Causal inference with incomplete confounders

Effect of tranexamic acid on mortality among traumatic brain injury patients

Imke Mayer, Julie Josse

May 19, 2019

CAMS, EHESS; CMAP, X

Overview

Introduction

Missing value website

More information and details on missing values: **R-miss-tastic** platform.

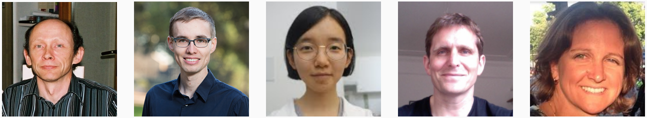
`https://rmisstastic.netlify.com`

→ Theoretical and practical tutorials, popular datasets, bibliography, workflows (in R), active contributors/researchers in the community, etc.

Interested in contribute to our platform? Feel free to contact us!

Collaborators

Jean-Pierre Nadal (ENS-EHESS), Stefan Wager (Stanford), Wei Jiang (X), Nicolas Prost (X)
Traumabase (APHP): Tobias Gauss, Sophie Hamada, Jean-Denis Moyer
Capgemini



Traumabase

20000 patients/ 244 variables/ 16 hospitals, from 2012 (4000 new patients/ year)

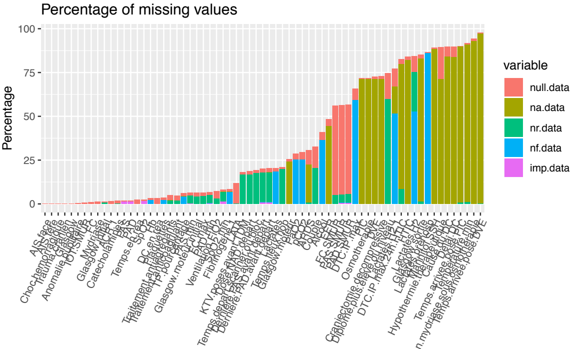
```
              Center       Accident Age Sex Weight Height   BMI BP SBP
1            Beaujon        Fall    54    m    85     NR    NR 180 110
2              Lille        Other   33    m    80    1.8 24.69 130  62
3  Pitie Salpetriere        Gun     26    m    NR     NR    NR 131  62
4            Beaujon      AVP moto  63    m    80    1.8 24.69 145  89
6  Pitie Salpetriere   AVP bicycle  33    m    75     NR    NR 104  86
7  Pitie Salpetriere AVP pedestrian 30    w    NR     NR    NR 107  66
9               HEGP   White weapon 16    m    98   1.92 26.58 118  54
10            Toulon  White weapon  20    m    NR     NR    NR 124  73
...................

   SpO2 Temperature Lactates   Hb  Glasgow Transfusion ..........
1    97        35.6     <NA> 12.7       12          yes
2   100        36.5      4.8 11.1       15          no
3   100          36      3.9 11.4        3          no
4   100        36.7     1.66   13       15          yes
6   100          36       NM 14.4       15          no
7   100        36.6       NM 14.3       15          yes
9   100        37.5       13 15.9       15          yes
10  100        36.9       NM 13.7       15          no
```

⇒ **Estimate causal effect**: administration of the **treatment** "tranexamic acid" (within the first 3 hours after the accident) on mortality (**outcome**) for traumatic brain injury (TBI) patients.

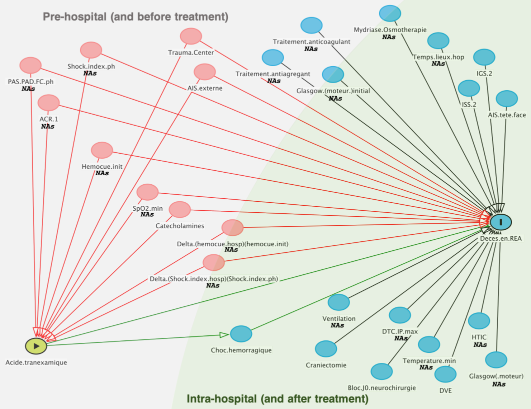
Causal inference for traumatic brain injury with missing values

- 3050 patients with a brain injury (a lesion visible on the CT scan)
- Treatment: tranexamic acid (binary)
- Outcome: in-ICU death (binary), causes: brain death, withdrawal of care, head injury and multiple organ failure.
- 45 **quantitative** & **categorical** covariates selected by experts (Delphi process). Pre-hospital (blood pressure, patients reactivity, type of accident, anamnesis, etc.) and hospital data



Causal inference for traumatic brain injury with missing values

- 3050 patients with a brain injury (a lesion visible on the CT scan)
- Treatment: tranexamic acid (binary)
- Outcome: in-ICU death (binary), causes: brain death, withdrawal of care, head injury and multiple organ failure
- 45 **quant.** & **categorical** covariates (pre-hosp & hosp) selected by experts (Delphi process)
- Treatment target: hemorrhagic shock (HS) → mediator (issue?: translates an underlying processus beginning before treatment (bleeding)).



Graph produced using DAGitty (?)

Outline

⇒ Causal inference

Causal inference methodology: estimate causal relationships between an intervention (acid administration) and an outcome (mortality), when the study is potentially confounded by treatment bias due to the absence of randomization.

⇒ Causal inference with missing values in the covariates

Unconfoundedness assumption possibly violated → modify estimand (generalized propensity score) and add assumptions on relationship between missing values and treatment assignment/outcome.

Causal inference: classical framework

Potential outcome framework (Rubin, 1974)

Causal effect

Binary treatment $w \in \{0, 1\}$ on *i-th* individual with potential outcomes $Y_i(1)$ and $Y_i(0)$. Individual causal effect of the treatment:

$$\Delta_i = Y_i(1) - Y_i(0)$$

- Problem: Δ_i never observed (only observe one outcome/indiv). Causal inference as a missing value pb?

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
1.1	20	F	1	NA	T
-6	45	F	0	F	NA
0	15	M	1	NA	F
	...		...	...	...
-2	52	M	0	T	NA

Potential outcome framework (Rubin, 1974)

Causal effect

Binary treatment $w \in \{0, 1\}$ on *i-th* individual with potential outcomes $Y_i(1)$ and $Y_i(0)$. Individual causal effect of the treatment:

$$\Delta_i = Y_i(1) - Y_i(0)$$

- Problem: Δ_i never observed (only observe one outcome/indiv). Causal inference as a missing value pb?
- **Average treatment effect (ATE)** $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$: The ATE is the difference of the average outcome had everyone gotten treated and the average outcome had nobody got treatment.

$\Rightarrow$ First solution: estimate τ with randomized controlled trials (RCT).

Average treatment effect estimation in RCTs

Assumptions:

Observe n iid samples (Y_i, W_i) each satisfying:

- $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$ (SUTVA)
- $W_i \perp\!\!\!\perp \{Y_i(0), Y_i(1)\}$ (random treatment assignment)

Difference-in-means estimator

$$\hat{\tau}_{DM} = \frac{1}{n_1} \sum_{W_1=1} Y_i - \frac{1}{n_0} \sum_{W_1=0} Y_i$$

Properties of $\hat{\tau}_{DM}$

$\hat{\tau}_{DM}$ is unbiased and $\sqrt{n}$-consistent. $\sqrt{n}\,(\hat{\tau}_{DM} - \tau) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_{DM})$, where $V_{DM} = \frac{Var(Y_i(0))}{\mathbb{P}(W_i=0)} + \frac{Var(Y_i(1))}{\mathbb{P}(W_i=1)}$.

