

Contents

0	Overview of the supplement document	1
1	Brief description of causal mediation analysis	1
1.1	Decomposition of total effect	1
1.2	Identification of natural effects	2
2	Implementation of the regression-based causal mediation analysis method	6
2.1	Parametrizing the mediation effect formulas	6
2.2	Obtaining standard errors via multivariate delta method	6
2.3	Linear mediator model, linear outcome model	7
2.3.1	Effect formulas	7
2.3.2	Variance formulas	8
2.4	Linear mediator model, non-linear outcome model	10
2.4.1	Effect formulas	10
2.4.2	Variance formulas	10
2.5	Logistic mediator model, linear outcome model	13
2.5.1	Effect formulas	13
2.5.2	Variance formulas	13
2.6	Logistic mediator model, non-linear outcome model	17
2.6.1	Effect formulas	17
2.6.2	Variance formulas	18
3	Software user guide and reproducible code	21
3.1	Pointers to <code>regmedint</code> online user guide	21
3.2	Generating the example data	22
3.3	Reproducing the data analysis example	23
4	Demonstration of covariate-dependence of natural effect estimates	25
4.1	Generating demonstration data	25
4.2	Demonstrating covariate dependence visually	27
4.3	Implications for marginal interpretation	29
4.3.1	Linear mediator model, linear outcome model	29
4.3.2	Linear mediator model, non-linear outcome model	29
4.3.3	Logistic mediator model, linear outcome model	30
4.3.4	Logistic mediator model, non-linear outcome model	31
4.4	Practical implications for reporting	33
4.4.1	Linear mediator model, linear outcome model	33
4.4.2	Linear mediator model, non-linear outcome model	33
4.4.3	Logistic mediator model, linear outcome model	34
4.4.4	Logistic mediator model, non-linear outcome model	34
5	Bibliography	34

0 Overview of the supplement document

This supplement document provides additional details for the causal mediation analysis method implemented in the R package `regmedint` (<https://kaz-yos.github.io/regmedint/>). The sections cover the following topics.

1. Definitions of the causal quantities interest and assumptions required for their identification.
2. Outline of the estimation method implemented in `regmedint` and mathematical formulas.
3. Links to software guide for `regmedint` and code to reproduce the example.
4. Visual demonstration of covariate-dependence of effect estimates.

Applied users of the `regmedint` package should benefit most from the sotware instructions in Section 3 and the causal assumptions stated in Section 1. Section 2 is meant for methodologists and developers looking for the implementation details of the `regmedint` package.

Section 4 visually demonstrate a somewhat overlooked aspect of the regression-based causal mediation analysis using an added functionality of the `regmedint` package (re-evaluation of effect estimates at various covariate levels without model refitting). Practical implications for reporting are discussed.

The latest version of this supplement document as well as the corresponding `org-mode` and `LATEX` documents can be found at: <https://github.com/kaz-yos/regmedint-supplement>.

1 Brief description of causal mediation analysis

The literature on causal mediation analysis is vast [VanderWeele, 2015] and is evolving, thus, only the pieces relevant for the current software are reviewed here.

1.1 Decomposition of total effect

Let Y be the outcome variable of interest, A be the treatment variable of interest, M be the mediator variable of interest, and \mathbf{C} be the potentially vector-valued pre-treatment baseline covariates necessary for exchangeability (See the section on identification). The treatment contrast of interest is a_1 vs a_0 , the second being the reference level. The counterfactual $Y_{a,m}$ is the value of Y for an individual when, possibly contrary to the fact, the treatment level a and mediator level m are assigned.

The total effect (TE), the causal effect of chaning the treatment level from the reference level a_0 to the level of interest a_1 , is defined as follows at the covariate level $\mathbf{C} = \mathbf{c}$.

$$TE = E[Y_{a_1} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0} | \mathbf{C} = \mathbf{c}]$$

The conditional direct effect (CDE(m)), the effect of chaning the treatment level from the reference level a_0 to the level of interest a_1 while fixing the mediator at m is the following at the covariate level $\mathbf{C} = \mathbf{c}$.

$$CDE(m) = E[Y_{a_1,m} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0,m} | \mathbf{C} = \mathbf{c}]$$

The TE can be decomposed into the direct (non-mediated) effect and indirect (mediated) effect in two different ways [Robins and Greenland, 1992, VanderWeele, 2013].

The decomposition of TE into the pure natural direct effect (PNDE) and the total natural indirect effect (TNIE) is the usual decomposition [Pearl, 2001]. Note that the treatment value indexing the mediator M is fixed at a_0 in the PNDE, whereas the treatment value indexing the outcome Y is fixed at a_1 in the TNIE. These are emphasized with \sim throughout the document.

$$PNDE = E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]$$

$$TNIE = E[Y_{\widetilde{a_1}, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{\widetilde{a_1}, M_{a_0}} | \mathbf{C} = \mathbf{c}]$$

The other decomposition of TE is into the pure natural indirect effect (PNIE) and the total natural direct effect (TNDE) [Robins and Greenland, 1992]. Note that the treatment value indexing the mediator M is fixed at a_1 in the TNDE, whereas the treatment value indexing the outcome Y is fixed at a_0 in the PNIE. That is, these flipped in this decomposition. These are emphasized with \sim throughout the document.

$$TNDE = E[Y_{a_1, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]$$

$$PNIE = E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}]$$

More intuitively, these decomposition differs in the cross-world counterfactual state that is used as the partition. The usual PNDE+TNIE decomposition uses $Y_{a_1, M_{a_0}}$ (treatment indexing on the outcome goes up to a_1 from a_0 first), whereas the PNIE+TNDE decomposition uses $Y_{a_0, M_{a_1}}$ (treatment indexing the mediator goes up to a_1 from a_0 first).

In either case, the effect that has the reference counterfactual outcome $Y_{a_0} = Y_{a_0, M_{a_0}}$ is the *pure* natural direct/indirect effect and the effect that has $Y_{a_1} = Y_{a_1, M_{a_1}}$ is the *total* natural direct/indirect effect. See [VanderWeele, 2013] for the meaning of these two decompositions in terms of causal interaction.

More in general, we can consider the effects on the outcome model link function (g_Y) scale as follows [Starkopf et al., 2017].

$$TE = g_Y(E[Y_{a_1} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_0} | \mathbf{C} = \mathbf{c}])$$

$$CDE(m) = g_Y(E[Y_{a_1, m} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_0, m} | \mathbf{C} = \mathbf{c}])$$

$$PNDE = g_Y(E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$TNIE = g_Y(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$TNDE = g_Y(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}])$$

$$PNIE = g_Y(E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - g_Y(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

1.2 Identification of natural effects

Several conditional exchangeabilities must be assumed for identification of effects in the causal mediation framework. See [VanderWeele, 2015] (p463) for details.

$$\begin{aligned} & \text{A1} \\ & Y_{a, m} \perp\!\!\!\perp A | \mathbf{C} \\ & \text{A2} \\ & Y_{a, m} \perp\!\!\!\perp M | \{A, \mathbf{C}\} \\ & \text{A3} \\ & M_a \perp\!\!\!\perp A | \mathbf{C} \\ & \text{A4} \\ & Y_{a, m} \perp\!\!\!\perp M_{a^*} | \mathbf{C} \end{aligned}$$

The controlled direct effect (CDE) is identified with only Assumptions A1 and A2. The natural effects require all four assumptions. Intuitively, the identification of CDE involves handling the treatment and the mediator as a sequence of exposures, whose causal effects on the outcome must be identified. Thus, the conditional

exchangeabilities for the treatment as an exposure (A1) and the mediator as an exposure (A2) are required.

Additionally, the identification of the natural effects require identifying the causal effect of the treatment on the mediator acting as an "outcome" (thus, Assumption A3). The identification of the partitioning counterfactual state mentioned above, in which the treatment value indexing the outcome Y and the treatment value indexing the mediator M differ requires Assumption A4.

Given these four assumptions, the mean counterfactual with different treatment values indexing the outcome Y and the mediator M can be identified as follows. See [VanderWeele, 2015] (p465) for the proof. For a continuous M , the summation is replaced with an integration.

By iterative expectation

$$E[Y_{a,M_{a^*}} | \mathbf{C} = \mathbf{c}] = E[E[Y_{a,M_{a^*}} | M_{a^*}, \mathbf{C} = \mathbf{c}] | \mathbf{C} = \mathbf{c}]$$

Rewrite outer expectation as Riemann-Stieltjes integral

$$= \int_m E[Y_{a,M_{a^*}} | M_{a^*} = m, \mathbf{C} = \mathbf{c}] dF(M_{a^*} = m | \mathbf{C} = \mathbf{c})$$

$$\because M_{a^*} = m$$

$$= \int_m E[Y_{a,m} | M_{a^*} = m, \mathbf{C} = \mathbf{c}] dF(M_{a^*} = m | \mathbf{C} = \mathbf{c})$$

$$\because \text{A4. } Y_{a,m} \perp\!\!\!\perp M_{a^*} | \mathbf{C}$$

$$= \int_m E[Y_{a,m} | \mathbf{C} = \mathbf{c}] dF(M_{a^*} = m | \mathbf{C} = \mathbf{c})$$

$$\because \text{A3. } M_{a^*} \perp\!\!\!\perp A | \mathbf{C}$$

$$= \int_m E[Y_{a,m} | \mathbf{C} = \mathbf{c}] dF(M_{a^*} = m | A, \mathbf{C} = \mathbf{c})$$

$$\because \text{Any given } A \text{ is equivalent, use stratum } a^*$$

$$= \int_m E[Y_{a,m} | \mathbf{C} = \mathbf{c}] dF(M_{a^*} = m | A = a^*, \mathbf{C} = \mathbf{c})$$

By causal consistency for M

$$= \int_m E[Y_{a,m} | \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

$$\because \text{A1. } Y_{a,m} \perp\!\!\!\perp A | \mathbf{C}$$

$$= \int_m E[Y_{a,m} | A, \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

$$\because \text{Any given } A \text{ is equivalent, use stratum } a$$

$$= \int_m E[Y_{a,m} | A = a, \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

$$\because \text{A3. } Y_{a,m} \perp\!\!\!\perp M | \{A, \mathbf{C}\}$$

$$= \int_m E[Y_{a,m} | A = a, M, \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

$$\because \text{Any given } M \text{ is equivalent, use stratum } m$$

$$= \int_m E[Y_{a,m} | A = a, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

By causal consistency for Y

$$= \int_m E[Y | A = a, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = a^*, \mathbf{C} = \mathbf{c})$$

Written in terms of expectation

$$= E[E[Y | A = a, M, \mathbf{C} = \mathbf{c}] | A = a^*, \mathbf{C} = \mathbf{c}]$$

We can observe that the first treatment value a in the counterfactual $Y_{a, M_{a^*}}$ indexes the outcome model $E[Y|A = \underline{a}, M = m, \mathbf{C} = \mathbf{c}]$, whereas the second treatment value a^* in the counterfactual $Y_{a, \underline{M}_{a^*}}$ indexes the mediator model $P(M = m|A = \underline{a^*}, \mathbf{C} = \mathbf{c})$.

Thus, this expression takes the expectation of the inner expectation, Y given (a, m, \mathbf{c}) (seen as a function of m only), using the distribution of M given (a^*, \mathbf{c}) .

