Summary

Baoyi Shi

1 Supported Data Types and Functionalities

rb: Regression-based Approach

wb: Weighting-based Approach

msm: Marginal Structural Model

iorw: Inverse Odds Ratio Weighting Approach

ne: Natural Effect Model

g-formula: G-formula Approach

Table 1: Supported Data Types and Functionalities

	rb	wb	msm	iorw	ne	g-formula
Linear Y	√	√	√	√	√	√
Logistic Y	√	√	√	√	√	√
Loglinear Y	√	√	√	√	√	√
Poisson Y	√	√	√	√	√	√
Quasipoisson Y	√	√	√	√	√	√
NegBin Y	√	√	√	√	×	√
Coxph Y	√	√	√	√	×	√
AFT Exp Y	√	√	√	√	×	√
AFT Weibull Y	√	√	√	√	×	√
Linear M	√	√	√	√	√	√
Logistic M	√	√	√	√	√	√
Categorical M	√	√	√	√	√	√
Any type M	×	√	×	√	√	×
User-defined Y/M Models	√	√	√	√	×	√
Single M	√	√	√	√	√	√
Multiple M	√	√	×	√	√	√
Pre-exposure Confounding	√	√	√	√	√	√
Post-exposure Confounding	×	×	√	×	×	√
2-way Decomposition	√	√	√	√	√	√
4-way Decomposition	√	√	√	×	×	√
Estimation: Closed-form Parameters Function	√	×	√	√	×	×
Estimation: Direct Imputation	√	√	√	√	√	√
Inference: Delta Method	√	×	√	√	×	×
Inference: Bootstrapping	√	√	√	√	√	√

2 Estimation and Inference

2.1 Estimation Method

Closed-form Parameters Function

Effect estimates are calculated using regression parameters.

Direct Conterfacturals imputation

1. For subject i, simulate $M_{a,i}$ by the predicted values of mediator regression models under $A = a, C = C_i$ and simulate $M_{a^*,i}$ by the predicted values of mediator regression models under $A = a^*, C = C_i$.
2. For subject i, simulate $Y_{aMa,i}$ by the predicted value of the outcome regression model under $A = a, M = M_{a,i}, C = C_i$, simulate $Y_{aMa^*,i}$ by the predicted value of the outcome regression model under $A = a, M = M_{a^*,i}, C = C_i$, simulate $Y_{a^*Ma,i}$ by the predicted value of the outcome regression model under $A = a^*, M = M_{a,i}, C = C_i$ and simulate $Y_{a^*Ma^*,i}$ by the predicted value of the outcome regression model under $A = a^*, M = M_{a^*,i}, C = C_i$.
3. Estimate $E[Y_{aMa}]$, $E[Y_{a^*Ma^*}]$, $E[Y_{aMa^*}]$, and $E[Y_{a^*Ma}]$ by $\frac{\sum_{i=1}^{N} Y_{aMa,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{a^*Ma^*,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{aMa^*,i}}{n}$ and $\frac{\sum_{i=1}^{N} Y_{a^*Ma,i}}{n}$ respectively.