Average treatment effect estimation with Difference-of-Means

Difference-of-Means estimator

- conceptually simple estimator and simple to estimate,
- consistent estimator with asymptotically valid inference,
- but is it the optimal way to use the data for fixed finite n?

Average treatment effect estimation with Difference-of-Means

Difference-of-Means estimator

- conceptually simple estimator and simple to estimate,
- consistent estimator with asymptotically valid inference,
- but is it the optimal way to use the data for fixed finite n?

Average Treatment effect

τ is a **causal parameter**, i.e. property we wish to know about a population. It is not related to the study design or the estimation method.

Randomized trials in the linear model

Idea: assume linearity of the responses $Y_i(0)$ and $Y_i(1)$ in the covariates.

Assumptions

- n iid samples (X_i, Y_i, W_i),
- $Y_i(w) = c_{(w)} + X_i\beta_{(w)} + \varepsilon_i(w)$, $w \in \{0, 1\}$,
- $\mathbb{E}[\varepsilon_i(w)|X_i] = 0$ and $Var(\varepsilon_i(w)|X_i) = \sigma^2$.

and without loss of generality we additionally assume:

- $\mathbb{P}(W_i = 0) = \mathbb{P}(W_i = 1) = \frac{1}{2}$,
- $\mathbb{E}[X] = 0$.

Randomized trials in the linear model

OLS estimator

$$
\begin{aligned}
\hat{\tau}_{OLS} &:= \hat{c}_{(1)} - \hat{c}_{(0)} + \bar{X}(\hat{\beta}_{(1)} - \hat{\beta}_{(0)}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((\hat{c}_{(1)} + X_i\hat{\beta}_{(1)}) - (\hat{c}_{(0)} - X_i\hat{\beta}_{(0)}) \Big),
\end{aligned}
$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the estimators are obtained by OLS for the two linear models.

Randomized trials in the linear model

OLS estimator

$$\begin{aligned}\hat{\tau}_{OLS} &:= \hat{c}_{(1)} - \hat{c}_{(0)} + \bar{X}(\hat{\beta}_{(1)} - \hat{\beta}_{(0)}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left((\hat{c}_{(1)} + X_i\hat{\beta}_{(1)}) - (\hat{c}_{(0)} - X_i\hat{\beta}_{(0)})\right),\end{aligned}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the estimators are obtained by OLS for the two linear models.

Properties of $\hat{\tau}_{OLS}$

- Asymptotic independence of $\hat{c}_{(w)}$, $\hat{\beta}_{(w)}$ and $\bar{X}$ and also

  $$\hat{\tau}_{OLS} - \tau = (\hat{c}_{(1)} - c_{(1)}) - (\hat{c}_{(0)} - c_{(0)}) + \bar{X}(\beta_{(1)} - \beta_{(0)}) + \bar{X}(\hat{\beta}_{(1)} - \hat{\beta}_{(0)} - \beta_{(1)} + \beta_{(0)}).$$

- Noting $V_{OLS} = 4\sigma^2 + (\beta_{(0)} - \beta_{(1)})^T Var(X)(\beta_{(0)} - \beta_{(1)})$, by central limit theorem we get

  $$\sqrt{n}\,(\hat{\tau}_{OLS} - \tau) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_{OLS}).$$

Randomized trials in the linear model

Properties of $\hat{\tau}_{OLS}$

- Noting $V_{OLS} = 4\sigma^2 + \|\beta_{(0)} - \beta_{(1)}\|_A^2$, by central limit theorem we get

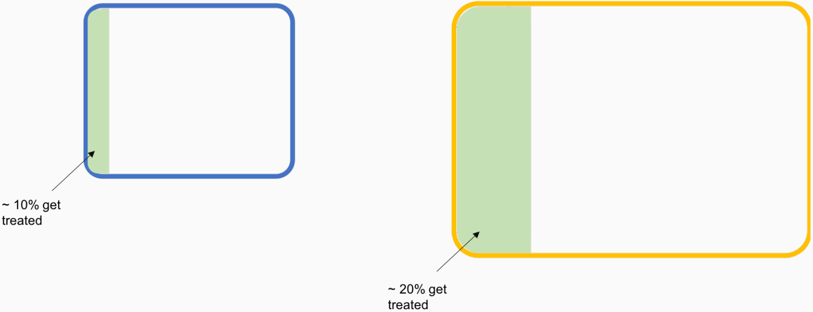
$$\sqrt{n}\,(\hat{\tau}_{OLS} - \tau) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_{OLS}).$$

Remark

- Under the linearity assumption, $V_{DM} = 4\sigma^2 + \|\beta_{(0)} - \beta_{(1)}\|_A^2 + \|\beta_{(0)} + \beta_{(1)})\|_A^2$.
  $\Rightarrow$ $\hat{\tau}_{OLS}$ is always at least as good as $\hat{\tau}_{DM}$ in terms of asymptotic variance.
- This still holds in case of model mis-specification. *(proof uses Huber-White linear regression analysis)*

How to combine different experiments or data sets

Study the effect of a cash incentive to discourage teenagers from smoking in two different cities.



~ 10% get treated

~ 20% get treated

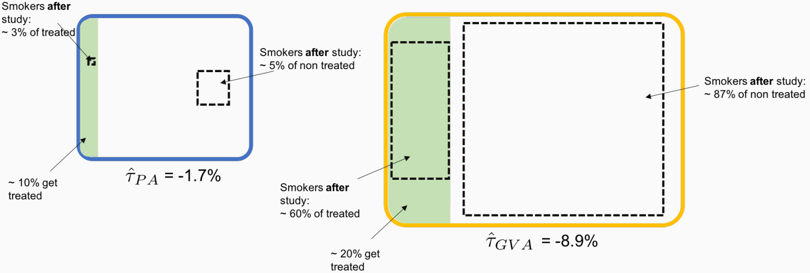
🟦 5% of all smoking teenagers in Palo Alto **before** study (#=2741)

🟨 20% of all smoking teenagers in Geneva **before** study (#=5188)

🟩 Treated teenagers

How to combine different experiments or data sets

Study the effect of a cash incentive to discourage teenagers from smoking in two different cities.



