C Austin. "An introduction to propensity score methods for reducing the effects of confounding in observational studies". In: *Multivariate behavioral research* 46.3 (2011), pp. 399–424.

William G Cochran. "Observational studies". In: *Introduction to Observational Studies and the Reprint of Cochran's paper and Comments* (1972), p. 126.

K Ellicott Colson et al. "Optimizing matching and analysis combinations for estimating causal effects". In: *Scientific reports* 6 (2016), p. 23222.

Sander Greenland and James M Robins. "Identifiability, exchangeability, and epidemiological confounding". In: *International journal of epidemiology* 15.3 (1986), pp. 413–419.

Références II

MA Hernan and JM Robins. *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC, 2020. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2019/11/ci_hernanrobins_5nov19.pdf.

Ryu Komatsu et al. "Association of thoracic epidural analgesia with risk of atrial arrhythmias after pulmonary resection: a retrospective cohort study". In: *Journal of anesthesia* 29.1 (2015), pp. 47–55.

Jersey Neyman. "Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes". In: *Roczniki Nauk Rolniczych* 10 (1923), pp. 1–51.

Sharon-Lise T Normand et al. "Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores". In: *Journal of clinical epidemiology* 54.4 (2001), pp. 387–398.

Judea Pearl. *Causality: models, reasoning and inference*. Vol. 29. Springer, 2000.

Références III

Paul R Rosenbaum and Donald B Rubin. "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1 (1983), pp. 41–55.

Donald B Rubin. *Matched sampling for causal effects*. Cambridge University Press, 2006.

Donald B Rubin. "Estimating causal effects of treatments in randomized and nonrandomized studies.". In: *Journal of educational Psychology* 66.5 (1974), p. 688.

Phillip J Schulte and Edward J Mascha. "Propensity score methods: theory and practice for anesthesia research". In: *Anesthesia & Analgesia* 127.4 (2018), pp. 1074–1084.

Mark J Van der Laan and Sherri Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

Sewall Wright. "Correlation and causation". In: *J. agric. Res.* 20 (1921), pp. 557–580.

Estimateur sandwich

En régression linéaire $Y_i = X_i\beta + \epsilon_i$,

- dans le cas homo-scédastique $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, l'estimateur de la variance de $\hat{\beta}$ est donné par
$$\hat{\sigma}^2 (X^\top X)^{-1}$$
- dans le cas hétéro-scédastique $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_i^2)$, l'estimateur sandwich de la variance de $\hat{\beta}$ est donné par
$$(X^\top X)^{-1}(X^\top \Sigma X)(X^\top X)^{-1}$$
avec $\Sigma = \text{diag}(e_1^2, \ldots, e_n^2)$.

Rappels sur l'estimation dans un GLM logistique

Notations : pour $i = 1, \ldots, n$

- Z_i est la réponse (outcome)
- U_i sont les covariables.

On suppose que $Z_i \sim \beta(p_i = \frac{\exp(U_i\beta)}{1+\exp(U_i\beta)})$, qui a alors la densité

$$f(z_i) = \frac{\big(\exp(U_i\beta)\big)^{z_i}}{1+\exp(U_i\beta)}$$

A partir de l'échantillon (Z_i, U_i), on forme alors la log-vraisemblance

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \log(f(Z_i)) = \sum_{i=1}^{n} \Big\{ Z_i U_i\beta - \big(1+\exp(U_i\beta)\big) \Big\},$$

et on définit l'estimateur de β au maximum de vraisemblance

$$\nabla \log \mathcal{L}(\hat{\beta}) = 0 \ \text{ (équation d'estimation).}$$

Modèle quasi-binomial

Avec les mêmes notations, on définit le modèle quasi-binomial via sa log-vraisemblance

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \log(f(Z_i)) = \sum_{i=1}^{n} \left\{ Z_i U_i \beta - \left(1 + \exp(U_i \beta)\right) \right\} \text{ (logistique)},$$

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \log(f(Z_i)) = \sum_{i=1}^{n} \frac{Z_i U_i \beta - \left(1 + \exp(U_i \beta)\right)}{\phi_i} \text{ (quasi-binomial)},$$

MSM logistique

On revient maintenant à notre MSM logistique

$$Y^a \sim \beta\left(\frac{\exp(\psi_0 + \psi_1 a)}{1 + \exp(\psi_0 + \psi_1 a)}\right)$$

on se place dans la pseudo-population est donc défini donc la log-vraisemblance comme

$$\log \mathcal{L}_{\text{weighted}}(\psi_0, \psi_1) = \sum_{i=1}^{n} W_i \Big\{ Y_i \psi_0 + \psi_1 A_i - \Big(1 + \exp(\psi_0 + \psi_1 A_i)\Big) \Big\},$$

avec les poids

$$W_i = \frac{1}{\hat{\pi}_i} \quad \text{pour les traités et} \quad W_i = \frac{1}{1 - \hat{\pi}_i} \quad \text{pour les contrôles.}$$

On définit l'estimateur de ψ_1 au maximum de vraisemblance pondéré par l'équation

$$\nabla \log \mathcal{L}_{\text{weighted}}(\hat{\psi}_0, \hat{\psi}_1) = 0 \ \text{(équation structurelle)}.$$

On retrouve ici la log-vraisemblance d'un **modèle quasi-binomial**.