Analysis of Direct and Indirect Mediation Effects in Causal Mediation Analysis

July 15, 2024

In this document, we develop definitions and formulas for the direct and indirect mediation effects to complement the total effect given by B & B. We follow the work of Pearl [2012] and the Samoilenko and Lefebvre group (e.g., Samoilenko and Lefebvre, 2023).

1 Expected Nested Counterfactuals

To start, we define the nested counterfactual $Y(x, M(x'))$ as the value that Y would assume when X is set to x and M is set to whatever value it would have assumed if we had set X to x'. Pearl [2012] identifies the expected value of this nested counterfactual with the following expression based on conditional expectations:

$$\mathbb{E}Y(x, M(x')) = \int \mathbb{E}(Y|M = m, X = x)\mathbb{P}(M = dm|X = x') \tag{1}$$

Focusing now on the case with binary response and binary mediator, 1 becomes

$$\begin{aligned}\mathbb{E}Y(x, M(x')) = \mathbb{P}(Y = 1|M = 1, X = x)\mathbb{P}(M = 1|X = x')+ \\ \mathbb{P}(Y = 1|M = 0, X = x)\mathbb{P}(M = 0|X = x')\end{aligned} \tag{2}$$

Consider using logistic regression models for Y and M. Write $\text{logit}(\mathbb{E}(Y|M = m, X = x)) = \beta_0 + \beta_m m + \beta_x x$ and $\text{logit}(\mathbb{E}(M|X = x)) = \alpha_0 + \alpha_x x$. Then (2) becomes

$$\begin{aligned}\mathbb{E}Y(x, M(x')) = \frac{1}{1 + \exp(-\beta_0 - \beta_m - \beta_x x)} \cdot \frac{1}{1 + \exp(-\alpha_0 - \alpha_x x')}+ \\ \frac{1}{1 + \exp(-\beta_0 - \beta_x x)} \cdot \frac{1}{1 + \exp(\alpha_0 + \alpha_x x')}\end{aligned} \tag{3}$$

Equation (3) holds for logistic regression with fixed-effects only. If we instead use mixed-effects logistic regressions for Y and M, then (1) and (2) still hold, but (3) must be modified. For the mixed-effects models, first write $V = (V_0, V_m, V_x) \sim$

$\mathrm{N}(0, \Sigma_V)$ and $U = (U_0, U_x) \sim \mathrm{N}(0, \Sigma_U)$ for the random-effects in our models for Y and M respectively. Next, write $\mathrm{logit}(\mathbb{E}(Y|V, M = m, X = x)) = (\beta_0 + V_0) + (\beta_m + V_m)m + (\beta_x + V_x)x$ and $\mathrm{logit}(\mathbb{E}(M|U, X = x)) = (\alpha_0 + U_0) + (\alpha_x + U_x)x$. Returning now to identification of the expected counterfactual for Y, we get

$$
\begin{aligned}
\mathbb{E}Y(x, M(x')) = &\left[\int \mathbb{P}(Y = 1|V = v, M = 1, X = x)\mathbb{P}(V = dv)\right] \cdot && (4)\\
&\left[\int \mathbb{P}(M = 1|U = u, X = x')\mathbb{P}(U = du)\right] + && (5)\\
&\left[\int \mathbb{P}(Y = 1|V = v, M = 0, X = x)\mathbb{P}(V = dv)\right] \cdot && (6)\\
&\left[\int \mathbb{P}(M = 0|U = u, X = x')\mathbb{P}(U = du)\right] && (7)
\end{aligned}
$$