The identification formulas for the two natural *direct* effects are the following. Note the change in the treatment value indexing the *mediator model* from a_0 in the PNDE (usual NDE) to a_1 in the TNDE (annotated with \sim). Within each effect, only the treatment values indexing the *outcome model* vary (ones not annotated) because each NDE represent the direct effect of the treatment when the mediator is fixed at the natural value it would take under one treatment value (annotated with \sim).

$$\begin{aligned}
 PNDE &= E[Y_{a_1, \underline{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \underline{M}_{a_0}} | \mathbf{C} = \mathbf{c}] \\
 &= E \left[E[Y|A = a_1, M, \mathbf{C} = \mathbf{c}] | A = \underline{a_0}, \mathbf{C} = \mathbf{c} \right] \\
 &\quad - E \left[E[Y|A = a_0, M, \mathbf{C} = \mathbf{c}] | A = \underline{a_0}, \mathbf{C} = \mathbf{c} \right] \\
 &= \int_m E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = \underline{a_0}, \mathbf{C} = \mathbf{c}) \\
 &\quad - \int_m E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = \underline{a_0}, \mathbf{C} = \mathbf{c}) \\
 &= \int_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\
 &\quad \times dF(M = m | A = \underline{a_0}, \mathbf{C} = \mathbf{c}) \\
 &= \begin{cases} \text{For continuous } M \\ \int_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\ \quad \times f(M = m | A = \underline{a_0}, \mathbf{C} = \mathbf{c}) dm \\ \\ \text{For discrete } M \\ \sum_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\ \quad \times P(M = m | A = \underline{a_0}, \mathbf{C} = \mathbf{c}) \end{cases}
 \end{aligned}$$

$$\begin{aligned}
 TNDE &= E[Y_{a_1, \underline{M}_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \underline{M}_{a_1}} | \mathbf{C} = \mathbf{c}] \\
 &= E \left[E[Y|A = a_1, M, \mathbf{C} = \mathbf{c}] | A = \underline{a_1}, \mathbf{C} = \mathbf{c} \right] \\
 &\quad - E \left[E[Y|A = a_0, M, \mathbf{C} = \mathbf{c}] | A = \underline{a_1}, \mathbf{C} = \mathbf{c} \right] \\
 &= \int_m E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = \underline{a_1}, \mathbf{C} = \mathbf{c}) \\
 &\quad - \int_m E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}] dF(M = m | A = \underline{a_1}, \mathbf{C} = \mathbf{c}) \\
 &= \int_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\
 &\quad \times dF(M = m | A = \underline{a_1}, \mathbf{C} = \mathbf{c})
 \end{aligned}$$

$$= \begin{cases} \text{For continuous } M \\ \int_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\ \quad \times f(M = m|A = \underline{a}_1, \mathbf{C} = \mathbf{c}) dm \\ \\ \text{For discrete } M \\ \sum_m \{E[Y|A = a_1, M = m, \mathbf{C} = \mathbf{c}] - E[Y|A = a_0, M = m, \mathbf{C} = \mathbf{c}]\} \\ \quad \times P(M = m|A = \underline{a}_1, \mathbf{C} = \mathbf{c}) \end{cases}$$

The identification formulas for the two natural *indirect* effects are the following. Note the change in the treatment value indexing the *outcome model* from a_1 in the TNIE (usual NIE) to a_0 in the PNIE (annotated with $\underline{\quad}$). Within each effect, only the treatment values indexing the *mediator model* vary (ones not annotated) because each NIE represent the indirect effect of the treatment when its effect on the mediator is "turned on", while the treatment value representing the direct path is fixed at the natural value it would take under one treatment value (annotated with $\underline{\quad}$).

$$\begin{aligned} TNIE &= E[Y_{\underline{a}_1, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{\underline{a}_1, M_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= E \left[E[Y|A = \underline{a}_1, M, \mathbf{C} = \mathbf{c}] | A = a_1, \mathbf{C} = \mathbf{c} \right] \\ &\quad - E \left[E[Y|A = \underline{a}_1, M, \mathbf{C} = \mathbf{c}] | A = a_0, \mathbf{C} = \mathbf{c} \right] \\ &= \int_m E[Y|A = \underline{a}_1, M = m, \mathbf{C} = \mathbf{c}] dF(M = m|A = a_1, \mathbf{C} = \mathbf{c}) \\ &\quad - \int_m E[Y|A = \underline{a}_1, M = m, \mathbf{C} = \mathbf{c}] dF(M = m|A = a_0, \mathbf{C} = \mathbf{c}) \\ &= \int_m E[Y|A = \underline{a}_1, M = m, \mathbf{C} = \mathbf{c}] \\ &\quad \times \{dF(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - dF(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} \\ &= \begin{cases} \text{For continuous } M \\ \int_m E[Y|A = \underline{a}_1, M = m, \mathbf{C} = \mathbf{c}] \\ \quad \times \{f(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - f(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} dm \\ \\ \text{For discrete } M \\ \sum_m E[Y|A = \underline{a}_1, M = m, \mathbf{C} = \mathbf{c}] \\ \quad \times \{P(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - P(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} \end{cases} \end{aligned}$$

$$\begin{aligned} PNIE &= E[Y_{\underline{a}_0, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{\underline{a}_0, M_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= E \left[E[Y|A = \underline{a}_0, M, \mathbf{C} = \mathbf{c}] | A = a_1, \mathbf{C} = \mathbf{c} \right] \\ &\quad - E \left[E[Y|A = \underline{a}_0, M, \mathbf{C} = \mathbf{c}] | A = a_0, \mathbf{C} = \mathbf{c} \right] \\ &= \int_m E[Y|A = \underline{a}_0, M = m, \mathbf{C} = \mathbf{c}] dF(M = m|A = a_1, \mathbf{C} = \mathbf{c}) \end{aligned}$$

$$\begin{aligned}
& - \int_m E[Y|A = \underline{a}_0, M = m, \mathbf{C} = \mathbf{c}] dF(M = m|A = a_0, \mathbf{C} = \mathbf{c}) \\
& = \int_m E[Y|A = \underline{a}_0, M = m, \mathbf{C} = \mathbf{c}] \\
& \quad \times \{dF(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - dF(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} \\
& = \begin{cases} \text{For continuous } M \\ \int_m E[Y|A = \underline{a}_0, M = m, \mathbf{C} = \mathbf{c}] \\ \quad \times \{f(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - f(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} dm \\ \\ \text{For discrete } M \\ \sum_m E[Y|A = \underline{a}_0, M = m, \mathbf{C} = \mathbf{c}] \\ \quad \times \{P(M = m|A = a_1, \mathbf{C} = \mathbf{c}) - P(M = m|A = a_0, \mathbf{C} = \mathbf{c})\} \end{cases}
\end{aligned}$$

2 Implementation of the regression-based causal mediation analysis method

Here we describe the formulas implemented in `regmedint`, using the notational convention in [VanderWeele, 2015, Valeri and Vanderweele, 2013, Valeri and VanderWeele, 2015]. In addition to what can be found in the appendix of [VanderWeele, 2015], we provide the explicit expressions for the alternative TNDE-PNIE decomposition. The differences are highlighted with $\underline{\sim}$.

2.1 Parametrizing the mediation effect formulas

As seen above, there are two models involved in identification of natural effects: the outcome model ($E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}]$) and the mediator model ($p(M|A = a, \mathbf{C} = \mathbf{c})$). The identification formulas do not specify any particular model structure (non-parametric). In the method described in [Valeri and Vanderweele, 2013, Valeri and VanderWeele, 2015], a simple parametric model is proposed for each.

The mediator model with a link function g_M is parametrized as follows.

$$g_M(E[M|A = a, \mathbf{C} = \mathbf{c}]) = \beta_0 + \beta_1 a + \beta_2^T \mathbf{c}$$

The outcome model with a link function g_Y is parametrized as follows.

$$g_Y(E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}]) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4^T \mathbf{c}$$

Under these parametric modeling assumptions, each effect of interest can be written as a function of the parameters (coefficients) of the mediator model (β and sometimes σ^2) and the outcome model (θ). Because of the product configuration (outcome model \times mediator model), each natural effect is a non-linear function of the parameters (model coefficients). The maximum likelihood estimates (MLE) of these effects are the ones with these parameters replaced with their respective MLEs from the two models.

2.2 Obtaining standard errors via multivariate delta method

Each effect of interest is estimated as a scalar-valued, non-linear function of estimated coefficients for the mediator model and the outcome model. Thus, we can obtain the standard error of each effect estimate using the variance covariance matrix for the coefficients and multivariate delta method [Hoef, 2012].

Let the scalar quantity of interest be Q , a function of parameter vector $(\beta^T, \theta^T)^T$. Then, its gradient (vector of partial derivatives) with respect to the parameter vector $(\beta^T, \theta^T)^T$ is the following.

$$\nabla Q = \frac{\partial Q}{\partial (\beta^T, \theta^T)^T}$$

$$= \begin{bmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \\ \frac{\partial Q}{\partial \beta_2} \\ \frac{\partial Q}{\partial \theta_0} \\ \frac{\partial Q}{\partial \theta_1} \\ \frac{\partial Q}{\partial \theta_2} \\ \frac{\partial Q}{\partial \theta_3} \\ \frac{\partial Q}{\partial \theta_4} \end{bmatrix}$$

In the case of a linear mediator model and a non-linear outcome model, there is an additional element $\frac{\partial Q}{\partial \sigma^2}$ at the bottom of the gradient vector.

By the large sample approximation using the multivariate delta method, the variance of the quantity of interest evaluated at the MLEs $(\hat{\beta}^T, \hat{\theta}^T)^T$ is the following.

$$\underbrace{Var \left[Q \left\{ (\hat{\beta}^T, \hat{\theta}^T)^T \right\} \right]}_{\text{scalar}} \approx \underbrace{\left[\nabla Q \left((\hat{\beta}^T, \hat{\theta}^T)^T \right) \right]^T}_{\text{row vector}} \underbrace{Var((\hat{\beta}^T, \hat{\theta}^T)^T)}_{\text{matrix}} \underbrace{\left[\nabla Q \left((\hat{\beta}^T, \hat{\theta}^T)^T \right) \right]}_{\text{column vector}}$$

This expression is abbreviated as $\Gamma \Sigma \Gamma'$ in [VanderWeele, 2015, Valeri and Vanderweele, 2013, Valeri and VanderWeele, 2015]. In these references, the treatment contrast $(a_1 - a_0)$ is factored out from $\nabla Q((\hat{\beta}^T, \hat{\theta}^T)^T)$ when possible. In the following, we define Γ as a column vector to be consistent with the implementation of `regmedint`, thus, the corresponding expression appears as $\Gamma^T \Sigma \Gamma$.

2.3 Linear mediator model, linear outcome model

2.3.1 Effect formulas

The function `calc_myreg_mreg_linear_yreg_linear_est()` implements the effect formulas in [VanderWeele, 2015] (p466).

Models

$$\begin{aligned} E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta}_4^T \mathbf{c} \\ E[M|A = a, \mathbf{C} = \mathbf{c}] &= \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^T \mathbf{c} \end{aligned}$$

Effects

$$\begin{aligned} CDE(m) &= E[Y_{a_1, m} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, m} | \mathbf{C} = \mathbf{c}] \\ &= (\theta_1 + \theta_3 m)(a_1 - a_0) \end{aligned}$$

$$\begin{aligned} PNDE &= E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c}) \right\} (a_1 - a_0) \\ TNIE &= E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= \beta_1(\theta_2 + \theta_3 a_1)(a_1 - a_0) \end{aligned}$$

$$TNDE = E[Y_{a_1, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]$$

$$\begin{aligned}
&= \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c}) \right\} (a_1 - a_0) \\
PNIE &= E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}] \\
&= \beta_1(\theta_2 + \theta_3 a_0)(a_1 - a_0)
\end{aligned}$$

$$TE = PNDE + TNIE$$

$$PM = \frac{TNIE}{PNDE + TNIE}$$

2.3.2 Variance formulas

The function `calc_myreg_mreg_linear_yreg_linear_se()` implements the standard error formulas in [VanderWeele, 2015] (p466).

$$\begin{aligned}
(a_1 - a_0)\Gamma_{CDE(m)} &= \frac{\partial CDE(m)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= (a_1 - a_0) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \end{bmatrix} \\
(a_1 - a_0)\Gamma_{PNDE} &= \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= (a_1 - a_0) \begin{bmatrix} \theta_3 \\ \theta_3 a_0 \\ \theta_3 \mathbf{c} \\ 0 \\ 1 \\ 0 \\ \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c} \\ 0 \end{bmatrix} \\
(a_1 - a_0)\Gamma_{TNIE} &= \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= (a_1 - a_0) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a_1 \\ 0 \\ 0 \\ 0 \\ \beta_1 \\ \beta_1 a_1 \\ 0 \end{bmatrix} \\
(a_1 - a_0)\Gamma_{TNDE} &= \frac{\partial TNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}
\end{aligned}$$

$$= (a_1 - a_0) \begin{bmatrix} \theta_3 \\ \theta_3 a_{1\sim} \\ \theta_3 \mathbf{c} \\ 0 \\ 1 \\ 0 \\ \beta_0 + \beta_1 a_{1\sim} + \beta_2^T \mathbf{c} \\ 0 \end{bmatrix}$$

$$(a_1 - a_0) \Gamma_{PNIE} = \frac{\partial PNIE}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a_{0\sim} \\ 0 \\ 0 \\ 0 \\ \beta_1 \\ \beta_1 a_{0\sim} \\ 0 \end{bmatrix}$$

$$\begin{aligned} (a_1 - a_0) \Gamma_{TE} &= \frac{\partial TE}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial (PNDE + TNIE)}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= (a_1 - a_0) (\Gamma_{PNDE} + \Gamma_{TNIE}) \end{aligned}$$

$$\begin{aligned} (a_1 - a_0) \Gamma_{PM} &= \frac{\partial PM}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &\text{By multivariate chain rule} \\ &= \frac{\partial PM}{\partial PNDE} \frac{\partial PNDE}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE} \frac{\partial TNIE}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial PM}{\partial PNDE} (a_1 - a_0) \Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE} (a_1 - a_0) \Gamma_{TNIE} \\ &= \frac{-TNIE}{(PNDE + TNIE)^2} (a_1 - a_0) \Gamma_{PNDE} + \frac{PNDE}{(PNDE + TNIE)^2} (a_1 - a_0) \Gamma_{TNIE} \\ &= (a_1 - a_0) \frac{-TNIE \Gamma_{PNDE} + PNDE \Gamma_{TNIE}}{(PNDE + TNIE)^2} \end{aligned}$$

