Analysis of Direct and Indirect Mediation Effects in Causal Mediation Analysis

May 28, 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

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

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. 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 $\mathscr{Y}(x, x') = \mathbb{E}Y(x, M(x'))$. 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. Combined with the formulas given throughout Section 1, we now have everything we need to do point estimation for the various mediation effects. It still remains however, to address uncertainty quantification.

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 Uncertainty Quantification

Note that all mediation effects are defined in terms of expected nested counterfactuals $\mathcal{Y}$. Thus, if we can produce a formula for the asymptotic covariance matrix of two (or more) values of $\mathcal{Y}$, we can use a simple application of the δ-method to obtain the variance of any 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.

Toward a variance formula for $\mathcal{Y}$, we first 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

[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.

chosen to match my code (and to avoid refactoring thereof).

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

There are two problems left to solve for the quantification of uncertainty in $\mathcal{Y}$. First, we need the (asymptotic) covariance matrix of θ. Second, we need the θ-gradient of the function $\mathcal{Y}$. Note that $\mathcal{Y}$ is also a function of x and x', but we treat these values as fixed for our analysis.

3 To Do

- Incorporate an interaction term between X and M in the model for Y.

References

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.