and, in the logistic regression context,

$$
\begin{aligned}
\mathbb{E}Y(x, M(x')) = &\left[\int \frac{\phi(v; 0, \Sigma_V)}{1 + \exp(-(\beta_0 + v_0) - (\beta_m + v_m) - (\beta_x + v_x)x)} dv\right] \cdot \qquad (8)\\
&\left[\int \frac{\phi(u; 0, \Sigma_U)}{1 + \exp(-(\alpha_0 + u_0) - (\alpha_x + u_x)x')} du\right] +\\
&\left[\int \frac{\phi(v; 0, \Sigma_V)}{1 + \exp(-(\beta_0 + v_0) - (\beta_x + v_x)x)} dv\right] \cdot\\
&\left[\int \frac{\phi(u; 0, \Sigma_U)}{1 + \exp((\alpha_0 + u_0) + (\alpha_x + u_x)x')} du\right],
\end{aligned}
$$

where $\phi(.; \mu, \Sigma)$ is the multivariate normal density with mean μ and covariance matrix Σ.

Note that the four integrals in (8) are all multivariate, but can be transformed to univariate integrals by suitable changes of variables. To this end, write $\eta = \alpha_0 + \alpha_x x$ and $\zeta = \beta_0 + \beta_x x$ for two linear predictors (note that ζ does not contain β_m), and $\gamma^2_\Sigma(c_1, \ldots, c_r) = (c_1, \ldots, c_r)\Sigma(c_1, \ldots, c_r)^T$, where Σ is an r-by-r covariance matrix. We will generally set $\Sigma = \Sigma_V$ or $\Sigma = \Sigma_U$, in which case we write γ^2_V or γ^2_U respectively. We now define the function ψ as a template for the four integrals in (8).

$$\psi(\mu, \sigma^2) := \int \frac{\phi(z; 0, 1)}{1 + \exp(-\mu - \sigma z)} dz.$$

Note that ψ is a univariate integral, so we can expect it to be well-approximated by routine numerical quadrature techniques. We now re-write (8) in terms of ψ as follows,

$$
\begin{aligned}
\mathbb{E}Y(x, M(x')) = &\psi(\zeta + \beta_m, \gamma^2_V(1, 1, x)) \cdot \psi(\eta, \gamma^2_U(1, x)) + \qquad (9)\\
&\psi(\zeta, \gamma^2_V(1, 0, x)) \cdot \psi(-\eta, \gamma^2_U(1, x)).
\end{aligned}
$$

2 Mediation Effects

Denote the expected nested counterfactual defined in (1) by $\mathcal{Y}(x, x') = \mathbb{E}Y(x, M(x'))$. We can define the various mediation effects in terms of expected counterfactuals. Note that mediation effects for a binary outcome are commonly defined on three different scales: risk difference, risk ratio and odds ratio. Table 1 gives all such definitions explicitly. In the rest of this section, we outline the processes of estimation and uncertainty quantification for the mediation effects defined in Table 1.

Table 1: Definitions of various mediation effects; x and x' denote different values of the exposure.

Risk Difference	Total Effect	$\mathcal{Y}(x, x) - \mathcal{Y}(x', x')$
	Direct Effect	$\mathcal{Y}(x, x') - \mathcal{Y}(x', x')$
	Indirect Effect	$\mathcal{Y}(x, x) - \mathcal{Y}(x, x')$
Risk Ratio	Total Effect	$\mathcal{Y}(x, x)/\mathcal{Y}(x', x')$
	Direct Effect	$\mathcal{Y}(x, x')/\mathcal{Y}(x', x')$
	Indirect Effect	$\mathcal{Y}(x, x)/\mathcal{Y}(x, x')$
Odds Ratio	Total Effect	$\frac{\mathcal{Y}(x,x)}{1-\mathcal{Y}(x,x)} \Big/ \frac{\mathcal{Y}(x',x')}{1-\mathcal{Y}(x',x')}$
	Direct Effect	$\frac{\mathcal{Y}(x,x')}{1-\mathcal{Y}(x,x')} \Big/ \frac{\mathcal{Y}(x',x')}{1-\mathcal{Y}(x',x')}$
	Indirect Effect	$\frac{\mathcal{Y}(x,x)}{1-\mathcal{Y}(x,x)} \Big/ \frac{\mathcal{Y}(x,x')}{1-\mathcal{Y}(x,x')}$