Smokers **after** study: ~ 3% of treated

Smokers **after** study: ~ 5% of non treated

~ 10% get treated

$\hat{\tau}_{PA}$ = -1.7%

Smokers **after** study: ~ 87% of non treated

Smokers **after** study: ~ 60% of treated

~ 20% get treated

$\hat{\tau}_{GVA}$ = -8.9%

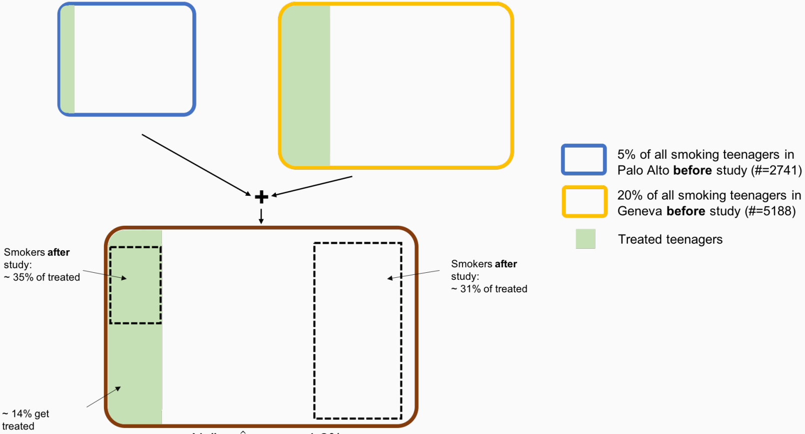
□ 5% of all smoking teenagers in Palo Alto **before** study (#=2741)

□ 20% of all smoking teenagers in Geneva **before** study (#=5188)

🟩 Treated teenagers

How to combine different experiments or data sets

Study the effect of a cash incentive to discourage teenagers from smoking in two different cities.



5% of all smoking teenagers in Palo Alto **before** study (#=2741)

20% of all smoking teenagers in Geneva **before** study (#=5188)

🟩 Treated teenagers

Smokers **after** study:
~ 35% of treated

Smokers **after** study:
~ 31% of treated

~ 14% get treated

Naïve $\hat{\tau}_{both}$ = 4.2%

How to combine different experiments or data sets

Study the effect of a cash incentive to discourage teenagers from smoking in two different cities.

Correct aggregation of the two studies:

$$\hat{\tau}_{both} = \frac{\square}{\square}\,\hat{\tau}_{PA} + \frac{\square}{\square}\,\hat{\tau}_{GVA} = \text{-}6.5\%$$

Aggregating several ATE estimators

How to combine several trials testing the same treatment but on different populations?

Assumptions

- n iid samples (X_i, Y_i, W_i),
- Covariates X_i take values in a **finite discrete** space $\mathcal{X}$ (i.e. $|\mathcal{X}| = p$).
- Treatment assignment is random conditionally on X_i:

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \,|\, X_i = x, \quad \forall x \in \mathcal{X}.$$

Bucket-wise ATE

$$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) \,|\, X_i = x]$$

Results for aggregated difference-in-means estimators

Aggregated difference-in-means estimator

$$\hat{\tau} := \sum_{x \in \mathcal{X}} \frac{n_x}{n} \hat{\tau}(x) = \sum_{x \in \mathcal{X}} \frac{n_x}{n} \left(\frac{1}{n_{x1}} \sum_{\{X_i = x,\, W_i = 1\}} Y_i - \frac{1}{n_{x0}} \sum_{\{X_i = x,\, W_i = 0\}} Y_i \right)$$

- Denoting $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x)$ and adding simplifying assumption $Var(Y(w) \mid X = x) = \sigma^2(x)$ we can show that

$$\sqrt{n_x} \left(\hat{\tau}(x) - \tau(x) \right) \xrightarrow[n \to \infty]{d} \mathcal{N}\left(0, \frac{\sigma^2(x)}{e(x)(1 - e(x))}\right)$$

- Finally, denoting $V_{BUCKET} = Var(\tau(X)) + \mathbb{E}\left[\frac{\sigma^2(X)}{e(X)(1-e(X))}\right]$,

$$\sqrt{n} \left(\hat{\tau} - \tau \right) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, V_{BUCKET})$$

no dependence in p, # of buckets!

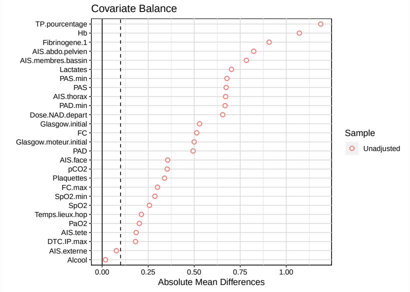
Observational data. Non random assignment: confounding

Mortality rate 16% - treated 28 - not treated 13: treatment kills?

```
              Died        P(Outcome | Treatment)
Treated      0     1           0          1
   FALSE  2225   340       0.867      0.133
    TRUE   436   168       0.722      0.278
```

Strong indication for confounding factors that need to be controlled for.

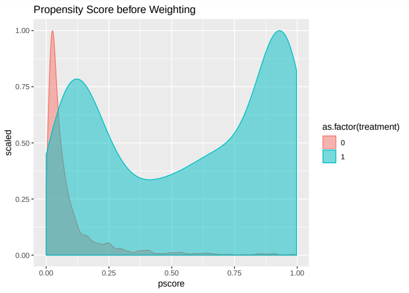
Standardized mean differences between treated and control.

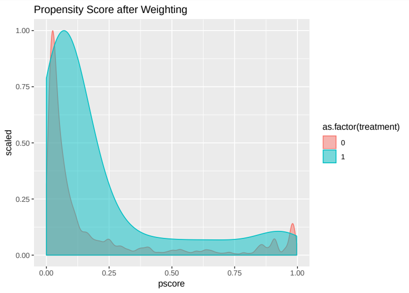


Treated patients are more severe with higher risk of death (graphical model)

Solutions to estimate ATE with observational data

- **Matching**: pair each treated (resp. untreated) patient with one or more similar untreated (resp. treated) patient (R package `Match`)
- **Inverse-propensity weighting**: to adjust for biases in the treatment assignment





- **Double robust methods** for model misspecifications: covariate balancing propensity score, augmented IPW. (Robins *et al.*, 1994)
- **Regression adjustment**, **regression-adjusted matching**, etc.

Unconfoundedness and the propensity score

Assumptions

- n iid samples (X_i, Y_i, W_i),
- Treatment assignment is random conditionally on X_i:
  $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i \quad \equiv$ **unconfoundedness** assumption.

Measure enough covariates to capture any dependence between W_i and the PO

Propensity score and overlap assumption

$$e(x) \triangleq \mathbb{P}(W_i = 1 \mid X_i = x) \quad \forall\, x \in \mathcal{X}.$$

We will assume overlap, i.e. $0 < e(x) < 1 \quad \forall\, x \in \mathcal{X}$.

Key property

e is a balancing score, i.e. under unconfoundedness, it satisfies

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid e(X_i)$$

As a consequence, it suffices to control for $e(X)$ (rather than X), to remove biases associated with non-random treatment assignment.

Unconfoundedness and the propensity score

Propensity score

$e(x) \triangleq \mathbb{P}(W_i = 1 \mid X_i = x) \quad \forall x \in \mathcal{X}.$

Key property

Under unconfoundedness, $e(x)$ satisfies $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid e(X_i)$.

