Review of the Literature on Exact Formulas for Mediation Effects

May 7, 2024

1 Summary and Future Work

We have a nice expression for the total effect of X on Y. It would be nice to also do direct and indirect effects. These are given by the so-called "mediation formula" [Pearl, 2012]. There is the whole business of linking our analysis to causal inference, but I'm happy to leave that for later. For now, we can just use the expressions in terms of conditional expectations which result when the counterfactual-based definition is identifiable[1].

Revisit after I review the relevant paper some more. There is an approach to UQ for GLMMs based on automatic differentiation. See Section 5. I'm still working to understand Zheng and Cadigan [2021], but it would be helpful if one of you could look over this paper. For our current purposes, we could just use the automatic differentiation-based asymptotic standard errors for the GLMM parameters. Later, when we want to do uncertainty quantification for the predicted random effects, I think that this paper will be useful.

2 The Samoilenko and Lefebvre group

This group does essentially what B&B are proposing with exact formulas and δ-method standard errors. In particular, they avoid any requirement for rare events. They also do direct and indirect effects. However, I haven't seen any mention of random-effects/multilevel-models.

These methods are implemented in the `R` package `ExactMed` [Caubet et al., 2023]. This package does not address mixed-effects models.

2.1 Samoilenko et al. [2018]

Gives exact formulas for direct and indirect mediation effects, as well as δ-method standard errors. Binary outcome, binary mediator. No analytical SE

[1] Pearl [2012] points out that, in the absence of random-effects, his mediation effects can be estimated directly from a contingency table, with no parametric model. If there are confounders (as in our setting), this approach may break down due to the large number of possible values. Uncertainty quantification? Extension to multilevel models?

when exposure-mediator interaction is present.

2.2 Samoilenko and Lefebvre [2021]

Extends analytical results of Samoilenko et al. [2018] to handle exposure-mediator interaction term in model for outcome.

2.3 Samoilenko and Lefebvre [2023]

As Samoilenko et al. [2018], but with continuous mediator and binary response. No mixed-effects. Very nice simulation study.

2.4 Caubet et al. [2024]

Extends the work of Samoilenko and Lefebvre [2023] to case-control data.

3 Derivative-Based

Defines mediation effects in terms of derivatives. Originally proposed by Yu et al. [2014], and later extended to multilevel models by Yu and Li [2020]. They use the name "third-variable effect analysis". Documented extensively in a book by Yu and Li [2022]. Implemented in the `R` packages `mma` and `mlma` for single level and multilevel models respectively. The latter uses the bootstrap for UQ.

3.1 Maria et al. [2024]

Incorporates random-effects in the derivative-based approach. Mostly does Bayesian inference. Mentions bootstrap.

3.2 Geldhof et al. [2018]

Defines mediation effects as partial derivatives. Does UQ by bootstrap or by a very rough method called MCCI. This MCCI simulates parameter values from a normal distribution parameterized by the estimates, then evaluates the mediation effects on these simulated parameters. This gives a distribution, but it seems to me that they would be better off just using the δ-method.

3.3 Doretti et al. [2022]

Gives exact formulas for mediation effects and δ-method SEs. Decomposes total mediation effect more finely than just direct and indirect (i.e. path-specific effects). No mixed-effects.

4 Structural Equation Modelling

There are a bunch of papers which use SEM to do mediation analysis. Many of them include multilevel/mixed-effects models. I'm not very familiar with SEM methodology, but it does seem to be connected with our work. See Valeri and VanderWeele [2013] for a review and comparison with the causal-inference framework. Generally limited to linear models (i.e. continuous response and mediator), where the mediation effects' formulas are very simple. They seem to only focus on indirect effects[2].

4.1 Zigler and Ye [2019]

Review of multilevel mediation analysis using SEM. Does a Monte Carlo study comparing different approaches. UQ is based on Bauer et al. [2006].

4.2 Bauer et al. [2006]

Does mixed-effects mediation analysis based on SEMs. Only handles continuous outcome and mediator. Mediation effects are simple enough that variances are computed directly (in terms of asymptotic covariances of the regression parameter estimates). Fits the two regression models (for M and Y) simultaneously by concatenating the response vectors and multiplying covariates by indicators.

4.3 Preacher et al. [2010]

Does mediation analysis using structural equation models. Does UQ by bootstrap or based on the exact distribution of the product of two dependent normals [MacKinnon et al., 2007].

5 GLMM UQ

5.1 Zheng and Cadigan [2021]

Does joint uncertainty quantification for the parameters and random effects. Based on maximizing the Laplace Approximation to the marginal likelihood. Gives δ-method SEs with gradients from automatic differentiation.