2.1 Estimation

Mediation effects are all defined in terms of the expected nested counterfactuals, $\mathcal{Y}(x, x')$. As such, the estimation of any mediation effect given in Table 1 centers around estimation of $\mathcal{Y}$. To this end, write θ for all parameters upon which $\mathcal{Y}$ depends. That is, θ contains both sets of regression coefficients, $\beta_0, \beta_m, \beta_x$ and α_0, α_x, as well as both sets of covariance parameters. For consistency with B & B, we parameterize these as τ_0, τ_m, τ_x for the standard deviations of V_0, V_m, V_x, and $\tau_{0,m}, \tau_{0,x}, \tau_{m,x}$ for the corresponding correlations[1]. Similarly, we use σ_0, σ_x for the standard deviations of U_0, U_x, and $\sigma_{0,x}$ for their correlation. We are now equipped to write-out θ in full. The order of parameters is chosen to match my code (and to avoid refactoring thereof).

$$\theta = (\alpha_0, \alpha_x, \sigma_0, \sigma_x, \sigma_{0,x}, \beta_0, \beta_m, \beta_x, \tau_0, \tau_m, \tau_x, \tau_{0,m}, \tau_{0,x}, \tau_{m,x})$$

In order to estimate θ, we fit two mixed-effects logistic regression models. The first model predicts the mediator, M, using the exposure, X, while the second

[1] While our notation doesn't match that given by B & B, parameterizing in terms of the standard deviations and correlations does. Alternative choices include the variances and covariances, or the unique components of the Cholesky factorizations of Σ_V and Σ_U [Wang and Merkle, 2018].

predicts the outcome, Y, using M and X. Both models also contain intercepts. We include random effects for all regression coefficients in both models. Either or both models may also contain one or more confounders, $\mathbf{C}$, although we assign these confounders only fixed effects. Following the algorithm used by the `lme4` package in `R` [Bates et al., 2015, Walker et al., 2023], we estimate the parameters in both regression models by maximizing an approximation to the marginal likelihood of the response (M or Y). This approximation is obtained using Laplace's Method, with a penalized iteratively re-weighted least squares algorithm maximizing to maximize the joint likelihood of the observed data and the random effects. Applying this procedure to both of our regression models yields estimates for all components of θ.

Combining our estimate of θ with Equation 8 from Section 1 and the formulas given in Table 1, we are now able to estimate all mediation effects on the three scales we consider. It remains however, to address uncertainty quantification.

2.2 Uncertainty Quantification

As mentioned above, all mediation effects are defined in terms of expected nested counterfactuals $\mathcal{Y}$, which themselves depend on the parameter vector θ. Quantification of uncertainty for our estimator of a mediation effect thus proceeds in three stages. First, we obtain a covariance matrix for our estimator of θ. Second, we apply the δ-method to get a joint covariance matrix for all values of $\mathcal{Y}$ required to evaluate the mediation effect of interest. Finally, we use the δ-method again to translate uncertainty in the $\mathcal{Y}$s to uncertainty in the mediation effect. In fact, we can obtain the covariance matrix for all three mediation effects defined on a particular scale (e.g., risk ratio), or, indeed, between all nine mediation effects given in Table 1. Note that the variances and covariances discussed above are asymptotic.

We now proceed with the three steps discussed above, starting with an asymptotic covariance matrix for θ. For now, we treat the two regression models as independent. That is, we assume that the covariance between all parameter estimators in our M model and all parameter estimators in our Y model are zero. This is likely not true; call it a working assumption. See Bauer et al. [2006] for a method to model this inter-model dependence. This step is fairly straightforward, since both regression models are fit using (approximate) maximum likelihood. We simply compute the negative Hessian of the marginal log-likelihood for the observed data in both models, and stack the results into a block-diagonal matrix. Call the resulting covariance matrix $\hat{\Sigma}_\theta$. This computation is performed by the `merDeriv` package in `R` [Wang and Merkle, 2018].