Proof

To prove this balancing property, we note that the distribution of W is fully specified by its mean. Therefore we need to prove that:

$\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, X_i] = \mathbb{E}[W_i|X_i] \Rightarrow \mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, e(X_i)] = \mathbb{E}[W_i|e(X_i)]$

Unconfoundedness and the propensity score

Propensity score

$e(x) \triangleq \mathbb{P}(W_i = 1 \mid X_i = x) \quad \forall x \in \mathcal{X}.$

Key property

Under unconfoundedness, $e(x)$ satisfies $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid e(X_i)$.

Proof

To prove this balancing property, we note that the distribution of W is fully specified by its mean. Therefore we need to prove that:

$$\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, X_i] = \mathbb{E}[W_i|X_i] \Rightarrow \mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, e(X_i)] = \mathbb{E}[W_i|e(X_i)]$$

a) By the law of total expectation we have:

$$\mathbb{E}[W_i|e(X_i)] = \mathbb{E}[\mathbb{E}[W_i|X_i, e(X_i)]|e(X_i)] = \mathbb{E}[\mathbb{E}[W_i|X_i]|e(X_i)] = e(X_i)$$

Unconfoundedness and the propensity score

Propensity score

$e(x) \triangleq \mathbb{P}(W_i = 1 \mid X_i = x) \quad \forall x \in \mathcal{X}.$

Key property

Under unconfoundedness, $e(x)$ satisfies $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid e(X_i)$.

Proof

To prove this balancing property, we note that the distribution of W is fully specified by its mean. Therefore we need to prove that:

$\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, X_i] = \mathbb{E}[W_i|X_i] \Rightarrow \mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, e(X_i)] = \mathbb{E}[W_i|e(X_i)]$

a) By the law of total expectation we have:

$$\mathbb{E}[W_i|e(X_i)] = \mathbb{E}[\mathbb{E}[W_i|X_i, e(X_i)]|e(X_i)] = \mathbb{E}[\mathbb{E}[W_i|X_i]|e(X_i)] = e(X_i)$$

b) And again using the law of total expectation we have the following:

$$\begin{aligned}
\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, e(X_i)] &= \mathbb{E}[\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, X_i, e(X_i)]|\{Y_i(0), Y_i(1)\}, e(X_i)] \\
&= \mathbb{E}[\mathbb{E}[W_i|\{Y_i(0), Y_i(1)\}, X_i]|\{Y_i(0), Y_i(1)\}, e(X_i)] \\
&= \mathbb{E}[\mathbb{E}[W_i|X_i]|\{Y_i(0), Y_i(1)\}, e(X_i)] \ (\textit{unconfoundedness}) \\
&= \mathbb{E}[e(X_i)|\{Y_i(0), Y_i(1)\}, e(X_i)] = e(X_i)
\end{aligned}$$

■

Inverse-propensity weighting estimation of ATE

$$\hat{\tau}_{IPW} \triangleq \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{\hat{e}(X_i)} - \frac{(1-W_i)Y_i}{1-\hat{e}(X_i)}\right)$$

$\Rightarrow$ Balance the differences between the two groups

The quality of this estimator depends on the estimation quality of $\hat{e}(x)$/on the postulated propensity score model. Indeed we have:

$$\begin{aligned}\mathbb{E}\left[\frac{WY}{e(X)}\right] &= \mathbb{E}\left[\frac{WY(1)}{e(X)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{WY(1)}{e(X)} | \, Y(1), X\right]\right] \\ &= \mathbb{E}\left[\frac{Y(1)}{e(X)}\mathbb{E}[W|Y(1),X]\right] = \mathbb{E}\left[\frac{Y(1)}{e(X)}\mathbb{E}[W|X]\right] \\ &= \mathbb{E}\left[\frac{Y(1)}{e(X)}e(X)\right] = \mathbb{E}[Y(1)].\end{aligned}$$

This holds if $e(X) = \mathbb{P}(W=1|X)$, therefore if $\hat{e}(X)$ is not the true propensity score then $\hat{\tau}_{IPW}$ is not necessarily a (consistent) estimate of τ.
Remark: Variance of the oracle estimate is bad!

Covariate balancing propensity score (CBPS)

Assume a linear-logistic model:

1. $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x) = \frac{1}{1+e^{-x^T\alpha}}$
2. $\mu_{(w)}(x) = x^T\beta_{(w)}$ (for $w \in \{0,1\}$).
3. $Y_i(w) = \mu_{(W_i)}(X_i) + \varepsilon_i$.

Decompose ATE $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\gamma}_{(1)}(X_i)W_iY_i - \hat{\gamma}_{(0)}(X_i)(1-W_i)Y_i\right)$:

$$\hat{\tau} = \bar{X}(\beta_{(1)} - \beta_{(0)}) + [\text{term for } \varepsilon] + \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_{(1)}(X_i)W_iX_i - \bar{X}\right)\beta_{(1)} - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_{(0)}(X_i)(1-W_i)X_i - \bar{X}\right)$$

$$= \bar{X}(\beta_{(1)} - \beta_{(0)}) + \frac{W_i(Y_i - \mu_{(1)}(X_i))}{e(X_i)} - \frac{(1-W_i)(Y_i - \mu_{(0)}(X_i))}{1-e(X_i)}$$

What happens when models are mis-specified? Double robustness

For specific $\hat{\gamma}_{(1)}$ and $\hat{\gamma}_{(0)}$ (functions of α), $\hat{\tau}$ is the CPBS and it is doubly robust, i.e. it is consistent in either one of the following cases:

1. Outcome model is linear but propensity score $e(x)$ is not logistic.
2. Propensity score $e(x)$ is logistic but outcome model is not linear.

Propensity score estimation and inverse-propensity weighting

Covariate balancing propensity score (CBPS)

- Use $\hat{\gamma}_{(1)} = \frac{1}{\hat{e}(x)} = 1 + e^{-x^T \hat{\alpha}_{(1)}}$ and solve for $\alpha_{(1)}$ by moment matching:
$$\frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_{(1)}(X_i) W_i X_i - \bar{X} = 0$$
- Same for $\hat{\gamma}_{(0)} = \frac{1}{1-\hat{e}(x)} = \frac{e^{-x^T \hat{\alpha}_{(0)}}}{1+e^{-x^T \hat{\alpha}_{(0)}}}$.

Note that $\hat{\gamma}_{(1)}$ and $\hat{\gamma}_{(0)}$ do not use the same propensity model but we can verify that both $\hat{\alpha}_{(1)}$ and $\hat{\alpha}_{(0)}$ are $\sqrt{n}$-consistent:

$$\|\hat{\alpha}_{(w)} - \alpha\|_2 = \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{for } w \in \{0, 1\}$$

Propensity score estimation and inverse-propensity weighting

IPW with covariate balancing propensity score (CBPS)

Under regularity assumptions (including overlap, i.e. $\exists\eta > 0$ such that $\eta \leq e(x) \leq 1 - \eta$ for all $x \in \mathcal{X}$), we have:

$$\hat{\tau}_{CBPS} = \bar{X}(\beta_{(1)} - \beta_{(0)}) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i\varepsilon_i}{\hat{e}(X_i)} - \frac{(1-W_i)\varepsilon_i}{1-\hat{e}(X_i)}\right) + \mathcal{O}_P\left(\frac{1}{n}\right)$$

And this estimator has same asymptotic variance as for bucketing.