2.2 Inference Method

Delta Method

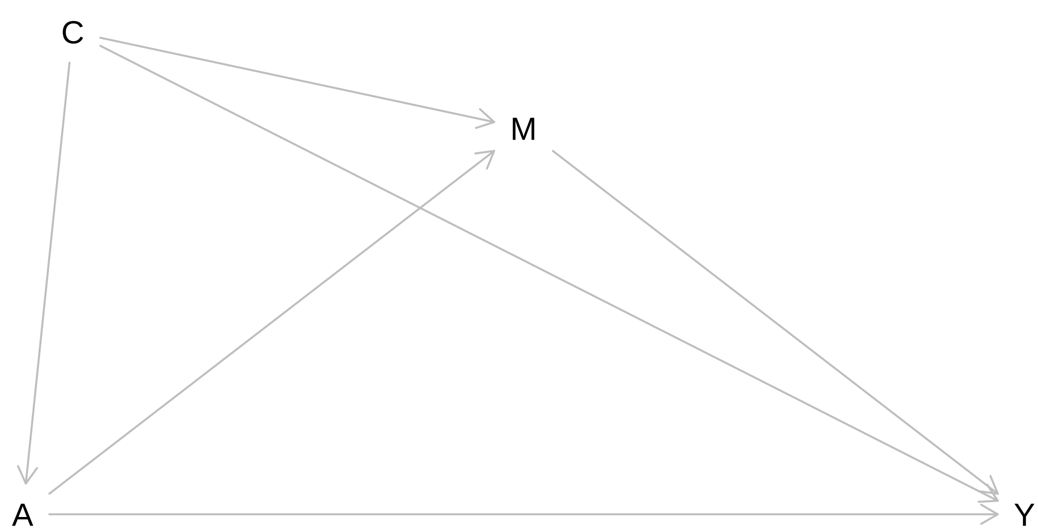
Standard errors of effects are estimated using the standard errors of regression parameters and delta method based on the closed-form parameters function.

Bootstrapping

Bootstrap the data, refit the regression models, and calculate a bootstrap estimate. Repeat the bootstrapping K times and calculate the standard error of these K boostrap estimates for each estimand, which is the estimated standard error of this estimand.

3 Pre-treatment Confounding

3.1 DAG



In the DAG, A denotes the treatment, Y denotes the outcome, M denotes a set of mediators, and C denotes a set of pre-treatment covariates.

3.2 Estimand

2-way decomposition in additive scale

$CDE = E[Y_{am} - Y_{a^*m}|C]$

$PNDE = E[Y_{aM_a^*} - Y_{a^*M_a^*}|C]$

$TNDE = E[Y_{aM_a} - Y_{a^*M_a}|C]$

$PNIE = E[Y_{a^*M_a} - Y_{a^*M_a^*}|C]$

$TNIE = E[Y_{aM_a} - Y_{aM_a^*}|C]$

$TE = PNDE + TNDE$

2-way decomposition in RR scale

$rr\_CDE = E[Y_{am}|C]/E[Y_{a^*m}|C]$

$rr\_PNDE = E[Y_{aM_a^*}|C]/E[Y_{a^*M_a^*}|C]$

$rr\_TNDE = E[Y_{aM_a}|C]/E[Y_{a^*M_a}|C]$

$rr\_PNIE = E[Y_{a^*M_a}|C]/E[Y_{a^*M_a^*}|C]$

$rr\_TNIE = E[Y_{aM_a}|C]/E[Y_{aM_a^*}|C]$

$rr\_TE = rr\_PNDE \times rr\_TNDE$

4-way decomposition in additive scale

CDE is defined above

$INT_{ref} = PNDE - CDE$

$INT_{med} = TNIE - PNIE$

$PIE = PNIE$

4-way decomposition in RR scale

$err\_CDE = (E[Y_{am} - Y_{a^*m}|C])/E[Y_{a^*M_a^*}|C]$

$err\_INT_{ref} = rr\_PNDE - 1 - err\_CDE$

$err\_INT_{med} = rr\_TNIE * rr\_PNDE - rr\_PNDE - rr\_PNIE + 1$

$err\_PIE = rr\_PNIE - 1$

$err\_TE = err\_CDE + err\_INT_{ref} + err\_INT_{med} + err\_PIE$

3.3 Estimation Approaches Can be Used

3.3.1 Regression-based Approach

Reference:

https://www.ncbi.nlm.nih.gov/pubmed/23379553

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287269/

Procedures

1. Fit a regression model for Y on A, M and C.
2. Fit a regression model for each mediator in M on A and C.

3. Estimate the point estimates of causal effects using closed-form parameters functions or direct Conterfacturals imputation.
4. Estimate the standard errors of causal effects using delta method or bootstrapping.