Variance-covariance matrix from two models

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix}$$

$$\begin{aligned} SE(\widehat{CDE}(m)) &= \sqrt{\Gamma_{CDE(m)}^T \boldsymbol{\Sigma} \Gamma_{CDE(m)}} |a_1 - a_0| \\ SE(\widehat{PNDE}) &= \sqrt{\Gamma_{PNDE}^T \boldsymbol{\Sigma} \Gamma_{PNDE}} |a_1 - a_0| \\ SE(\widehat{TNIE}) &= \sqrt{\Gamma_{TNIE}^T \boldsymbol{\Sigma} \Gamma_{TNIE}} |a_1 - a_0| \\ SE(\widehat{TNDE}) &= \sqrt{\Gamma_{TNDE}^T \boldsymbol{\Sigma} \Gamma_{TNDE}} |a_1 - a_0| \\ SE(\widehat{PNIE}) &= \sqrt{\Gamma_{PNIE}^T \boldsymbol{\Sigma} \Gamma_{PNIE}} |a_1 - a_0| \end{aligned}$$

$$SE(\widehat{TE}) = \sqrt{\Gamma_{TE}^T \boldsymbol{\Sigma} \Gamma_{TE}} |a_1 - a_0|$$

$$SE(\widehat{PM}) = \sqrt{\Gamma_{PM}^T \boldsymbol{\Sigma} \Gamma_{PM}} |a_1 - a_0|$$

2.4 Linear mediator model, non-linear outcome model

These formulas are used for all non-linear outcome models, including logistic (rare outcome assumption), log-linear, Poisson, negative binomial [Valeri and Vanderweele, 2013], accelerated failure time, and Cox (rare outcome assumption) [Valeri and VanderWeele, 2015].

2.4.1 Effect formulas

The function `calc_myreg_mreg_linear_yreg_logistic_est()` implements the effect formulas in [VanderWeele, 2015] (p468).

Models

$$\text{logit}(E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}]) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^T \mathbf{c}$$

$$E[M|A = a, \mathbf{C} = \mathbf{c}] = \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^T \mathbf{c}$$

Effects on outcome model link function scale

$$CDE(m) = \text{logit}(E[Y_{a_1, m} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, m} | \mathbf{C} = \mathbf{c}])$$

$$= (\theta_1 + \theta_3 m)(a_1 - a_0)$$

$$PNDE = \text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$\approx \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c} + \theta_2 \sigma^2) \right\} (a_1 - a_0) + \frac{1}{2} \theta_3^2 \sigma^2 (a_1^2 - a_0^2)$$

$$TNIE = \text{logit}(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$\approx \beta_1(\theta_2 + \theta_3 a_1)(a_1 - a_0)$$

$$TNDE = \text{logit}(E[Y_{a_1, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}])$$

$$\approx \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c} + \theta_2 \sigma^2) \right\} (a_1 - a_0) + \frac{1}{2} \theta_3^2 \sigma^2 (a_1^2 - a_0^2)$$

$$PNIE = \text{logit}(E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$\approx \beta_1(\theta_2 + \theta_3 a_0)(a_1 - a_0)$$

$$TE = PNDE + TNIE$$

$$PM = \frac{\exp(PNDE)(\exp(TNIE) - 1)}{\exp(PNDE)\exp(TNIE) - 1}$$

2.4.2 Variance formulas

The function `calc_myreg_mreg_linear_yreg_logistic_se()` implements the standard error formulas in [VanderWeele, 2015] (p468).

$$(a_1 - a_0) \Gamma_{CDE(m)} = \frac{\partial CDE(m)}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \\ 0 \end{bmatrix}$$

$$(a_1 - a_0)\Gamma_{PNDE} = \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} \theta_3 \\ \theta_3 a_0 \\ \theta_3 \mathbf{c} \\ 0 \\ 1 \\ \theta_3 \sigma^2 \\ \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c} + \theta_2 \sigma^2 + \theta_3 \sigma^2 (a_0 + a_1) \\ 0 \\ \theta_3 \theta_2 + \frac{1}{2} \theta_3^2 (a_1 + a_0) \end{bmatrix}$$

$$(a_1 - a_0)\Gamma_{TNIE} = \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a_1 \\ 0 \\ 0 \\ 0 \\ \beta_1 \\ \beta_1 a_1 \\ 0 \\ 0 \end{bmatrix}$$

$$(a_1 - a_0)\Gamma_{TNDE} = \frac{\partial TNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} \theta_3 \\ \theta_3 a_1 \\ \theta_3 \mathbf{c} \\ 0 \\ 1 \\ \theta_3 \sigma^2 \\ \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c} + \theta_2 \sigma^2 + \theta_3 \sigma^2 (a_0 + a_1) \\ 0 \\ \theta_3 \theta_2 + \frac{1}{2} \theta_3^2 (a_1 + a_0) \end{bmatrix}$$

$$(a_1 - a_0)\Gamma_{PNIE} = \frac{\partial PNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a_0 \\ 0 \\ 0 \\ 0 \\ \beta_1 \\ \beta_1 a_0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} (a_1 - a_0)\Gamma_{TE} &= \frac{\partial TE}{\partial(\beta^T, \theta^T, \sigma^2)^T} \\ &= \frac{\partial(PNDE + TNIE)}{\partial(\beta^T, \theta^T, \sigma^2)^T} \\ &= (a_1 - a_0)(\Gamma_{PNDE} + \Gamma_{TNIE}) \end{aligned}$$

$$\begin{aligned} (a_1 - a_0)\Gamma_{PM} &= \frac{\partial PM}{\partial(\beta^T, \theta^T, \sigma^2)^T} \\ &\text{By multivariate chain rule} \\ &= \frac{\partial PM}{\partial PNDE} \frac{\partial PNDE}{\partial(\beta^T, \theta^T, \sigma^2)^T} + \frac{\partial PM}{\partial TNIE} \frac{\partial TNIE}{\partial(\beta^T, \theta^T, \sigma^2)^T} \\ &= \frac{\partial PM}{\partial PNDE} (a_1 - a_0)\Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE} (a_1 - a_0)\Gamma_{TNIE} \\ &= -\frac{\exp(PNDE) \{\exp(TNIE) - 1\}}{\{\exp(PNDE) \exp(TNIE) - 1\}^2} (a_1 - a_0)\Gamma_{PNDE} \\ &\quad + \frac{\exp(PNDE) \exp(TNIE) \{\exp(PNDE) - 1\}}{\{\exp(PNDE) \exp(TNIE) - 1\}^2} (a_1 - a_0)\Gamma_{TNIE} \end{aligned}$$

Variance-covariance matrix from two models

$$\Sigma = \begin{bmatrix} \Sigma_\beta & 0 & 0 \\ 0 & \Sigma_\theta & 0 \\ 0 & 0 & \Sigma_{\sigma^2} \end{bmatrix}$$

$$\Sigma_{\sigma^2} = \frac{2(\sigma^2)^2}{n-p} \text{ where } p = \text{length}(\beta)$$

$$SE(\widehat{CDE}(m)) = \sqrt{\Gamma_{CDE(m)}^T \Sigma \Gamma_{CDE(m)}} |a_1 - a_0|$$

$$SE(\widehat{PNDE}) = \sqrt{\Gamma_{PNDE}^T \Sigma \Gamma_{PNDE}} |a_1 - a_0|$$

$$SE(\widehat{TNIE}) = \sqrt{\Gamma_{TNIE}^T \Sigma \Gamma_{TNIE}} |a_1 - a_0|$$

$$SE(\widehat{TNDE}) = \sqrt{\Gamma_{TNDE}^T \Sigma \Gamma_{TNDE}} |a_1 - a_0|$$

$$SE(\widehat{PNIE}) = \sqrt{\Gamma_{PNIE}^T \Sigma \Gamma_{PNIE}} |a_1 - a_0|$$

$$SE(\widehat{TE}) = \sqrt{\Gamma_{TE}^T \Sigma \Gamma_{TE}} |a_1 - a_0|$$

$$SE(\widehat{PM}) = \sqrt{\Gamma_{PM}^T \Sigma \Gamma_{PM}} |a_1 - a_0|$$

2.5 Logistic mediator model, linear outcome model

2.5.1 Effect formulas

The function `calc_myreg_mreg_logistic_yreg_linear_est()` implements the effect formulas in [VanderWeele, 2015] (p471).

Models

$$E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^T \mathbf{c}$$

$$\text{logit}(E[M|A = a, \mathbf{C} = \mathbf{c}]) = \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^T \mathbf{c}$$

Effects

$$CDE(m) = E[Y_{a_1, m} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, m} | \mathbf{C} = \mathbf{c}]$$

$$= (\theta_1 + \theta_3 m)(a_1 - a_0)$$

$$PNDE = E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]$$

$$= \{\theta_1(a_1 - a_0)\} + \{\theta_3(a_1 - a_0)\} \frac{\exp(\beta_0 + \beta_1 \widetilde{a_0} + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 \widetilde{a_0} + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$TNIE = E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}]$$

$$= (\theta_2 + \theta_3 \widetilde{a_1}) \left\{ \frac{\exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})} - \frac{\exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})} \right\}$$

$$TNDE = E[Y_{a_1, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]$$

$$= \{\theta_1(a_1 - a_0)\} + \{\theta_3(a_1 - a_0)\} \frac{\exp(\beta_0 + \beta_1 \widetilde{a_1} + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 \widetilde{a_1} + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$PNIE = E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}]$$

$$= (\theta_2 + \theta_3 \widetilde{a_0}) \left\{ \frac{\exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})} - \frac{\exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})} \right\}$$

$$TE = PNDE + TNIE$$

$$PM = \frac{TNIE}{PNDE + TNIE}$$

2.5.2 Variance formulas

The function `calc_myreg_mreg_logistic_yreg_linear_se()` implements the standard error formulas in [VanderWeele, 2015] (p471).

$$(a_1 - a_0) \Gamma_{CDE(m)} = \frac{\partial CDE(m)}{\partial (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \end{bmatrix}$$

$$\begin{aligned}
d_{1,PNDE} &= \theta_3 \frac{\exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})\}^2} \\
d_{2,PNDE} &= \tilde{a}_0 \theta_3 \frac{\exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})\}^2} \\
d_{3,PNDE} &= \mathbf{c} \theta_3 \frac{\exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})\}^2} \\
d_{4,PNDE} &= 0 \\
d_{5,PNDE} &= 1 \\
d_{6,PNDE} &= 0 \\
d_{7,PNDE} &= \frac{\exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 \tilde{a}_0 + \beta_2^T \mathbf{c})} \\
d_{8,PNDE} &= 0 \\
(a_1 - a_0) \Gamma_{PNDE} &= \frac{\partial PNDE}{\partial (\beta^T, \theta^T)^T} \\
&= (a_1 - a_0) \begin{bmatrix} d_{1,PNDE} \\ d_{2,PNDE} \\ d_{3,PNDE} \\ d_{4,PNDE} \\ d_{5,PNDE} \\ d_{6,PNDE} \\ d_{7,PNDE} \\ d_{8,PNDE} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
Q_{TNIE} &= \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})\}^2} \\
B_{TNIE} &= \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})\}^2} \\
K_{TNIE} &= \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})} \\
D_{TNIE} &= \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})} \\
d_{1,TNIE} &= (\theta_2 + \theta_3 \tilde{a}_1)(Q_{TNIE} - B_{TNIE}) \\
d_{2,TNIE} &= (\theta_2 + \theta_3 \tilde{a}_1)(a_1 Q_{TNIE} - a_0 B_{TNIE}) \\
d_{3,TNIE} &= (\theta_2 + \theta_3 \tilde{a}_1) \mathbf{c} (Q_{TNIE} - B_{TNIE}) \\
d_{4,TNIE} &= 0 \\
d_{5,TNIE} &= 0 \\
d_{6,TNIE} &= K_{TNIE} - D_{TNIE} \\
d_{7,TNIE} &= \tilde{a}_1 (K_{TNIE} - D_{TNIE}) \\
d_{8,TNIE} &= 0
\end{aligned}$$