Based on the `R` package `TMB`, which does automatic differentiation [Kristensen et al., 2016]. There is a related package, `glmmTMB`, which is specifically designed to apply automatic differentiation methods to GLMMs [Brooks et al., 2017] (their paper focuses on zero-inflated count GLMMs, but their methodology applies more generally). Zheng and Cadigan [2021] raises some concerns about the two `TMB` packages, and presents a solution.

[2] There is a short paragraph in Zitzmann and Helm [2021] near the top of column 1 on page 532 which says "the direct effect is only of little interest". They don't give a reference for this.

6 Others

6.1 Kenny et al. [2003]

Gives a simplistic definition of mediation effects for continuous models. Does UQ based on exact distribution of mediation effects obtained as a transformation of Gaussians. A nice approach, but only really works because of the simple setting.

There is a book by MacKinnon [2017] on mediation analysis, which has a whole chapter called "Multilevel Mediation Models". However, it is mostly based on the Kenny et al. [2003] and small extensions.

References

Daniel J. Bauer, Kristopher J. Preacher, and Karen M. Gil. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological Methods*, 11 (2), 2006.

Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Mächler, and Benjamin M. Bolker. `glmmTMB` balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The `R` Journal*, 9(2), 2017.

Miguel Caubet, Mariia Samoilenko, Jesse Gervais, and Geneviève Lefebvre. *ExactMed: Exact Mediation Analysis for Binary Outcomes*, 2023. URL `https://caubm.github.io/ExactMed/`. R package version 0.2.0.9000.

Miguel Caubet, Kevin L'Espérance, Anita Koushik, and Geneviève Lefebvre. An empirical evaluation of approximate and exact regression-based causal mediation approaches for a binary outcome and a continuous or a binary mediator for case-control study designs. *BMC Medical Research Methodology*, 24(1), 2024.

Marco Doretti, Martina Raggi, and Elena Stanghellini. Exact parametric causal mediation analysis for a binary outcome with a binary mediator. *Statistical Methods & Applications*, 31(1), 2022.

G. John Geldhof, Katherine P. Anthony, James P. Selig, and Carolyn A. Mendez-Luck. Accommodating binary and count variables in mediation: A case for conditional indirect effects. *International Journal of Behavioral Development*, 42(2), 2018.

David A. Kenny, Josephine D. Korchmaros, and Niall Bolger. Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 2003.

Kasper Kristensen, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. `TMB`: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 2016.

David P. MacKinnon. *Introduction to statistical mediation analysis*. Routledge, 2017.

David P. MacKinnon, Matthew S. Fritz, Jason Williams, and Chondra M. Lockwood. Distribution of the product confidence limits for the indirect effect: program `PRODCLIN`. *Behaviour Research Methods*, 39(3), 2007.

Chiara Di Maria, Claudio Rubino, and Alessandro Albano. The derivative-based approach to nonlinear mediation models: insights and applications. *Quality and Quantity*, 2024.

Judea Pearl. The causal mediation formula - a guide to the assessment of pathways and mechanisms. *Prevention Science*, 13(4), 2012.

Kristopher J. Preacher, Michael J. Zyphur, and Zhen Zhang. A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 2010.

Mariia Samoilenko and Geneviève Lefebvre. Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*, 190(9), 2021.

Mariia Samoilenko and Geneviève Lefebvre. An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*, 42(3), 2023.

Mariia Samoilenko, Lucie Blais, and Geneviève Lefebvre. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Observational Studies*, 4(1), 2018.

Linda Valeri and Tyler J. VanderWeele. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions with `SAS` and `SPSS` macros. *Psychological Methods*, 18(2), 2013.

Qingzhao Yu and Bin Li. Third-variable effect analysis with multilevel additive models. *PLoS ONE*, 15(10), 2020.

Qingzhao Yu and Bin Li. *Statistical methods for mediation, confounding and moderation analysis using `R` and `SAS`*. CRC Press, 2022.

Qingzhao Yu, Ying Fan, and Xiaocheng Wu. General multiple mediation analysis with an application to explore racial disparities in breast cancer survival. *Journal of Biometrics & Biostatistics*, 5(2), 2014.

Nan Zheng and Noel Cadigan. Frequentist delta-variance approximations with mixed-effects models and TMB. *Computational Statistics and Data Analysis*, 160, 2021.

Christina K. Zigler and Feifei Ye. A comparison of multilevel mediation modeling methods: recommendations for applied researchers. *Multivariate Behavioral Research*, 54(3), 2019.

Steffen Zitzmann and Christoph Helm. Multilevel analysis of mediation, moderation, and nonlinear effects in small samples, using expected a posteriori estimates of factor scores. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(4), 2021.