Double robustness of CBPS

Under linear-logistic specification, $\hat{\tau}_{CBPS}$ has "good" asymptotic variance. What happens if the model is mis-specified?

Double robustness

$\hat{\tau}_{CBPS}$ remains consistent in either one of the following cases:

1. Outcome model is linear but propensity score $e(x)$ is not logistic.
2. Propensity score $e(x)$ is logistic but outcome model is not linear.

Note that the asymptotic variance might be different in these cases.

Doubly robust ATE estimation

Define $\mu_{(w)}(x) := \mathbb{E}[Y_i(w) \mid X_i = x]$ and $e(x) := \mathbb{P}(W_i = 1 \mid X_i = x)$.

Doubly robust estimator

$$\hat{\tau}_{DR} := \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + W_i \frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1 - W_i) \frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1 - \hat{e}(X_i)} \right)$$

is consistent if either the $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent.

Furthermore, the oracle $\hat{\tau}_{DR^*}$ has good asymptotic variance (assuming $\mu_{(w)}(\cdot)$ and $e(\cdot)$ known).

Doubly robust ATE estimation

Define $\mu_{(w)}(x) := \mathbb{E}[Y_i(w) \mid X_i = x]$ and $e(x) := \mathbb{P}(W_i = 1 \mid X_i = x)$.

Doubly robust estimator

$$\hat{\tau}_{DR} := \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + W_i \frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1 - W_i) \frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1 - \hat{e}(X_i)} \right)$$

is consistent if either the $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent.

Furthermore, the oracle $\hat{\tau}_{DR^*}$ has good asymptotic variance (assuming $\mu_{(w)}(\cdot)$ and $e(\cdot)$ known).

Example: linear-logistic model

1. $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x) = (1 + e^{-x^T\alpha})^{-1}$
2. $\mu_{(w)}(x) = x^T \beta_{(w)}$ (for $w \in \{0, 1\}$). $\quad Y_i(w) = \mu_{(W_i)}(X_i) + \varepsilon_i$.

What happens when models are mis-specified? Double robustness

$\hat{\tau}_{DR}$ is doubly robust, i.e. it is consistent in either one of the following cases:

1. Outcome model is linear but propensity score $e(x)$ is not logistic.

Doubly robust ATE estimation

Define $\mu_{(w)}(x) := \mathbb{E}[Y_i(w) \,|\, X_i = x]$ and $e(x) := \mathbb{P}(W_i = 1 \,|\, X_i = x)$.

Doubly robust estimator

$$\hat{\tau}_{DR} := \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + W_i \frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1 - W_i) \frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1 - \hat{e}(X_i)} \right)$$

is consistent if either the $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent.

Furthermore, the oracle $\hat{\tau}_{DR^*}$ has good asymptotic variance (assuming $\mu_{(w)}(\cdot)$ and $e(\cdot)$ known).

Remark 1: Possibility to use **any (machine learning) procedure** such as random forests, deep nets, etc. to estimate $\hat{e}(x)$ and $\hat{\mu}_{(w)}(x)$ without harming the interpretability of the causal effect estimation.

Remark 2: In case of overparametrization or non-parametric estimation $\hat{\mu}_{(w)}(x)$ and $\hat{e}(x)$ need be learned/estimated by **cross-splitting** to achieve same performance as oracle $\hat{\tau_{DR^*}}$. Package `grf`. **?**

Semiparametric efficiency for ATE estimation

Efficient score estimator

Given unconfoundedness ($\{Y_i(1), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i$) but no further parametric assumptions on $\mu_{(w)}(x)$ and $e(x)$, the previously attained asymptotic variance,

$$V^* := Var(\tau(X)) + \mathbb{E}\left[\frac{\sigma^2(X)}{e(X)(1 - e(X))}\right],$$

is optimal and any estimator τ^* that attains it is asymptotically equivalent to $\hat{\tau}_{DR^*}$.

V^* is the semiparametric efficient variance for ATE estimation.

Semiparametric*: we are interested in a parametric estimand, τ, which we estimate using nonparametric estimates ($\hat{\tau}_{DR}$ depends on nonparametric estimates $\hat{\mu}_{(w)}(x)$ and $\hat{e}(x)$). See* **??** *for more details.*

Cross-fitting for ATE estimation

Cross-fitted ATE estimator

Assume we divide the data into K folds.

$$\hat{\tau}_{CF} = \frac{1}{n}\sum_{i=1}^{n}\Bigg(\hat{\mu}_{(1)}^{(-k(i))}(X_i) - \hat{\mu}_{(0)}^{(-k(i))}(X_i) + W_i \frac{Y_i - \hat{\mu}_{(1)}^{(-k(i))}(X_i)}{\hat{e}^{(-k(i))}(X_i)} - (1 - W_i)\frac{Y_i - \hat{\mu}_{(0)}^{(-k(i))}(X_i)}{1 - \hat{e}^{(-k(i))}(X_i)}\Bigg),$$

where $k(i)$ maps an observation X_i to one of the K folds and $\hat{\mu}^{(-j)}$ indicates that the estimator has been learned on all the folds except the j-th fold.

Assuming overlap, sup-norm consistency of all used machine learning adjustments and risk decay, we have

$$\sqrt{n}\left(\hat{\tau}_{CF} - \hat{\tau}_{DR^*}\right) \xrightarrow[n\to\infty]{p} 0.$$

And we can prove that we can build level-α confidence intervals for τ.

Causal inference: with missing confounder values?

Unconfoundedness with missing confounder values?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
NA	20	F	1	NA	T
-6	45	NA	0	F	NA
0	NA	M	1	NA	F
NA	32	F	1	NA	T
1	63	M	1	F	NA
-2	NA	M	0	T	NA

Unconfoundedness with missing confounder values?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

Covariates			Treatment	Outcome
X_1	X_2	X_3	W	Y
NA	20	F	1	T
-6	45	NA	0	F
0	NA	M	1	F
NA	32	F	1	T
1	63	M	1	F
-2	NA	M	0	T

Unconfoundedness with missing confounder values?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

Covariates			Treatment	Outcome
X_1	X_2	X_3	W	Y
NA	20	F	1	T
-6	45	NA	0	F
0	NA	M	1	F
NA	32	F	1	T
1	63	M	1	F
-2	NA	M	0	T

Unconfoundedness with missing confounder values?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

→ Often not a good idea! What are the alternatives?

Idea 1

Assume MAR mechanism and do multiple imputation using (X, W, Y) (???).
Problem: can we use Y for propensity score estimation?

Underlying assumptions (?):

- $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i \,|\, X_i$
- overlap (i.e. treatment probability bounded away from 0 and 1)
- X is MAR given Y and W (i.e. $\mathbb{P}(R|X, Y, W) = \mathbb{P}(R|X^{obs}, Y, W)$)

Notations

- covariates $X \in \mathcal{X}$ (wheredim$(\mathcal{X}) = p$),
- response pattern $R \in \{0,1\}^p$ which is defined as $R_j = \mathbb{1}_{\{X_j \text{ is observed}\}}$,
- $X = (X^{obs}, X^{mis})$ where $X^{obs} = \{X_j : R_j = 1\}$

Unconfoundedness with missing confounder values?