Note the lack of the common factor $(a_1 - a_0)$

$$\Gamma_{TNIE} = \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= \begin{bmatrix} d_{1,TNIE} \\ d_{2,TNIE} \\ d_{3,TNIE} \\ d_{4,TNIE} \\ d_{5,TNIE} \\ d_{6,TNIE} \\ d_{7,TNIE} \\ d_{8,TNIE} \end{bmatrix}$$

$$d_{1,TNDE} = \theta_3 \frac{\exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{\{1 + \exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})\}^2}$$

$$d_{2,TNDE} = a_1 \theta_3 \frac{\exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{\{1 + \exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})\}^2}$$

$$d_{3,TNDE} = \mathbf{c} \theta_3 \frac{\exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{\{1 + \exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})\}^2}$$

$$d_{4,TNDE} = 0$$

$$d_{5,TNDE} = 1$$

$$d_{6,TNDE} = 0$$

$$d_{7,TNDE} = \frac{\exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_1 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$d_{8,TNDE} = 0$$

$$(a_1 - a_0) \Gamma_{TNDE} = \frac{\partial TNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= (a_1 - a_0) \begin{bmatrix} d_{1,TNDE} \\ d_{2,TNDE} \\ d_{3,TNDE} \\ d_{4,TNDE} \\ d_{5,TNDE} \\ d_{6,TNDE} \\ d_{7,TNDE} \\ d_{8,TNDE} \end{bmatrix}$$

$$Q_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})\}^2}$$

$$B_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{\{1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})\}^2}$$

$$K_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$D_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$d_{1,PNIE} = (\theta_2 + \theta_3 a_0)(Q_{PNIE} - B_{PNIE})$$

$$d_{2,PNIE} = (\theta_2 + \theta_3 a_0)(a_1 Q_{PNIE} - a_0 B_{PNIE})$$

$$d_{3,PNIE} = (\theta_2 + \theta_3 a_0) \mathbf{c}(Q_{PNIE} - B_{PNIE})$$

$$d_{4,PNIE} = 0$$

$$d_{5,PNIE} = 0$$

$$d_{6,PNIE} = K_{PNIE} - D_{PNIE}$$

$$d_{7,PNIE} = a_0(K_{PNIE} - D_{PNIE})$$

$$d_{8,PNIE} = 0$$

Note the lack of the common factor $(a_1 - a_0)$

$$\begin{aligned} \Gamma_{PNIE} &= \frac{\partial PNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \begin{bmatrix} d_{1,PNIE} \\ d_{2,PNIE} \\ d_{3,PNIE} \\ d_{4,PNIE} \\ d_{5,PNIE} \\ d_{6,PNIE} \\ d_{7,PNIE} \\ d_{8,PNIE} \end{bmatrix} \end{aligned}$$

Note the lack of the common factor $(a_1 - a_0)$

$$\begin{aligned} \Gamma_{TE} &= \frac{\partial TE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= (a_1 - a_0)\Gamma_{PNDE} + \Gamma_{TNIE} \end{aligned}$$

Note the lack of the common factor $(a_1 - a_0)$

$$\begin{aligned} \Gamma_{PM} &= \frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &\text{By multivariate chain rule} \\ &= \frac{\partial PM}{\partial PNDE} \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE} \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial PM}{\partial PNDE} (a_1 - a_0)\Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE} \Gamma_{TNIE} \\ &= \frac{-TNIE}{(PNDE + TNIE)^2} (a_1 - a_0)\Gamma_{PNDE} + \frac{PNDE}{(PNDE + TNIE)^2} \Gamma_{TNIE} \\ &= \frac{-TNIE (a_1 - a_0)\Gamma_{PNDE} + PNDE \Gamma_{TNIE}}{(PNDE + TNIE)^2} \end{aligned}$$

Variance-covariance matrix from two models

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix} \\ SE(\widehat{CDE}(m)) &= \sqrt{\Gamma_{CDE(m)}^T \boldsymbol{\Sigma} \Gamma_{CDE(m)}} |a_1 - a_0| \\ SE(\widehat{PNDE}) &= \sqrt{\Gamma_{PNDE}^T \boldsymbol{\Sigma} \Gamma_{PNDE}} |a_1 - a_0| \\ SE(\widehat{TNIE}) &= \sqrt{\Gamma_{TNIE}^T \boldsymbol{\Sigma} \Gamma_{TNIE}} \end{aligned}$$

$$\begin{aligned}
SE(\widehat{TNDE}) &= \sqrt{\Gamma_{TNDE}^T \boldsymbol{\Sigma} \Gamma_{TNDE}} |a_1 - a_0| \\
SE(\widehat{PNIE}) &= \sqrt{\Gamma_{PNIE}^T \boldsymbol{\Sigma} \Gamma_{PNIE}} \\
SE(\widehat{TE}) &= \sqrt{\Gamma_{TE}^T \boldsymbol{\Sigma} \Gamma_{TE}} \\
SE(\widehat{PM}) &= \sqrt{\Gamma_{PM}^T \boldsymbol{\Sigma} \Gamma_{PM}}
\end{aligned}$$

2.6 Logistic mediator model, non-linear outcome model

These formulas are used for all non-linear outcome models, including logistic (rare outcome assumption), log-linear, Poisson, negative binomial [Valeri and Vanderweele, 2013], accelerated failure time, and Cox (rare outcome assumption) [Valeri and VanderWeele, 2015].

2.6.1 Effect formulas

The function `calc_myreg_mreg_logistic_yreg_logistic_est()` implements the effect formulas in [VanderWeele, 2015] (p473).

Models

$$\begin{aligned}
\text{logit}(E[Y|A = a, M = m, \mathbf{C} = \mathbf{c}]) &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^T \mathbf{c} \\
\text{logit}(E[M|A = a, \mathbf{C} = \mathbf{c}]) &= \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^T \mathbf{c}
\end{aligned}$$

Effects on outcome model link function scale

$$\begin{aligned}
CDE(m) &= \text{logit}(E[Y_{a_1, m} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, m} | \mathbf{C} = \mathbf{c}]) \\
&= (\theta_1 + \theta_3 m)(a_1 - a_0)
\end{aligned}$$

$$\begin{aligned}
PNDE &= \text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]) \\
&\approx \theta_1(a_1 - a_0) \\
&\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
TNIE &= \text{logit}(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}]) \\
&\approx \log(1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c}))
\end{aligned}$$

Note the $a_0 \rightarrow a_1$ changes associated with β_1 .

$$\begin{aligned}
TNDE &= \text{logit}(E[Y_{a_1, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_1}} | \mathbf{C} = \mathbf{c}]) \\
&\approx \theta_1(a_1 - a_0) \\
&\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c}))
\end{aligned}$$

Note the $a_1 \rightarrow a_0$ changes associated with θ_3 .

$$PNIE = \text{logit}(E[Y_{a_0, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$\begin{aligned}
&\approx \log(1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad + \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})) \\
&\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c}))
\end{aligned}$$

$$\begin{aligned}
TE &= PNDE + TNIE \\
PM &= \frac{\exp(PNDE)(\exp(TNIE) - 1)}{\exp(PNDE)\exp(TNIE) - 1}
\end{aligned}$$

2.6.2 Variance formulas

The function `calc_myreg_mreg_logistic_yreg_logistic_se()` implements the standard error formulas in [VanderWeele, 2015] (p473).

Note the lack of the common factor $(a_1 - a_0)$ throughout.

$$\begin{aligned}
\Gamma_{CDE(m)} &= \frac{\partial CDE(m)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ (a_1 - a_0) \\ 0 \\ m(a_1 - a_0) \\ 0 \end{bmatrix}
\end{aligned}$$

$$Q_{PNDE} = \frac{\exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$B_{PNDE} = \frac{\exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$d_{1,PNDE} = Q_{PNDE} - B_{PNDE}$$

$$d_{2,PNDE} = a_0(Q_{PNDE} - B_{PNDE})$$

$$d_{3,PNDE} = \mathbf{c}(Q_{PNDE} - B_{PNDE})$$

$$d_{4,PNDE} = 0$$

$$d_{5,PNDE} = a_1 - a_0$$

$$d_{6,PNDE} = Q_{PNDE} - B_{PNDE}$$

$$d_{7,PNDE} = a_1 Q_{PNDE} - a_0 B_{PNDE}$$

$$d_{8,PNDE} = 0$$

$$\Gamma_{PNDE} = \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= \begin{bmatrix} d_{1,PNDE} \\ d_{2,PNDE} \\ d_{3,PNDE} \\ d_{4,PNDE} \\ d_{5,PNDE} \\ d_{6,PNDE} \\ d_{7,PNDE} \\ d_{8,PNDE} \end{bmatrix}$$

$$Q_{TNIE} = \frac{\exp(\theta_2 + \theta_3 a_{\sim 1} + \beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_{\sim 1} + \beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}$$

$$B_{TNIE} = \frac{\exp(\theta_2 + \theta_3 a_{\sim 1} + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_{\sim 1} + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}$$

$$K_{TNIE} = \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}$$

$$D_{TNIE} = \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}$$

$$d_{1,TNIE} = (D_{TNIE} + Q_{TNIE}) - (K_{TNIE} + B_{TNIE})$$

$$d_{2,TNIE} = a_0(D_{TNIE} - B_{TNIE}) + a_1(Q_{TNIE} - K_{TNIE})$$

$$d_{3,TNIE} = \mathbf{c} \{ (D_{TNIE} + Q_{TNIE}) - (K_{TNIE} + B_{TNIE}) \}$$

$$d_{4,TNIE} = 0$$

$$d_{5,TNIE} = 0$$

$$d_{6,TNIE} = Q_{TNIE} - B_{TNIE}$$

$$d_{7,TNIE} = a_{\sim 1}(Q_{TNIE} - B_{TNIE})$$

$$d_{8,TNIE} = 0$$

$$\Gamma_{TNIE} = \frac{\partial TNIE}{\partial (\beta^T, \theta^T)^T}$$

$$= \begin{bmatrix} d_{1,TNIE} \\ d_{2,TNIE} \\ d_{3,TNIE} \\ d_{4,TNIE} \\ d_{5,TNIE} \\ d_{6,TNIE} \\ d_{7,TNIE} \\ d_{8,TNIE} \end{bmatrix}$$

$$Q_{TNDE} = \frac{\exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_{\sim 1} + \beta_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_{\sim 1} + \beta_2^T \mathbf{c})}$$

$$B_{TNDE} = \frac{\exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_{\sim 1} + \beta_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_{\sim 1} + \beta_2^T \mathbf{c})}$$

$$d_{1,TNDE} = Q_{TNDE} - B_{TNDE}$$

$$d_{2,TNDE} = a_{\sim 1}(Q_{TNDE} - B_{TNDE})$$

$$d_{3,TNDE} = \mathbf{c}(Q_{TNDE} - B_{TNDE})$$

$$d_{4,TNDE} = 0$$

$$d_{5,TNDE} = a_1 - a_0$$

$$d_{6,TNDE} = Q_{TNDE} - B_{TNDE}$$

$$d_{7,TNDE} = a_1 Q_{TNDE} - a_0 B_{TNDE}$$

$$d_{8,TNDE} = 0$$

$$\begin{aligned} \Gamma_{TNDE} &= \frac{\partial TNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \begin{bmatrix} d_{1,TNDE} \\ d_{2,TNDE} \\ d_{3,TNDE} \\ d_{4,TNDE} \\ d_{5,TNDE} \\ d_{6,TNDE} \\ d_{7,TNDE} \\ d_{8,TNDE} \end{bmatrix} \end{aligned}$$

$$Q_{PNIE} = \frac{\exp(\theta_2 + \theta_3 \tilde{a}_0 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 \tilde{a}_0 + \beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$B_{PNIE} = \frac{\exp(\theta_2 + \theta_3 \tilde{a}_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\theta_2 + \theta_3 \tilde{a}_0 + \beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$K_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$D_{PNIE} = \frac{\exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \boldsymbol{\beta}_2^T \mathbf{c})}$$