3.3.2 Weighting-based Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287269/

Procedures

1. Fit a regression model for Y on A, M, C.
2. For $E[Y_{a^*Ma^*}]$, estimate it by taking a weighted average of the subjects with $A = a^*$ and each subject i is given a weight $\frac{P(A=a^*)}{P(A=a^*|c_i)}$.
3. For $E[Y_{aMa}]$, estimate it by taking a weighted average of the subjects with $A = a$ and each subject i is given a weight $\frac{P(A=a)}{P(A=a|c_i)}$.
4. For $E[Y_{aMa^*}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a^*$ and each subject i is given a weight $\frac{P(A=a^*)}{P(A=a^*|c_i)}$.
5. For $E[Y_{a^*Ma}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a$ and each subject i is given a weight $\frac{P(A=a)}{P(A=a|c_i)}$.
6. Use bootsrapping to estimate the standard error for each estimand.

3.3.3 Inverse Odds Ratio Weighting Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3954805/

https://www.ncbi.nlm.nih.gov/pubmed/25693776

Procedures

1. Fit a model for A given M and C
2. Calculate the stablized weights for each subject, $w_i = \frac{f_{A|M,C}(A=0|M_i,C_i)}{f_{A|M,C}(A=A_i|M_i,C_i)}$.
3. Estimate the direct effect by a weighted regression model of Y on A and C using the weights calculated in 2. The estimated direct effect is the coefficient of A in this regression model.
4. Estimate the total effect by a regression model of Y on A and C. The estimated total effect is the coefficient of A in this regression model.
5. Calculate the indirect effect by subtracting the direct effect from the total effect.

3.3.4 Natural Effect Model

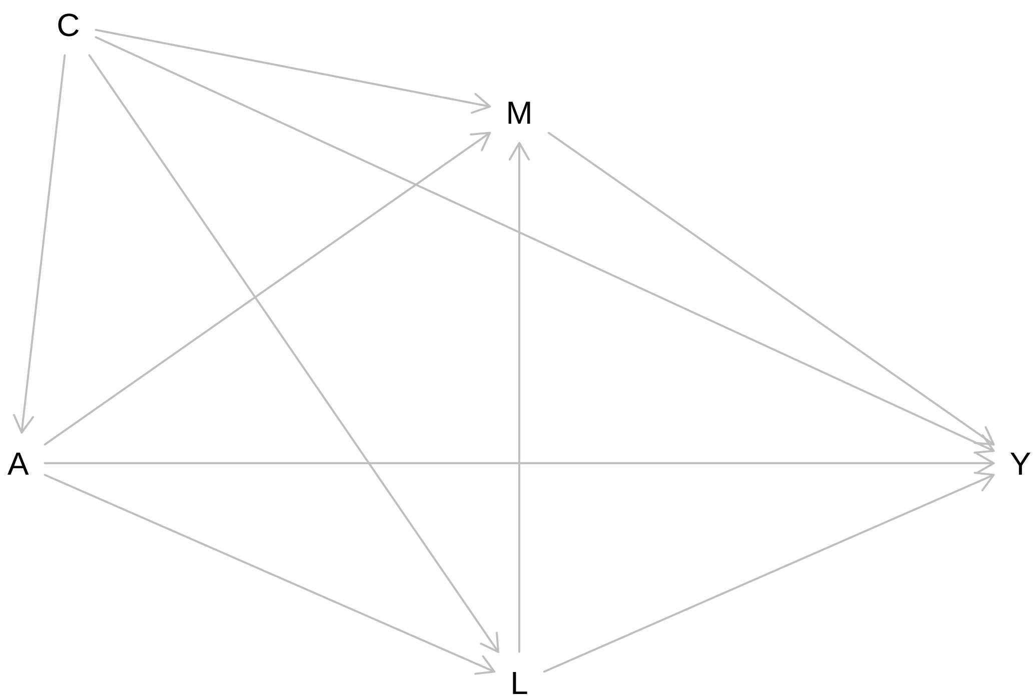
Incorporate the Medflex package.

3.3.5 Other Approaches

Approaches talked about later can also be used.