Idea 2

Assume logistic-linear generating model and MAR mechanism (i.e. $\mathbb{P}(R|X) = \mathbb{P}(R|X^{obs})$) and perform logistic and linear regressions handling missing values (?). $\hat{\tau}_{saem}$

Problem: Strong model dependence.

Details:

- SAEM on (X^{obs}, W) → estimate logistic regr. parameter by $\hat{\theta}$ → predict $\mathbb{P}(W = 1|X^{obs})$
- EM on (X^{obs}, Y) → estimate μ and Σ → impute X using $(\hat{\mu}, \hat{\Sigma})$ → X_{imp}.
- Linear regression of Y on X_{imp} in each group (treated and control).

Assumptions for Idea 2 to work for ATE estimation:

- $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i \,|\, X_i^{obs}, R$
- overlap
- $\mathbb{P}(R|X, Y(0), Y(1), W) = \mathbb{P}(R|X^{obs})$

Unconfoundedness with missing confounder values?

Idea 3

Adapt the initial assumptions s.t. treatment assignment is unconfounded given only the **observed** covariates and the **response pattern**.

? suggest adapting the propensity score to missing values and using pattern mixture model to estimate it:

Generalized propensity score

$e^*(X^{obs}, R) = \mathbb{P}(W = 1 \,|\, X^{obs}, R).$

Generalized propensity score

Generalized propensity score

$e^*(X^{obs}, R) = \mathbb{P}(W = 1 \,|\, X^{obs}, R)$.

→ Allows to balance treatment and control groups **on the observed covariates** in the case of missing values under:

Assumptions

Treatment is unconfounded given X^{obs} and R:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i \,|\, X_i^{obs}, R_i, \tag{1}$$

or alternatively:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i \,|\, (X_i^{obs}, X_i^{mis}), R_i,$$

$$\begin{cases} \text{CIT:} & W_i \perp\!\!\!\perp X_i^{mis} \,|\, X_i^{obs}, R_i \\ \textbf{or} & \\ \text{CIO:} & Y_i(t) \perp\!\!\!\perp X_i^{mis} \,|\, X_i^{obs}, R_i \quad \text{for } t \in \{0,1\} \end{cases} \tag{2}$$

Generalized propensity score

Under previous assumptions, estimate e^* using **missingness pattern approach** (MPA), i.e. estimate one propensity model per pattern.

$\rightarrow$ Often impossible in practice if p (moderately) large (w.r.t. n)

$\rightarrow$ too few samples per pattern.

Possible workarounds:

- smooth the model by using all available observations for estimating the model for a given pattern and only use the estimated propensity scores for units with this exact pattern (**?**).

  e.g. for $r = (1, 1, 0, 0, 1)$, use all x_i^{obs} with $r_i = (1, 1, *, *, 1)$ where $*$ can be either 0 or 1.

- missing indicator method (**?**): one approach by (**?**) model the joint distribution (X, W, R) by log-linear specification, using the general location model:

  ((W,R) are assumed to be iid multinomial r.v. (Π=cell $\mathbb{P}$ from multinomial), and conditional on (W_i, R_i), X_i is p-variate normal with mean depending on the cell but with common covariance (Γ=cell means, Ω=covariance matrix) $\rightarrow \theta = (\Pi, \Gamma, \Omega)$. Using EM $\rightarrow \hat{\theta}$ for joint distribution (X, W, R). One E-step with $\hat{\theta}$ and setting W_i to missing obtain $\hat{e^*}$.)

  another approach (?): adapting MIM II from (**?**)?

- random forests with **missingness incorporated in attributes**, MIA (**??**).

Random forests with missing values

Under previous assumptions, estimate e^* using **missingness pattern approach** (MPA), i.e. estimate one propensity model per pattern.
$\rightarrow$ Often impossible in practice if p (moderately) large (w.r.t. n)
$\rightarrow$ too few samples per pattern.

Random forests with missing values

Under previous assumptions, estimate e^* using **missingness pattern approach** (MPA), i.e. estimate one propensity model per pattern.
→ Often impossible in practice if p (moderately) large (w.r.t. n)
→ too few samples per pattern.

Random forests allow incorporating missing values directly since they allow semi-discrete variables (e.g. $X \in (\mathbb{R} \times \{\mathtt{NA}\})^p$).

With MIA or mean imputation, splits are possible either on observed variables or on response pattern. **?**
→ define doubly robust $\hat{\tau}_{mia}$

Nonparametric estimation → more modelling flexibility for propensity (and outcome) model.

Simulations: performance of $\hat{\tau}_{saem}$

Data generating model

We simulate continuous confounders Z, specify a logistic propensity model $W \sim Z$ and a linear outcome model $Y \sim Z$.

Missing values mechanisms: MCAR or MAR (given X^{obs}), with 50% of missing values.

We generate $Z^i = [Z_1^i \;\; Z_2^i \;\; Z_{10}^i]^T \sim \mathcal{N}(0, \Sigma)$, $i \in \{1, \ldots, n\}$, where $\Sigma = I - \rho(I - 1)$, with $\rho \in \{0.3, 0.6\}$. $\quad \rightarrow \mathbf{Z} = [Z^1 \ldots Z^n]^T \in \mathbb{R}^{n \times 10}$.

Covariates

Similar to (**?**), we define some nonlinear transformations X of Z, serving as covariates to assess the robustness to mis-specification: We define some nonlinear transformations of the covariates **Z** which are the actually observed covariates **X**.

$$X_{i1} = \exp(Z_{i1}/2) \qquad X_{i2} = \frac{Z_{i2}}{1+\exp(Z_{i1})} + 10 \qquad X_{i3} = \left(\frac{Z_{i1}Z_{i3}}{25} + 0.6\right)^3$$

Simulations: performance of $\hat{\tau}_{saem}$

Data generating model

Missing values mechanisms: MCAR or MAR given X^{obs}, with 50% of missing values.

We generate $Z^i = [Z^i_1 \;\; Z^i_2 \;\; Z^i_{10}]^T \sim \mathcal{N}(0, \Sigma)$, $i \in \{1, \ldots, n\}$, where $\Sigma = I - \rho(I - 1)$, with $\rho \in \{0.3, 0.6\}$. $\quad \rightarrow \mathbf{Z} = [Z^1 \; \ldots \; Z^n]^T \in \mathbb{R}^{n \times 10}$.

CIT: $W \sim Z \odot R$ (where $R_{ij} = \mathbb{1}_{\{Z_{ij} \textit{ is observed}\}}$ and $\odot$ = Hadamard product).