$$d_{1,PNIE} = (D_{PNIE} + Q_{PNIE}) - (K_{PNIE} + B_{PNIE})$$

$$d_{2,PNIE} = a_0(D_{PNIE} - B_{PNIE}) + a_1(Q_{PNIE} - K_{PNIE})$$

$$d_{3,PNIE} = \mathbf{c} \{ (D_{PNIE} + Q_{PNIE}) - (K_{PNIE} + B_{PNIE}) \}$$

$$d_{4,PNIE} = 0$$

$$d_{5,PNIE} = 0$$

$$d_{6,PNIE} = Q_{PNIE} - B_{PNIE}$$

$$d_{7,PNIE} = \tilde{a}_0(Q_{PNIE} - B_{PNIE})$$

$$d_{8,PNIE} = 0$$

$$\begin{aligned} \Gamma_{PNIE} &= \frac{\partial PNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \begin{bmatrix} d_{1,PNIE} \\ d_{2,PNIE} \\ d_{3,PNIE} \\ d_{4,PNIE} \\ d_{5,PNIE} \\ d_{6,PNIE} \\ d_{7,PNIE} \\ d_{8,PNIE} \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \Gamma_{TE} &= \frac{\partial TE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \end{aligned}$$

$$= \Gamma_{PNDE} + \Gamma_{TNIE}$$

$$\begin{aligned} \Gamma_{PM} &= \frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &\text{By multivariate chain rule} \\ &= \frac{\partial PM}{\partial PNDE} \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE} \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial PM}{\partial PNDE} \Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE} \Gamma_{TNIE} \\ &= -\frac{\exp(PNDE) \{\exp(TNIE) - 1\}}{\{\exp(PNDE) \exp(TNIE) - 1\}^2} \Gamma_{PNDE} \\ &\quad + \frac{\exp(PNDE) \exp(TNIE) \{\exp(PNDE) - 1\}}{\{\exp(PNDE) \exp(TNIE) - 1\}^2} \Gamma_{TNIE} \end{aligned}$$

Variance-covariance matrix from two models

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix} \\ SE(\widehat{CDE}(m)) &= \sqrt{\Gamma_{CDE(m)}^T \boldsymbol{\Sigma} \Gamma_{CDE(m)}} \\ SE(\widehat{PNDE}) &= \sqrt{\Gamma_{PNDE}^T \boldsymbol{\Sigma} \Gamma_{PNDE}} \\ SE(\widehat{TNIE}) &= \sqrt{\Gamma_{TNIE}^T \boldsymbol{\Sigma} \Gamma_{TNIE}} \\ SE(\widehat{TNDE}) &= \sqrt{\Gamma_{TNDE}^T \boldsymbol{\Sigma} \Gamma_{TNDE}} \\ SE(\widehat{PNIE}) &= \sqrt{\Gamma_{PNIE}^T \boldsymbol{\Sigma} \Gamma_{PNIE}} \\ SE(\widehat{TE}) &= \sqrt{\Gamma_{TE}^T \boldsymbol{\Sigma} \Gamma_{TE}} \\ SE(\widehat{PM}) &= \sqrt{\Gamma_{PM}^T \boldsymbol{\Sigma} \Gamma_{PM}} \end{aligned}$$

3 Software user guide and reproducible code

3.1 Pointers to regmedint online user guide

Further details of the software `regmedint` are given on its website (<https://kaz-yos.github.io/regmedint/index.html>). In particular, the online vignettes for the package should be useful.

1. Introduction to user interface functions

- URL: https://kaz-yos.github.io/regmedint/articles/vig_01_introduction.html
- Explains all user interface functions and their outputs.

2. Implementation of formulas

- URL: https://kaz-yos.github.io/regmedint/articles/vig_02_formulas.html
- Demonstrates internal functions implementing formulas.

3. Using bootstrapping with regmedint

- URL: https://kaz-yos.github.io/regmedint/articles/vig_03_bootstrap.html
- Implements bootstrap confidence intervals with the `boot` and `modelr` packages.

4. Using multiple imputation with regmedint

- URL: https://kaz-yos.github.io/regmedint/articles/vig_04_mi.html

- Shows how to use multiple imputation via the `mice` and `mitools` packages.

In the following, we included the minimum code to reproduce the example causal mediation analysis in the paper.

3.2 Generating the example data

The demonstration data used in the paper was generated as follows in R version 4.0.0 on macOS version 10.15.5.

$$n = 1000$$

Covariates

$$C_{1,i} \stackrel{\text{iid}}{\sim} N(0, 1)$$

$$C_{2,i} \stackrel{\text{iid}}{\sim} N(0, 1)$$

Binary treatment

$$\text{logit}(E[A_i|C_{1,i}, C_{2,i}]) = -0.5 + 0.1C_{1,i} + 0.2C_{2,i}$$

$$A_i \sim \text{Bernoulli}(p = E[A_i|C_{1,i}, C_{2,i}])$$

Continuous mediator

$$E[M_i|A_i, C_{1,i}, C_{2,i}] = 0 + 0.5A_i - 0.1C_{1,i} + 0.3C_{2,i}$$

$$M_i \sim N(E[M_i|A_i, C_{1,i}, C_{2,i}], 0.5^2)$$

Exponential event time

$$\lambda_{T_i} = 0 + 0.3A_i + 0.15M_i + 0.05A_iM_i - 0.2C_{1,i} - 0.4C_{2,i}$$

$$T_i \sim \text{Exponential}(\lambda_{T_i})$$

Exponential censoring time

$$C_i \sim \text{Exponential}(\exp(-0.5))$$

Observed time and event

$$\text{time}_i = \min(T_i, C_i, 10)$$

$$\text{event}_i = I(T_i < C_i)I(T_i < 10)$$

```
suppressMessages(library(tidyverse))
library(survival)
library(tableone)
set.seed(248361264)
## Sample size
n <- 1000
## Demo data
demo <-
  ## Covariates
  tibble(id = seq_len(n),
         c1 = rnorm(n = n, mean = 0, sd = 1),
         c2 = rnorm(n = n, mean = 0, sd = 1)) %>%
  ## Binary treatment
  mutate(logit_p_a = (-0.5) + (0.1 * c1) + (0.2 * c2),
         p_a = exp(logit_p_a) / (1 + exp(logit_p_a)),
         a = rbinom(n = n, size = 1, prob = p_a)) %>%
  ## Continuous mediator
  mutate(lp_m = 0 + (0.5 * a) + (-0.1 * c1) + (0.3 * c2),
         m = rnorm(n = n, mean = lp_m, sd = 0.5)) %>%
  ## Exponential event time
  mutate(lp_t = 0 + (0.3 * a) + (0.15 * m) + (0.05 * a * m) + (-0.2 * c1) + (-0.4 * c2),
         ## The "mean" argument is on the linear predictor scale.
         t = rsurvreg(n = n, mean = lp_t, scale = 1, distribution = "exponential")) %>%
  ## Exponential censoring time
```

```

mutate(log_rate_cens = -0.5,
       cens = rexp(n = n, rate = exp(log_rate_cens)),
       admin_cens = 10) %>%
## Observed time and event
mutate(time = pmin(t, cens, admin_cens),
       event = as.numeric((t < cens) & (t < admin_cens))) %>%
## Clean
select(id, c1, c2, a, m, time, event)
## Summarize generated data
cat("### Summary of the generated data\n")
CreateTableOne(vars = c("c1", "c2", "m", "time", "event"),
               strata = c("a"),
               data = demo,
               factorVars = c("event"),
               test = FALSE,
               addOverall = TRUE) %>%
print()

```

```

### Summary of the generated data
               Stratified by a
               Overall      0      1
n               1000      631      369
c1 (mean (SD))  0.02 (0.98) -0.01 (0.97) 0.08 (1.00)
c2 (mean (SD)) -0.01 (0.99) -0.07 (0.98) 0.08 (1.01)
m (mean (SD))   0.19 (0.66) -0.02 (0.60) 0.55 (0.60)
time (mean (SD)) 0.70 (0.74)  0.63 (0.65) 0.81 (0.86)
event = 1 (%)   588 (58.8)   385 (61.0) 203 (55.0)

```

3.3 Reproducing the data analysis example

The example data analysis included in the paper was conducted as follows. Please note all variables must be numeric. Multi-category variables must be recoded as multiple dichotomous (0, 1) variables.

```

library(regmedint)
## Model fitting
res_obj <- regmedint(data = demo,
                    yvar = "time", avar = "a", mvar = "m", cvar = c("c1", "c2"), eventvar = "event",
                    a0 = 0, a1 = 1, m_cde = 0, c_cond = c(0.02, -0.01),
                    mreg = "linear", yreg = "survAFT_weibull",
                    interaction = TRUE, casecontrol = FALSE)
## Result extraction
summary(res_obj, exponentiate = TRUE)

```

Mediator model

Call:

```
lm(formula = m ~ a + c1 + c2, data = data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.43491	-0.34054	0.01103	0.35847	1.78420

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.003209	0.019969	0.161	0.872
a	0.528667	0.032951	16.044	< 2e-16 ***
c1	-0.107149	0.016179	-6.623	5.75e-11 ***
c2	0.314042	0.016019	19.604	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5009 on 996 degrees of freedom

Multiple R-squared: 0.424, Adjusted R-squared: 0.4223

F-statistic: 244.4 on 3 and 996 DF, p-value: < 2.2e-16

Outcome model

Call:

```
survival::survreg(formula = Surv(time, event) ~ a * m + c1 +
                  c2, data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.01316	0.05113	-0.26	0.797
a	0.23465	0.10928	2.15	0.032
m	0.15233	0.09494	1.60	0.109
c1	-0.19270	0.04069	-4.74	0.0000022
c2	-0.45450	0.05026	-9.04	< 2e-16
a:m	0.23899	0.14572	1.64	0.101
Log(scale)	-0.00293	0.03224	-0.09	0.928

Scale= 0.997


```

Weibull distribution
Loglik(model)= -622.1   Loglik(intercept only)= -686.5
Chisq= 128.85 on 5 degrees of freedom, p= 4.2e-26
Number of Newton-Raphson Iterations: 5
n= 1000

### Mediation analysis
      est      se      Z      p      lower      upper
cde  0.23464922 0.10928376 2.147156 0.0317809020070 0.02045699 0.4488415
pnde 0.25045086 0.10264589 2.439950 0.0146892902740 0.04926860 0.4516331
tnie 0.20687841 0.06893865 3.000906 0.0026917751336 0.07176114 0.3419957
tnde 0.37679616 0.09968080 3.780027 0.0001568110487 0.18142538 0.5721669
pnie 0.08053311 0.05044362 1.596497 0.1103777869362 -0.01833457 0.1794008
te   0.45732926 0.08868248 5.156929 0.0000002510326 0.28351480 0.6311437
pm   0.50917486 0.15919090 3.198517 0.0013813617786 0.19716643 0.8211833
      exp(est) exp(lower) exp(upper)
cde  1.264465  1.0206677  1.566496
pnde 1.284604  1.0505025  1.570876
tnie 1.229833  1.0743987  1.407754
tnde 1.457607  1.1989251  1.772103
pnie 1.083865  0.9818325  1.196500
te   1.579849  1.3277885  1.879759
pm   NA        NA        NA

Evaluated at:
avar: a
  a1 (intervened value of avar) = 1
  a0 (reference value of avar) = 0
mvar: m
  m_cde (intervend value of mvar for cde) = 0
cvar: c1 c2
  c_cond (covariate vector value) = 0.02 -0.01

Note that effect estimates can vary over m_cde and c_cond values when interaction = TRUE.

```

As seen above, the mediator regression model, the outcome regression model, and the mediation analysis results are presented. The results also include the values of **a0** (reference treatment level), **a1** (treatment level of interest), **m_cde** (value at which the mediator is fixed; relevant for **cde** only), and **c_cond** (covariate vector value; relevant for natural direct and indirect effects only) at which the effect estimates are evaluated.