To quantify uncertainty in $\mathcal{Y}$ based on our estimator for θ, we use the δ-method [see, e.g., Chapter 3 of van der Vaart, 1998]. Briefly, given the asymptotic covariance of our estimator for θ, $\sqrt{n}(\hat{\theta} - \theta) \rightsquigarrow \mathrm{N}(0, \Sigma)$, the asymptotic covariance of some function of our estimator can be obtained by differentiating that function and multiplying by the limiting distribution. That is, $\sqrt{n}[f(\hat{\theta}) - f(\theta)] \rightsquigarrow \nabla f(\theta)\mathrm{N}(0, \Sigma) \overset{d}{=} \mathrm{N}(0, \nabla f(\theta)\Sigma\nabla f(\theta)^T)$. We estimate this limiting covariance by $\nabla f(\hat{\theta})\hat{\Sigma}\nabla f(\hat{\theta})^T$, where typically $\Sigma = \Sigma(\theta)$ and $\hat{\Sigma} = \Sigma(\hat{\theta})$.

Returning now to our problem, in order to apply the δ-method we need the gradient of $\mathcal{Y}$ with respect to θ. While tedious, this calculation is easily performed with the help of symbolic differentiation software like Maple [Maplesoft, a division of Waterloo Maple Inc., 2020]. Next, we evaluate our gradient formula at $\hat{\theta}$, $\nabla\mathcal{Y}|_{\theta=\hat{\theta}}$, then pre- and post-multiply $\hat{\Sigma}_\theta$ by this estimated gradient. If we require the covariance matrix for multiple values of $\mathcal{Y}$ (e.g., $\mathcal{Y}(x, x)$ and $\mathcal{Y}(x, x')$ for some $x \neq x'$), we frame the problem as a vector-valued transformation, and compute the Jacobian by stacking the gradient for each individual $\mathcal{Y}$. More precisely, to get a covariance matrix for $\mathcal{Y}_1, \ldots, \mathcal{Y}_r$, we first construct the r-by-$|\theta|$ Jacobian matrix $\mathcal{J} := [\nabla\mathcal{Y}_1, \ldots, \nabla\mathcal{Y}_r]$. We then evaluate $\mathcal{J}$ at $\theta = \hat{\theta}$, call the result $\hat{\mathcal{J}}$, and pre- and post-multiply $\hat{\Sigma}_\theta$ by $\hat{\mathcal{J}}$, giving $\hat{\mathcal{J}}\hat{\Sigma}_\theta\hat{\mathcal{J}}^T$.

Finally, to get the asymptotic variance of one or more estimated mediation effects, we just apply the δ-method to the corresponding formulas from Table 1. Since the mediation effects are all simple functions of the $\mathcal{Y}$, any gradients required for this step are easily obtained.

3 To Do

- Incorporate an interaction term between X and M in the model for Y.
- Explore dependence between the models for M and Y using the 'stacking' technique described in Bauer et al. [2006]. This will give non-zero covariance between parameter estimators from the two models.
- Incorporate covariates/confounders

References

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using `lme4`. *Journal of Statistical Software*, 67(1), 2015.

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.

Maplesoft, a division of Waterloo Maple Inc. Maple, 2020.

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

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.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Steven Walker, Rune Haubo Bojesen Christensen, Douglas Bates, Benjamin M. Bolker, and Martin Mächler. Fitting generalized linear mixed-effects models using `lme4`. *Unpublished*, 2023. URL `https://github.com/lme4/lme4/blob/master/misc/glmer_JSS/glmer.pdf`.

Ting Wang and Edgar C. Merkle. `merDeriv`: derivative computations for linear mixed effects models with applications to robust standard errors. *Journal of Statistical Software*, 87(1), 2018.