Example: for fixed $\alpha \in \mathbb{R}^4$ and $\tau \in \mathbb{R}$:

$r^i = (1, 1, 0, 0, 0, 1, 0, 0, 0, 1) \Rightarrow$

$logit(\mathbb{P}(W^i = 1 | Z^i_{obs} = z^i_{obs}, R^i = r^i)) = \alpha_0 + \alpha_1 z^i_1 + \alpha_2 z^i_2 + \alpha_6 z^i_6 + \alpha_{10} z^i_{10}$

$r^j = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0) \Rightarrow logit(\mathbb{P}(W^j = 1 | Z^j_{obs} = z^j_{obs}, R^i = r^j)) = \alpha_0 + \alpha_2 z^j_2$

$\neg$**CIT**: $logit(\mathbb{P}(W^i = 1 | Z^i = z^i)) = \alpha_0 + \alpha^T z^i$.

CIO: $Y \sim Z \odot R$.

Example: for fixed $\beta \in \mathbb{R}^4$ and $\tau \in \mathbb{R}$:

$r^i = (1, 1, 0, 0, 0, 1, 0, 0, 0, 1) \Rightarrow$

$\mathbb{E}(Y^i | Z^i_{obs} = z^i_{obs}, R^i = r^i, W^i = w^i) = \beta_0 + \beta_1 z^i_1 + \beta_2 z^i_2 + \beta_6 z^i_6 + \beta_{10} z^i_{10} + \tau w^i$

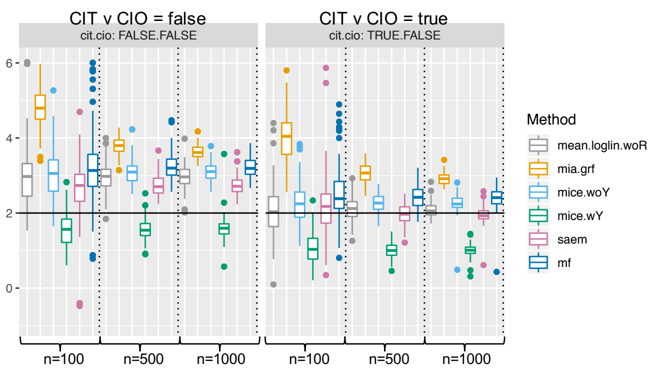
$r^j = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0) \Rightarrow$

$\mathbb{E}(Y^j | Z^j_{obs} = z^j_{obs}, R^i = r^j, W^j = w^j) = \beta_0 + \beta_2 z^j_2 + \tau w^j$

Simulations: performance of $\hat{\tau}_{saem}$

Similar results between $\rho = 0.3$ and $\rho = 0.6$ and also between MCAR and MAR.

Figure 1: Estimated and true average treatment effect ($\tau = 2$, MAR, $\rho = 0.6$)



Remarks:

1. $\hat{\tau}_{saem}$ unbiased under CIT/CIO assumption.
2. Slow convergence of $\hat{\tau}_{mia}$ under CIT/CIO assumption(?)
3. For MI (mice), bias by using/ignoring outcome Y depends on CIT/CIO assumption.

Simulations: performance of $\hat{\tau}_{mia}$

Data generating model

- Case 1: Latent confounder classes
  Class labels $C \sim \mathcal{M}(3) \in \{1, 2, 3\}$ (multinomial distribution) and covariates $Z^i \sim \mathcal{N}_p(\mu(c), \Sigma(c)) \in \mathbb{R}^{n \times p} \mid C^i = c$. ($\mu(c), \Sigma(c)$ are fixed a priori).
- Case 2: Deep latent variable model
  Codes $C \sim \mathcal{N}_d(0, 1)$ (here: $d = 3$) and covariates $Z^i =\sim \mathcal{N}_p(\mu(c), \Sigma(c)) \mid C^i = c$, $i \in \{1, \dots, n\}$, where $(\mu(c), \Sigma(c)) = (V \tanh(Wc + a) + b), \exp(\gamma^T(Wc + a) + \delta)I_p)$.
- $n \in \{100, 500, 1000\}$, $N = 100$ (# of replications).

Logistic propensity model $W \sim Z$, linear outcome model $Y \sim Z$.

Missing values mechanisms: MCAR, MNAR (50% missing values).

Application: Traumabase

Many choices, issues in practice....

- Coding issues: recode certain not really missing values, for ex Glasgow score ($\in \{3, \ldots, 15\}$) is missing for deceased patients. Recode by a category or a constant (lower bound min(GCS)=3).

Many choices, issues in practice....

- Coding issues: recode certain not really missing values, for ex Glasgow score ($\in \{3, \ldots, 15\}$) is missing for deceased patients. Recode by a category or a constant (lower bound min(GCS)=3).
- Impute with iterative FAMD (out-of-range imputation), Random forests (computational costly), mice (invertibility pbs with many categories)?
- Which observations? All individuals (TBI and no-TBI patients)
- Which variables ? All available variables or the experts' pre-selection?
- Impute with treatment, covariates and outcome?

Many choices, issues in practice....

- Coding issues: recode certain not really missing values, for ex Glasgow score ($\in \{3, \ldots, 15\}$) is missing for deceased patients. Recode by a category or a constant (lower bound min(GCS)=3).
- Impute with iterative FAMD (out-of-range imputation), Random forests (computational costly), mice (invertibility pbs with many categories)?
- Which observations? All individuals (TBI and no-TBI patients)
- Which variables ? All available variables or the experts' pre-selection?
- Impute with treatment, covariates and outcome?

Imputation (FAMD) + IPW: $\left(\sum_{i=1}^{n} \frac{W_i}{\hat{e}(X_i)}\right)^{-1} \sum_{i=1}^{n} \frac{W_i Y_i}{\hat{e}(X_i)} - \left(\sum_{i=1}^{n} \frac{1-W_i}{1-\hat{e}(X_i)}\right)^{-1} \frac{(1-W_i) Y_i}{1-\hat{e}(X_i)}$ Model treatment on covariates $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x)$ weights: **GLM, GRF, GBM**. Trimming (0.1% & 99.9% quantiles for weight thresholding).

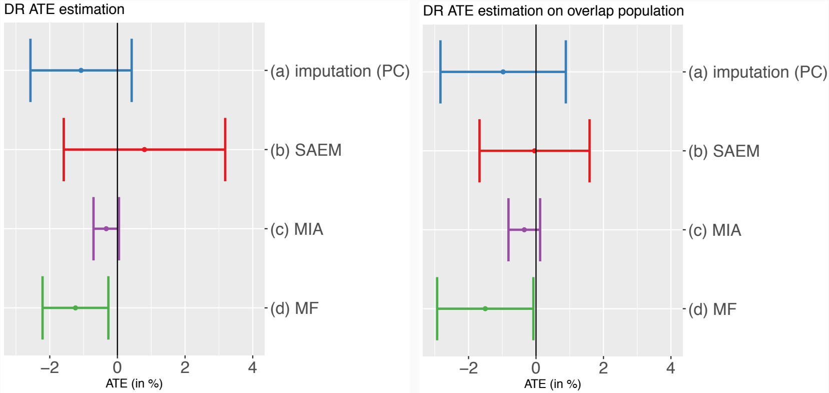
SAEM (quanti) + IPW (weights glm, trimming)

Imputation (FAMD) + double robust: models outcome on covariates and treatment on covariates (**GLM, RF, GBM**)

mia+grf + double robust: models outcome on covariates and treatment on covariates with mia.