The **summary** function has the corresponding arguments **a0**, **a0**, **m_cde**, and **c_cond** to override these without unnecessarily refitting the models. Here the **m_cde** value is overridden (**coef** is used to avoid printing the same mediator and outcome model results).

```

## Re-evaluate cde at m_cde = 0.5
coef(summary(res_obj, exponentiate = TRUE, m_cde = 0.5))

```

```

      est      se      Z      p      lower      upper
cde  0.35414333 0.09710871 3.646875 0.0002654488526 0.16381376 0.5444729
pnde 0.25045086 0.10264589 2.439950 0.0146892902740 0.04926860 0.4516331
tnie 0.20687841 0.06893865 3.000906 0.0026917751336 0.07176114 0.3419957
tnde 0.37679616 0.09968080 3.780027 0.0001568110487 0.18142538 0.5721669
pnie 0.08053311 0.05044362 1.596497 0.1103777869362 -0.01833457 0.1794008
te   0.45732926 0.08868248 5.156929 0.0000002510326 0.28351480 0.6311437
pm   0.50917486 0.15919090 3.198517 0.0013813617786 0.19716643 0.8211833
      exp(est) exp(lower) exp(upper)
cde  1.424959  1.1779949  1.723700
pnde 1.284604  1.0505025  1.570876
tnie 1.229833  1.0743987  1.407754
tnde 1.457607  1.1989251  1.772103
pnie 1.083865  0.9818325  1.196500
te   1.579849  1.3277885  1.879759
pm   NA        NA        NA

```

The following code was used to create the table included in the paper.

```

cols <- c("est", "lower", "upper", "exp(est)", "exp(lower)", "exp(upper)")
rows <- c("te", "pnde", "tnie", "pm")
## Construct a production-level table
tab <- rbind(coef(summary(res_obj, exponentiate = TRUE))[rows, cols],
             coef(summary(res_obj, exponentiate = TRUE, m_cde = 0.0))[rows, cols, drop = FALSE],
             coef(summary(res_obj, exponentiate = TRUE, m_cde = 0.5))[rows, cols, drop = FALSE]) %>%
  as.data.frame() %>%
  tibble::rownames_to_column() %>%
  mutate(rownames = case_when(rownames == "te" ~ "Total Effect (TE)",
                              rownames == "pnde" ~ "Pure Natural Direct Effect (PNDE)",
                              rownames == "tnie" ~ "Total Natural Indirect Effect (TNIE)",

```

```

    rowname == "pm" ~ " Proportion Mediated (PM)",
    rowname == "cde" ~ " CDE at m = 0",
    rowname == "cde.1" ~ " CDE at m = 0.5")) %>%
mutate_at(.vars = vars(-starts_with("rowname")), function(vec) {
  if_else(!is.na(vec), sprintf("%.2f", vec), "")
}) %>%
mutate(est_ci = sprintf("%s [%s, %s]", est, lower, upper),
  exp_est_ci = sprintf("%s [%s, %s]", `exp(est)` , `exp(lower)` , `exp(upper)`)) %>%
mutate(exp_est_ci = if_else(exp_est_ci == " [, ]", "-", exp_est_ci)) %>%
add_row(rowname = "Controlled Direct Effect (CDE)",
  est_ci = "", exp_est_ci = "",
  .before = 5) %>%
select(rowname, est_ci, exp_est_ci) %>%
rename(Effect = rowname,
  `Est. [95% CI]` = est_ci,
  `Exp(Est.) [95% CI]` = exp_est_ci)
## Write to an Excel file
openxlsx::write.xlsx(x = tab,
  file = "./supplement.xlsx")
## Print as a LaTeX table
tab %>%
mutate(Effect = str_replace_all(Effect, "^ ", "- ")) %>%
xtable::xtable() %>%
print(include.rownames = FALSE)

```

Effect	Est. [95% CI]	Exp(Est.) [95% CI]
Total Effect (TE)	0.46 [0.28, 0.63]	1.58 [1.33, 1.88]
- Pure Natural Direct Effect (PNDE)	0.25 [0.05, 0.45]	1.28 [1.05, 1.57]
- Total Natural Indirect Effect (TNIE)	0.21 [0.07, 0.34]	1.23 [1.07, 1.41]
- Proportion Mediated (PM)	0.51 [0.20, 0.82]	-
Controlled Direct Effect (CDE)		
- CDE at m = 0	0.23 [0.02, 0.45]	1.26 [1.02, 1.57]
- CDE at m = 0.5	0.35 [0.16, 0.54]	1.42 [1.18, 1.72]

All effect estimates are estimates of conditional effects, conditioning on the covariates in `c_var`. Since the outcome model is an accelerated failure time model (Weibull model), the estimates are on the log mean time ratio scale (positive values indicate beneficial effects). The exponentiated estimates are on the mean time ratio scale (values > 1.0 indicate beneficial effects). As the outcome model is a non-linear model, the proportion mediated is calculated using a transformation of the ratio scale [Vanderweele and Vansteelandt, 2010].

4 Demonstration of covariate-dependence of natural effect estimates

We here use the `regmedint` package to describe a somewhat overlooked aspect of the regression-based estimation method, *i.e.*, covariate-dependence of conditional effect estimates. Although neither the mediator model nor the outcome model has interaction terms between the treatment variable and baseline covariates, the resulting conditional effect estimates are dependent on the baseline covariate vector value at which they are evaluated (`c_cond` argument). This subtle point regarding the regression-based causal mediation method has previously been mentioned [Steen et al., 2017, Starkopf et al., 2017], but remains relatively unknown. Using simulated data, we visually demonstrate the issue, clarify their implications on estimation of marginal effects, and then state potential approaches to reporting of the results.

4.1 Generating demonstration data

As explained in Section 2, there are four sets of formulas to cover linear and logistic mediator models as well as linear and non-linear outcome models. Here we generate normal (linear) and Poisson (non-linear) outcomes for demonstration. For simplicity, only one continuous covariate (`c1`) is involved in the simulated data.

$$n = 1000$$

Covariates

$$C_{1,i} \stackrel{\text{iid}}{\sim} N(0, 1)$$

Binary treatment

$$\text{logit}(E[A_i|C_{1,i}]) = -0.5 + 0.1C_{1,i}$$

$$A_i \sim \text{Bernoulli}(p = E[A_i|C_{1,i}])$$

Mediator, continuous

$$E[M_{cont,i}|A_i, C_{1,i}, C_{2,i}] = 0 + 0.5A_i + 0.5C_{1,i}$$

$$M_{cont,i} \sim N(E[M_i|A_i, C_{1,i}, C_{2,i}], 0.5^2)$$

Mediator, binary

$$E[M_{bin,i}|A_i, C_{1,i}, C_{2,i}] = -0.9 + 0.5A_i + 0.5C_{1,i}$$

$$M_{bin,i} \sim \text{Bernoulli}(E[M_i|A_i, C_{1,i}, C_{2,i}])$$

Outcome, continuous (continuous mediator)

$$E[Y_{cont,Mcont,i}|A_i, M_{cont,i}, C_{1,i}] = 0 + 0.3A_i + 0.15M_{cont,i} + 0.05A_iM_{cont,i} - 0.2C_{1,i}$$

$$Y_{cont,Mcont,i} \sim N(E[Y_{cont,Mcont,i}|A_i, M_{cont,i}, C_{1,i}], 1)$$

Outcome, continuous (binary mediator)

$$E[Y_{cont,Mbin,i}|A_i, M_{bin,i}, C_{1,i}] = 0 + 0.3A_i + 0.15M_{bin,i} + 0.45A_iM_{bin,i} - 0.2C_{1,i}$$

$$Y_{cont,Mbin,i} \sim N(E[Y_{cont,Mbin,i}|A_i, M_{bin,i}, C_{1,i}], 1)$$

Outcome, Poisson (continuous mediator)

$$\log(E[Y_{pois,Mcont,i}|A_i, M_{cont,i}, C_{1,i}]) = 0 + 0.3A_i + 0.15M_{cont,i} + 0.05A_iM_{cont,i} - 0.2C_{1,i}$$

$$Y_{pois,Mcont,i} \sim \text{Poisson}(E[Y_{pois,Mcont,i}|A_i, M_{cont,i}, C_{1,i}])$$

Outcome, Poisson (binary mediator)

$$\log(E[Y_{pois,Mbin,i}|A_i, M_{bin,i}, C_{1,i}]) = 0 + 0.3A_i + 0.15M_{bin,i} + 0.45A_iM_{bin,i} - 0.2C_{1,i}$$

$$Y_{pois,Mbin,i} \sim \text{Poisson}(E[Y_{pois,Mbin,i}|A_i, M_{bin,i}, C_{1,i}])$$

```

suppressMessages(library(tidyverse))
library(tableone)
set.seed(248361264)
## Sample size
n <- 1000
## Demo data
demo2 <-
  ## Covariates
  tibble(id = seq_len(n),
    c1 = rnorm(n = n, mean = 0, sd = 1)) %>%
  ## Binary treatment
  mutate(logit_p_a = (-0.5) + (0.1 * c1),
    p_a = exp(logit_p_a) / (1 + exp(logit_p_a)),
    a = rbinom(n = n, size = 1, prob = p_a)) %>%
  ## Continuous and binary mediator
  mutate(lp_m_cont = 0 + (0.5 * a) + (0.5 * c1),
    m_cont = rnorm(n = n, mean = lp_m_cont, sd = 0.5),
    lp_m_bin = -0.9 + (0.5 * a) + (0.5 * c1),
    p_m_bin = exp(lp_m_bin) / (1 + exp(lp_m_bin)),
    m_bin = rbinom(n = n, size = 1, prob = p_m_bin)) %>%
  ## Outcomes
  mutate(lp_y_m_cont = 0 + (0.3 * a) + (0.15 * m_cont) + (0.05 * a * m_cont) + (-0.2 * c1),
    lp_y_m_bin = 0 + (0.3 * a) + (0.15 * m_bin) + (0.45 * a * m_bin) + (-0.2 * c1),
    ## Normal continuous outcomes
    y_cont_m_cont = rnorm(n = n, mean = lp_y_m_cont, sd = 1),
    y_cont_m_bin = rnorm(n = n, mean = lp_y_m_bin, sd = 1),
    ## Poisson count outcomes
    y_pois_m_cont = rpois(n = n, lambda = exp(lp_y_m_cont)),
    y_pois_m_bin = rpois(n = n, lambda = exp(lp_y_m_bin)),
    ) %>%
  ## Clean
  select(id, c1, a, m_cont, m_bin,
    y_cont_m_cont, y_cont_m_bin,
    y_pois_m_cont, y_pois_m_bin)
## Summarize generated data
CreateTableOne(vars = setdiff(names(demo2), "id"),
  strata = c("a"),
  data = demo2,
  test = FALSE,

```

```

      addOverall = TRUE) %>%
print()

```

	Stratified by a			
	Overall	0	1	
n	1000	668	332	
c1 (mean (SD))	0.02 (0.98)	0.03 (1.00)	-0.01 (0.94)	
a (mean (SD))	0.33 (0.47)	0.00 (0.00)	1.00 (0.00)	
m_cont (mean (SD))	0.18 (0.75)	0.05 (0.72)	0.45 (0.72)	
m_bin (mean (SD))	0.35 (0.48)	0.29 (0.46)	0.46 (0.50)	
y_cont_m_cont (mean (SD))	0.12 (1.05)	-0.01 (1.06)	0.38 (0.99)	
y_cont_m_bin (mean (SD))	0.26 (1.04)	0.08 (1.04)	0.62 (0.95)	
y_pois_m_cont (mean (SD))	1.15 (1.07)	1.03 (1.00)	1.39 (1.17)	
y_pois_m_bin (mean (SD))	1.37 (1.32)	1.09 (1.04)	1.93 (1.61)	