4 Post-treatment Confounding

4.1 DAG



In the DAG, A denotes the treatment, Y denotes the outcome, M denotes a set of mediators, C denotes a set of pre-treatment covariates and L denotes a set of post-treatment covariates.

4.2 Estimand

$rNDE = E(Y_{aGa^*}) - E(Y_{a^*Ga^*}) = \sum_{l,m,c}\{E[Y|a,l,m,c]P(l|a,c) - E[Y|a^*,l,m,c]P(l|a^*,c)\}P(m|a^*,c)P(c)$

$rNIE = E(Y_{aGa}) - E(Y_{aGa^*}) = \sum_{l,m,c}\{E[Y|a,l,m,c]P(l|a,c)\{P(m|a,c) - P(m|a^*,c)\}P(c)$

$rTE = rNDE + rNIE$

4.3 Estimation Approaches Can be Used

4.3.1 G-formula Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5285457/

Estimation Algorithm for $E(Y_{a1Ga2})$

Let $L = (L_1, L_2, ..., L_p)$ and $M = (M_1, M_2, ..., M_n)$.

1. Bootstrap the dataset.

2. Fit models for $E(L_1|A,C)$, $E(L_2|L_1,A,C)$,$\ldots$, $E(L_p|L_1,L_2,...,L_{p-1},A,C)$ using the bootstrapped dataset. Then, for each subject i, simulate the value of L^*_{1i} under A=a1 by the predicted value of $L_{1i}|A=a1, C=C_i$; simulate the value of L^*_{2i} under A=a1 by the predicted value of $L_2|L^*_{1i}, A=a1, C=C_i$; $\ldots$; simulate the value of L^*_{pi} under A=a1 by the predicted value of $L_p|L^*_{p-1,i},...,L^*_{2i},L^*_{1i},A=a1,C=C_i$.
3. Fit models for $E(M_1|A,C)$, $E(M_2|M_1,A,C)$,$\ldots$,$E(M_n|M_1,M_2,...,M_{n-1},A,C)$ using the bootstrapped dataset. Then, for each subject i, simulate the value of M^*_{1i} under A=a2 by the predicted value of $M_{1i}|A=a2, C=C_i$; simulate the value of M^*_{2i} under A=a2 by the predicted value of $M_2|M^*_{1i}, A=a2, C=C_i$; $\ldots$; simulate the value of M^*_{ni} under A=a2 by the predicted value of $M_n|M^*_{n-1,i},...,M^*_{2i},M^*_{1i},A=a2,C=C_i$.
4. Fit models for E(Y|A,L,M,C) using the bootstrapped dataset. Then, for each subject i, simulate the potential outcome Y^*_i by the predicted value of $Y|A=a1, L_p=L^*_{pi},...,L_2=L^*_{2i},L_1=L^*_{1i},M_n=M^*_{ni},...,M_2=M^*_{2i},M_1=M^*_{1i},C=C_i$.
5. Calculate mean value of Y^*, i.e, $\sum_{i=1}^{N} Y^*_i$
6. Repeat 1-5 K times and estimate $E(Y_{a1Ga2})$ as $\sum_{k=1}^{K}\sum_{i=1}^{N} Y^*_{ik}$.

4.3.2 Marginal Structual Model

Reference:

https://journals.lww.com/epidem/fulltext/2009/01000/marginal_structural_models_for_the_estimation_of.6.aspx

Procedures

For rCDE:

1. Fit a weighted regression model for Y on A and M where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)}\frac{P(M=m_i|A=a_i)}{P(M=m_i|A=a_i,C=c_i,L=l_i)}$.
2. Get the point estimate using $CDE = E[Y_{am} - Y_{a^*m}]$ and get the standard error using bootstrapping.

For rNDE and rNIE:

1. Fit a weighted regression model for Y on A, M and C where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)}\frac{P(M=m_i|A=a_i)}{P(M=m_i|A=a_i,C=c_i,L=l_i)}$.
2. Fit a weighted regression model for M on A and C where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)}$.
3. Get the point estimate and standard error using the simulation-based approach.