Results

ATE estimations for the effect of tranexamic acid on in-ICU mortality for TBI patients. Imputations/SAEM on all patients (TBI + no-TBI).



DR ATE estimation

DR ATE estimation on overlap population

(y-axis: estimation approach), (x-axis: ATE estimation with sandwich CI (see ? for details))
We compute the mortality rate in the treated group and the mortality rate in the control group (after covariate balancing). The obtained value corresponds to the **difference in percentage points between mortality rates in treatment and control**.

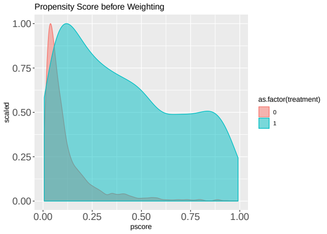
Overlap population: **overlap weights** to adjust for insufficient overlap (? for details).

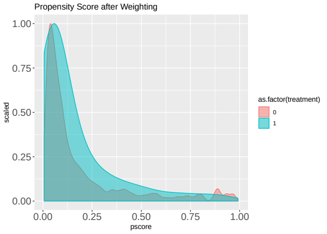
Results: Propensity scores before and after weighting

before

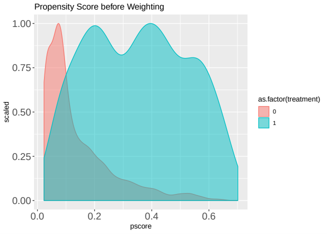
after

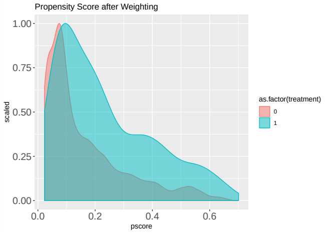
SAEM





MIA

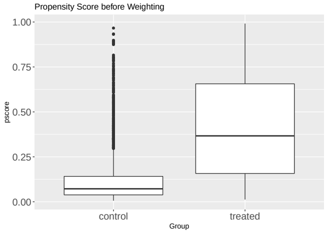


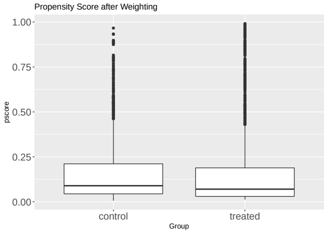


Results: Propensity scores before and after weighting

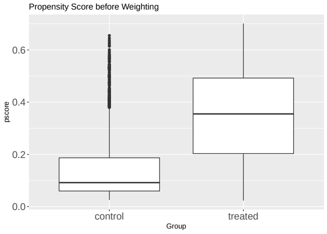
before

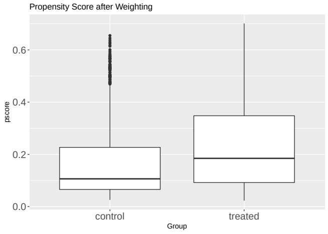
after


Propensity Score before Weighting


Propensity Score after Weighting

SAEM

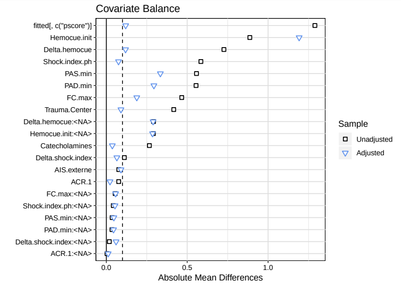

Propensity Score before Weighting


Propensity Score after Weighting

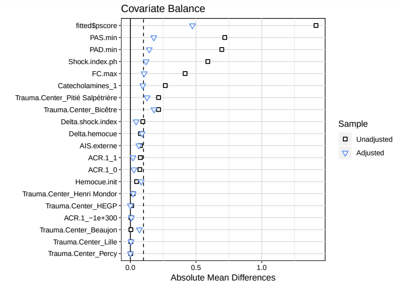
MIA

Results: Standardized mean differences

SAEM



MIA



In our case, covariate balance is better when reweighting with propensity scores estimated using random forests with MIA than those estimated with SAEM.

Conclusion

Conclusion

Today we have seen:

- that experimental data (from **randomized controlled trials**) are gold standard for answering causal questions,
- that drawing causal inferences from **observational data** is possible,
- that **missing confounder values** alter causal analyses,
- additional assumptions guaranteeing **unconfoundedness given missing values**,
- two solutions two handle missing values in causal inference,
- a first small application on real data.

On-going work, perspectives

Methodology/Theory

- Different missing values mechanism. Sportisse, A., Boyer, C. and Josse, J. Low-rank estimation with missing non at random data.
- Logistic regression for mixed variables. ongoing work, Jiang, W., Pichon, M., Josse, J.
- Identify subgroups of patients who could benefit from treatment? Optimal Prescription Trees (Bertsimas et al., 2018).
- Heterogeneous treatment effects (Athey and Imbens, 2015) and optimal policy learning (Imai and Ratkovic, 2013).
- Towards more complex treatment strategies: Do certain treatment strategies, i.e. bundles of treatments (administration of noradrenaline and SSH and tranexamic acid, etc.), have an effect on 24h mortality, on 14d mortality, etc.?
- Consistency of ATE estimator $\hat{\tau}_{mia}$ for missing confounder values.

On-going work, perspectives

Traumabase - Traumatic brain injury

- Bias of mortality (dead before receiving?)
- Plausibility of unconfoundedness?
- Role/Definition of hemorrhagic shock.
- Choice of pre- and post-treatment covariates. Depending on future application. Ideally real-time treatment decision → learning optimal treatment policies.
- Compare results to the ones from CRASH3 study (?, not published yet).

Do you have any questions or comments?

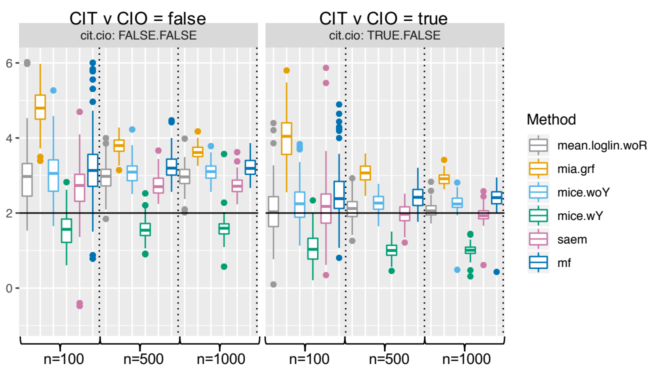
References

References i

Simulations 1

Similar results between $\rho = 0.3$ and $\rho = 0.6$ and also between MCAR and MAR.

Figure 2: Estimated ATE ($\tau = 2$, **MCAR**, $\rho = 0.6$)



Remarks:

1. $\hat{\tau}_{saem}$ unbiased under CIT/CIO assumption.
2. For MI (mice), amount of bias by using/ignoring outcome Y depends on CIT/CIO assumption.