4.2 Demonstrating covariate dependence visually

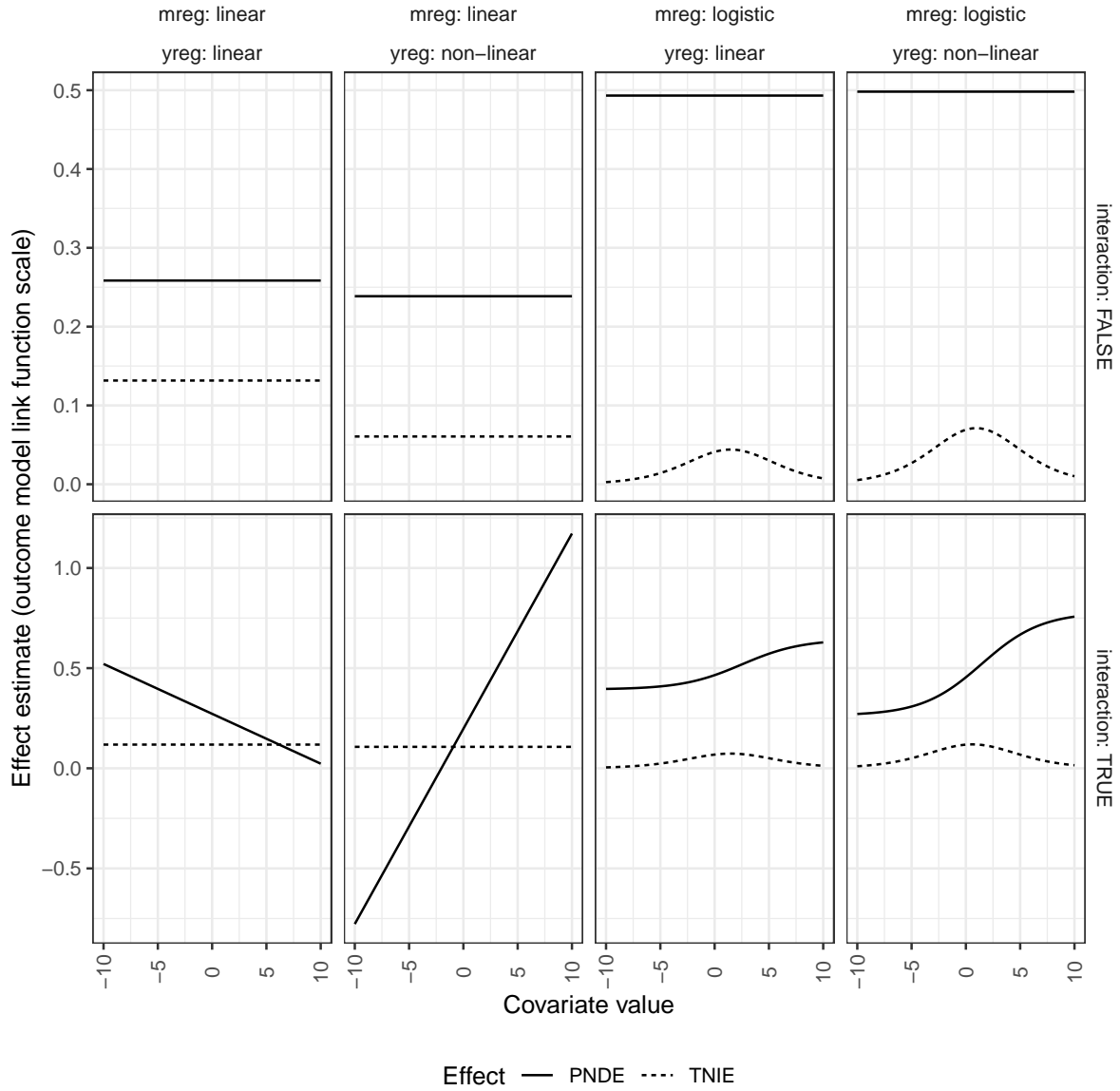
Here we fit models 8 times to cover the four patterns stated above with or without an treatment-mediator interaction term in the outcome model. The resulting effect estimates are re-evaluated at various covariate values to reveal their relationship.

```

suppressMessages(library(tidyverse))
## A grid of c_cond values to evaluate the effects at.
c_cond_grid <- seq(from = -10, to = +10, by = 0.1)
## Collect demonstration results
res <-
  ## Construct 8 patterns
  tibble(mreg = rep(c("linear", "linear", "logistic", "logistic"), 2),
         yreg = rep(c("linear", "poisson", "linear", "poisson"), 2),
         interaction = rep(c(TRUE, FALSE), each = 4)) %>%
  mutate(mvar = if_else(mreg == "linear", "m_cont", "m_bin"),
         yvar = case_when(mreg == "linear" & yreg == "linear" ~ "y_cont_m_cont",
                          mreg == "logistic" & yreg == "linear" ~ "y_cont_m_bin",
                          mreg == "linear" & yreg == "poisson" ~ "y_pois_m_cont",
                          mreg == "logistic" & yreg == "poisson" ~ "y_pois_m_bin"),
         ## Fit regression models
         fit = pmap(list(mreg, yreg, interaction, yvar, mvar),
                    function(mreg, yreg, interaction, yvar, mvar) {
                      regmedint(data = demo2,
                                yvar = yvar, avar = "a", mvar = mvar, cvar = "c1", eventvar = NULL,
                                a0 = 0, a1 = 1, m_cde = 0, c_cond = 0,
                                mreg = mreg, yreg = yreg,
                                interaction = interaction, casecontrol = FALSE)
                    }) %>%
  mutate(data = map(fit, function(fit) {
    ## Re-evaluate at each c_cond value using the coef method.
    map(c_cond_grid, function(c_cond) {
      ## Force in a row tibble.
      as_tibble(t(c(c_cond = c_cond, coef(fit, c_cond = c_cond))))
    }) %>%
    ## Collect all rows in one tibble
    bind_rows()
  }) %>%
  select(-yvar, -fit, -mvar) %>%
  unnest(data) %>%
  ## Long format
  pivot_longer(cols = c(-mreg, -yreg, -interaction, -c_cond))

## Plot
res %>%
  filter(name %in% c("pnde", "tnie")) %>%
  mutate(name = factor(name,
                       levels = c("pnde", "tnie"),
                       labels = c("PNDE", "TNIE")),
         yreg = if_else(yreg == "linear", "linear", "non-linear")) %>%
  ggplot(mapping = aes(x = c_cond, y = value, group = name, linetype = name)) +
  geom_line() +
  scale_y_continuous(name = "Effect estimate (outcome model link function scale)") +
  scale_x_continuous(name = "Covariate value") +
  scale_linetype(name = "Effect") +
  facet_grid(interaction ~ mreg + yreg, scales = "free_y", labeller = label_both) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5),
        legend.position = "bottom",
        legend.key = element_blank(),
        plot.title = element_text(hjust = 0.5),
        strip.background = element_blank())

```



`mreg` refers to the mediator model specification (linear or logistic). `yreg` refers to the outcome model specification (linear or non-linear).

The following relationship between the effect estimates and covariate value can be observed.

Interaction	Effect	mreg: linear yreg: linear	mreg: linear yreg: non-linear	mreg: logistic yreg: linear	mreg: logistic yreg: non-linear
FALSE ($\theta_3 = 0$)	PNDE	Constant	Constant	Constant	Constant
	TNIE	Constant	Constant	Varies	Varies
TRUE ($\theta_3 \neq 0$)	PNDE	Varies	Varies	Varies	Varies
	TNIE	Constant	Constant	Varies	Varies

The reasons for these specific covariate dependences can be understood by examining the PNDE and TNIE formulas in Section 2. Covariate dependence exists when the formula contains the term $\beta_2^T \mathbf{c}$, which is the covariate part of linear predictor for the mediator model. If either one of PNDE and TNIE contains this term, TE (PNDE + TNIE) is also dependent on the covariates.

4.3 Implications for marginal interpretation

These covariate-dependence of conditional natural effects have implications on whether and how marginal natural effects can be obtained with `regmedint`. In summary, only in the case of a linear mediator model and linear outcome model, one can obtain the marginal (population average) natural effects by evaluating the effect estimates at the mean covariate vector $E[\mathbf{C}]$ (mean covariate vector in practice).

4.3.1 Linear mediator model, linear outcome model

In the case with a linear mediator model with a linear (Section 2.3) outcome model, the PNDE (and also TNDE) formula contains the term $\beta_2^T \mathbf{c}$. Notably PNDE (and also TNDE) is linear in the covariate vector \mathbf{c} as well as the counterfactual outcomes. Thus, evaluating this PNDE (and also TNDE) formula at the mean covariate vector $E[\mathbf{C}]$ gives the marginal effect.

$$\begin{aligned} \text{Conditional PNDE} &= E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c}) \right\} (a_1 - a_0) \end{aligned}$$

$$\text{Marginal PNDE} = E[Y_{a_1, \widetilde{M}_{a_0}}] - E[Y_{a_0, \widetilde{M}_{a_0}}]$$

By iterative expectation

$$\begin{aligned} &= E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] \right] - E \left[E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \\ &= E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \end{aligned}$$

Inside is conditional PNDE at \mathbf{C}

$$\begin{aligned} &= E \left[\left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}) \right\} (a_1 - a_0) \right] \\ &= \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}]) \right\} (a_1 - a_0) \end{aligned}$$

i.e., Equal to conditional PNDE formula evaluated at $E[\mathbf{C}]$

When there is no treatment-mediator interaction term in the outcome model ($\theta_3 = 0$), the term $\beta_2^T \mathbf{c}$ drops out of conditional PNDE formula. Only $\theta_1(a_1 - a_0)$ (treatment coefficient in the outcome model; also agrees with CDE) remains as the direct effect, thus, eliminating the covariate dependence. Under this modeling assumption of no treatment-mediator interaction term in the outcome model, the marginal effects are the same as the conditional effects because the latter do not vary across covariate patterns.

4.3.2 Linear mediator model, non-linear outcome model

In the case with a linear mediator model with a non-linear (Section 2.4) outcome model, the PNDE (and also TNDE) formula contain the term $\beta_2^T \mathbf{c}$. The PNDE (and also TNDE) formula is linear in the covariate vector \mathbf{c} . However, the effect is defined on the outcome model link function scale (not linear in counterfactual outcomes). This non-linearity of the outcome model link function (logit in the example below) prevents the marginal interpretation of the conditional PNDE evaluated at $E[\mathbf{C}]$. The interpretation of the effect evaluated at $E[\mathbf{C}]$ is the *conditional* effect for an *average individual*.

$$\begin{aligned} \text{Conditional PNDE} &= \text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}]) \\ &\approx \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c} + \theta_2 \sigma^2) \right\} (a_1 - a_0) + \frac{1}{2} \theta_3^2 \sigma^2 (a_1^2 - a_0^2) \end{aligned}$$

$$\text{Marginal PNDE} = \text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}}])$$

By iterative expectation

$$= \text{logit} \left(E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \right) - \text{logit} \left(E \left[E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \right)$$

By non-linearity of outcome model link function

$$\neq E \left[\text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}]) \right]$$

By the conditional PNDE formula

$$\approx E \left[\left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C} + \theta_2 \sigma^2) \right\} (a_1 - a_0) + \frac{1}{2} \theta_3^2 \sigma^2 (a_1^2 - a_0^2) \right]$$

$$= \left\{ \theta_1 + \theta_3(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}] + \theta_2 \sigma^2) \right\} (a_1 - a_0) + \frac{1}{2} \theta_3^2 \sigma^2 (a_1^2 - a_0^2)$$

i.e., Not equal not conditional PNDE formula evaluated at $E[\mathbf{C}]$

When there is no treatment-mediator interaction term in the outcome model ($\theta_3 = 0$), the term $\beta_2^T \mathbf{c}$ drops out of the conditional PNDE (and also TNDE) formula. Only $\theta_1(a_1 - a_0)$ (treatment coefficient in the outcome model; also agrees with CDE) remains as the direct effect, thus, eliminating the covariate dependence. Under this modeling assumption of no treatment-mediator interaction term in the outcome model, the marginal effects are the same as the conditional effects because the latter do not vary across covariate patterns.

4.3.3 Logistic mediator model, linear outcome model

In the case with a logistic mediator model with a linear (Section 2.5) outcome model, both the PNDE (and also TNDE) formula and the TNIE (and also PNIE) formula are non-linear functions of $\beta_2^T \mathbf{c}$. Therefore, covariate dependence exists for both PNDE and TNIE. This non-linearity comes from the mediator model link function (logit). Because of this non-linearity, evaluation at $E[\mathbf{C}]$ does not give a marginal interpretation. The interpretation the effect evaluated at $E[\mathbf{C}]$ is the *conditional* effect for an *average individual*. The PNDE is examined first.

$$\begin{aligned} \text{Conditional PNDE} &= E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= \{\theta_1(a_1 - a_0)\} + \{\theta_3(a_1 - a_0)\} \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})} \end{aligned}$$

$$\text{Marginal PNDE} = E[Y_{a_1, \widetilde{M}_{a_0}}] - E[Y_{a_0, \widetilde{M}_{a_0}}]$$

By iterative expectation

$$\begin{aligned} &= E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] \right] - E \left[E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \\ &= E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] - E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \end{aligned}$$

Inside is conditional PNDE at \mathbf{C}

$$\begin{aligned} &= E \left[\left\{ \theta_1(a_1 - a_0) \right\} + \left\{ \theta_3(a_1 - a_0) \right\} \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})} \right] \\ &= \left\{ \theta_1(a_1 - a_0) \right\} + \left\{ \theta_3(a_1 - a_0) \right\} E \left[\frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})} \right] \end{aligned}$$

By non-linearity of mediator model link function

$$\neq \left\{ \theta_1(a_1 - a_0) \right\} + \left\{ \theta_3(a_1 - a_0) \right\} \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])}$$

i.e., Not equal not conditional PNDE formula evaluated at $E[\mathbf{C}]$

The TNIE formula is examined below.

$$\begin{aligned} \text{Conditional } TNIE &= E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}] - E[Y_{a_1, M_{a_0}} | \mathbf{C} = \mathbf{c}] \\ &= (\theta_2 + \theta_3 a_1) \left\{ \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})} - \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})} \right\} \end{aligned}$$

$$\text{Marginal } TNIE = E[Y_{a_1, M_{a_1}}] - E[Y_{a_1, M_{a_0}}]$$

By iterative expectation

$$\begin{aligned} &= E \left[E[Y_{a_1, M_{a_1}} | \mathbf{C}] \right] - E \left[E[Y_{a_1, M_{a_0}} | \mathbf{C}] \right] \\ &= E \left[E[Y_{a_1, M_{a_1}} | \mathbf{C}] - E[Y_{a_1, M_{a_0}} | \mathbf{C}] \right] \end{aligned}$$

Inside is conditional TNIE at \mathbf{C}

$$\begin{aligned} &= E \left[(\theta_2 + \theta_3 a_1) \left\{ \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})} - \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})} \right\} \right] \\ &= (\theta_2 + \theta_3 a_1) \left\{ E \left[\frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})} \right] - E \left[\frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})} \right] \right\} \end{aligned}$$

By non-linearity of mediator model link function

$$\neq (\theta_2 + \theta_3 a_1) \left\{ \frac{\exp(\beta_0 + \beta_1 a_1 + \beta_2^T E[\mathbf{C}])}{1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T E[\mathbf{C}])} - \frac{\exp(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])}{1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])} \right\}$$

i.e., Not equal not conditional TNIE formula evaluated at $E[\mathbf{C}]$

When there is no treatment-mediator interaction term in the outcome model ($\theta_3 = 0$), the term $\beta_2^T \mathbf{c}$ cancels out of the PNDE (and also TNDE) formulas. Only $\theta_1(a_1 - a_0)$ (treatment coefficient in the outcome model; also agrees with CDE) remains as the direct effect, thus, eliminating the covariate dependence for PNDE (and also TNDE). However, the TNIE (and also PNIE) formula remain as non-linear functions of $\beta_2^T \mathbf{c}$, thus, covariate dependence persists for TNIE. Under this modeling assumption of no treatment-mediator interaction term in the outcome model, the marginal *direct* effects are the same as the conditional direct effects because the latter do not vary across covariate patterns. However, the same does not hold for the *indirect* effects.

4.3.4 Logistic mediator model, non-linear outcome model

In the cases with a logistic mediator model with a non-linear (Section 2.6) outcome model, both the PNDE (and also TNDE) formula and the TNIE (and also PNIE) formula are non-linear functions of $\beta_2^T \mathbf{c}$. Therefore, covariate dependence exists for both PNDE and TNIE. This non-linearity comes from the mediator model link function (logit). Further, the effects are defined on the outcome model link function scale (not linear in counterfactual outcomes). Because of these non-linearity, evaluation at $E[\mathbf{C}]$ does not give a marginal interpretation. The interpretation the effect evaluated at $E[\mathbf{C}]$ is the *conditional* effect for an *average individual*. The PNDE is examined first.

$$\begin{aligned} \text{Conditional } PNDE &= \text{logit}(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}]) \\ &\approx \theta_1(a_1 - a_0) \\ &\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})) \\ &\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})) \end{aligned}$$

$$\text{Marginal } PNDE = \text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}}])$$

By iterative expectation

$$= \text{logit} \left(E \left[E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \right) - \text{logit} \left(E \left[E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}] \right] \right)$$

By non-linearity of outcome model link function

$$\begin{aligned} &\neq E \left[\text{logit}(E[Y_{a_1, \widetilde{M}_{a_0}} | \mathbf{C}]) - \text{logit}(E[Y_{a_0, \widetilde{M}_{a_0}} | \mathbf{C}]) \right] \\ &\approx E[\theta_1(a_1 - a_0) \\ &\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})) \\ &\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))] \end{aligned}$$

$$\begin{aligned} &= \theta_1(a_1 - a_0) \\ &\quad + E[\log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))] \\ &\quad - E[\log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))] \end{aligned}$$

By non-linearity of mediator model link function

$$\begin{aligned} &\neq \theta_1(a_1 - a_0) \\ &\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])) \\ &\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])) \end{aligned}$$

i.e., Not equal not conditional PNDE formula evaluated at $E[\mathbf{C}]$

The TNIE formula is examined below.

$$\text{Conditional } TNIE = \text{logit}(E[Y_{a_1, M_{a_1}} | \mathbf{C} = \mathbf{c}]) - \text{logit}(E[Y_{a_0, M_{a_0}} | \mathbf{C} = \mathbf{c}])$$

$$\begin{aligned} &\approx \log(1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})) \\ &\quad - \log(1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})) \\ &\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{c})) \\ &\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{c})) \end{aligned}$$

$$\text{Marginal } TNIE = \text{logit}(E[Y_{a_1, M_{a_1}}]) - \text{logit}(E[Y_{a_0, M_{a_0}}])$$

By iterative expectation

$$= \text{logit} \left(E \left[E[Y_{a_1, M_{a_1}} | \mathbf{C}] \right] \right) - \text{logit} \left(E \left[E[Y_{a_0, M_{a_0}} | \mathbf{C}] \right] \right)$$

By non-linearity of outcome model link function

$$\begin{aligned} &\neq E \left[\text{logit}(E[Y_{a_1, M_{a_1}} | \mathbf{C}]) - \text{logit}(E[Y_{a_0, M_{a_0}} | \mathbf{C}]) \right] \\ &\approx E[\log(1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C})) \\ &\quad - \log(1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})) \\ &\quad + \log(1 + \exp(\theta_2 + \theta_3 a_1 + \beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C})) \\ &\quad - \log(1 + \exp(\theta_2 + \theta_3 a_0 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))] \end{aligned}$$

$$\begin{aligned}
&= E[\log(1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))] \\
&\quad - E[\log(1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C}))] \\
&\quad + E[\log(1 + \exp(\theta_2 + \theta_3 \tilde{a}_1 + \beta_0 + \beta_1 a_1 + \beta_2^T \mathbf{C}))] \\
&\quad - E[\log(1 + \exp(\theta_2 + \theta_3 \tilde{a}_1 + \beta_0 + \beta_1 a_0 + \beta_2^T \mathbf{C}))]
\end{aligned}$$

By non-linearity of mediator model link function

$$\begin{aligned}
&\neq \log(1 + \exp(\beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}])) \\
&\quad - \log(1 + \exp(\beta_0 + \beta_1 a_1 + \beta_2^T E[\mathbf{C}])) \\
&\quad + \log(1 + \exp(\theta_2 + \theta_3 \tilde{a}_1 + \beta_0 + \beta_1 a_1 + \beta_2^T E[\mathbf{C}])) \\
&\quad - \log(1 + \exp(\theta_2 + \theta_3 \tilde{a}_1 + \beta_0 + \beta_1 a_0 + \beta_2^T E[\mathbf{C}]))
\end{aligned}$$

i.e., Not equal not conditional TNIE formula evaluated at $E[\mathbf{C}]$

When there is no treatment-mediator interaction term in the outcome model ($\theta_3 = 0$), the term $\beta_2^T \mathbf{c}$ cancels out of the PNDE (and also TNDE) expressions. Only $\theta_1(a_1 - a_0)$ (treatment coefficient in the outcome model; also agrees with CDE) remains as the direct effect, thus, eliminating the covariate dependence for PNDE (and also TNDE). However, the TNIE (and also PNIE) expressions remain as non-linear functions of $\beta_2^T \mathbf{c}$, thus, covariate dependence persists for TNIE. Under this modeling assumption of no treatment-mediator interaction term in the outcome model, the marginal *direct* effects are the same as the conditional direct effects because the latter do not vary across covariate patterns. However, the same does not hold for the *indirect* effects.

4.4 Practical implications for reporting

Evaluating the natural direct and indirect effects at the mean covariate level is the default behavior of the SAS macro [Valeri and Vanderweele, 2013]. With the `regmedint` R package, the same can be done via setting the `c_cond` argument to the mean covariate vector.

The literature is somewhat confusing regarding when the effect estimates evaluated at the mean covariate vector have marginal interpretation. [Valeri and Vanderweele, 2013] (p139) states "For continuous outcomes, if \mathbf{C} were set at its average level we would obtain marginal effects on the entire population." [VanderWeele, 2015] (p37) states "... evaluated at the mean covariate levels are shown. For a continuous mediator, these effects will also be equal to the marginal effects on average for the population."

As shown above, both the mediator model and the outcome model have to be linear regression models. We outlined reasonable approaches to reporting in specific models below.

4.4.1 Linear mediator model, linear outcome model

Evaluation at the mean covariate vector gives the effect estimates marginal (population average) interpretation. Thus, reporting these marginal quantities for a well-defined population (set `c_cond` to the average in that population) is useful.

If there is no treatment-mediator interaction term in the outcome model, uniform conditional effects are implied and the marginal effects agree with the conditional effects.

4.4.2 Linear mediator model, non-linear outcome model

Evaluation at the mean covariate level does *not* give the effect estimates marginal (population average) interpretations. It is still possible to report the result evaluated at the mean covariate vector, however, the results should be interpreted as *conditional* effects for an *average individual*. If categorical covariates (e.g., country) are involved, an average individual can be ambiguous or non-existent (e.g., individual residing in the average country). Reporting conditional effects evaluated at several meaningful covariate patterns may be preferred.

If there is no treatment-mediator interaction term in the outcome model, uniform conditional effects are implied and the marginal effects agree with the conditional effects.

4.4.3 Logistic mediator model, linear outcome model

Evaluation at the mean covariate level does *not* give the effect estimates marginal (population average) interpretations. It is still possible to report the result evaluated at the mean covariate vector, however, the results should be interpreted as *conditional* effects for an *average individual*. If categorical covariates (e.g., country) are involved, an average individual can be ambiguous or non-existent (e.g., individual residing in the average country). Reporting conditional effects evaluated at several meaningful covariate patterns may be preferred.

The same is true even if there is no treatment-mediator interaction term in the outcome model.

4.4.4 Logistic mediator model, non-linear outcome model

Evaluation at the mean covariate level does *not* give the effect estimates marginal (population average) interpretations. It is still possible to report the result evaluated at the mean covariate vector, however, the results should be interpreted as *conditional* effects for an *average individual*. If categorical covariates (e.g., country) are involved, an average individual can be ambiguous or non-existent (e.g., individual residing in the average country). Reporting conditional effects evaluated at several meaningful covariate patterns may be preferred.

The same is true even if there is no treatment-mediator interaction term in the outcome model.

5 Bibliography

- [Hoef, 2012] Hoef, J. M. V. (2012). Who Invented the Delta Method? *The American Statistician*, 66(2):124–127.
- [Pearl, 2001] Pearl, J. (2001). Direct and Indirect Effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI’01, pages 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- [Robins and Greenland, 1992] Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155.
- [Starkopf et al., 2017] Starkopf, L., Andersen, M. P., Gerds, T. A., Torp-Pedersen, C., and Lange, T. (2017). Comparison of five software solutions to mediation analysis. page 77.
- [Steen et al., 2017] Steen, J., Loeys, T., Moerkerke, B., and Vansteelandt, S. (2017). Flexible Mediation Analysis With Multiple Mediators. *Am. J. Epidemiol.*, 186(2):184–193.
- [Valeri and Vanderweele, 2013] Valeri, L. and Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*, 18(2):137–150.
- [Valeri and VanderWeele, 2015] Valeri, L. and VanderWeele, T. J. (2015). SAS macro for causal mediation analysis with survival data. *Epidemiology*, 26(2):e23–24.
- [VanderWeele, 2015] VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, 1 edition edition.
- [VanderWeele, 2013] VanderWeele, T. J. (2013). A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology*, 24(2):224–232.
- [Vanderweele and Vansteelandt, 2010] Vanderweele, T. J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *Am. J. Epidemiol.*, 172(12):1339–1348.