Verification of SE Formulas for Mediation Analysis

May 17, 2024

I'm doing some Monte Carlo to verify the new SE formulas. Recall that we're doing mediation analysis, so we've got a response, Y, an exposure, X, and a mediation, M. We also have some number of confounders, which will be grouped together in the matrix W. Broadly speaking, we fit two regression models, one to predict M using X and W, the other to predict Y using M, X and W. We then compute the mediation effect (specifically, the total effect of X on Y) as a function of the coefficients from these two regression models. An asymptotic SE for our mediation effect estimator can then be obtained from the asymptotic standard errors of our fitted regression coefficients using the δ-method.

So far, so simple. There are a few places that things start to get more complicated. First, each of the two regression models can be either linear of logistic depending on whether the corresponding response variable is continuous or binary[1]. Furthermore, we can add random effects to our regression models. In the trust study, we have random effects for the intercept, X and M (naturally, the latter only applies when predicting Y). The problem is slightly simpler with a single confounder (mostly, the bookkeeping is a bit easier), although I don't want to stick to this for long.

We will address each of these extra layers of complexity in turn. First though, we start with the simplest version of the problem.

1 Continuous Response, Continuous Mediation, Fixed-Effects

I set the sample size, n, to 1000. Each regression coefficient is 1 and the residual standard deviation in both regression models is 0.2. I use a single confounder and generate both it, W, and the exposure, X, as iid $\mathrm{N}(1, 1)$. I generate 1000 datasets, each with different values for X and W.

[1] In principle, we could have Y and/or M follow any distribution with a suitable GLM formulation. I don't think I've ever seen count data (i.e. Poisson regression) used here, much less anything more exotic.

Empirical	Mean Analytical	Median Analytical
0.00833	0.00896	0.00895

Table 1: Standard errors for estimated mediation effect with continuous response and mediation, fixed-effects.

On each dataset, we fit the two regression models, then extract coefficients and standard errors. Next, we compute our estimate of the mediation effect and its δ-method standard error (see Overleaf for details). After repeating this process 1000 times, we compute the empirical standard error (SD of our estimates), as well as the mean and median estimated standard error. Values are given in Table 1. As you can see, our δ-method formula works very well.

The results are very similar if we use multiple confounders (specifically, I use 3). Henceforth, I will use 3 confounders in my analysis.

2 Continuous Response, Binary Mediation, Fixed-Effects

Here things start getting messier. I will return to documenting the analysis, but first I want to discuss the proposed estimand. Note that the total effect given in Case 2 of Section 1.1, $\gamma(a, \alpha, \beta_1, \beta_2)$ depends on our choice of levels for X and W. If X is binary then its choice is easy since we're computing the effect of a one-unit increase in X. For W however, I don't see any natural choice of reference value. One option is to evaluate γ at each W in the observed dataset and average to "marginalize out" W. I have a few references which discuss doing something along these lines, but it might be worth discussing on Thursday. For now, I'm just going to choose a value of W and study "fixed-confounder" mediation effects.

I repeated the analysis in Section 1, this time with a binary mediator. Datasets are generated with $n = 1000$ observations, 3 confounders. Regression slopes are set to 1, and the intercept is set so that the expected linear predictor is approximately zero (this helps keep the number of 0s and 1s for M approximately balanced). The residual variance for Y is set to 0.2^2.

Following the analysis outlined by Bruno, I use a logistic regression model for the mediator and linear regression for the response. Formulas for the δ-method are messier, but conceptually this case is nearly identical to the continuous response setting. One difference here is that the total mediation effect depends on levels of the covariates, specifically X and W. I report results for $X = 0$ and $W = [1, 1, 1]^T$. This value of X is very natural when the exposure is binary. I have no such reason for choosing this value of W. See the opening paragraph of this section for more details.

I generated 1000 datasets, fit regression models for M (logistic) and Y (linear), then computed our estimate for the mediation effect. I also evaluated our SE formula on each dataset. Table 2 gives the empirical standard error of our

Empirical	Mean Analytical	Median Analytical
0.0200	0.0203	0.0203

Table 2: Standard errors for estimated mediation effect with continuous response and binary mediation, fixed-effects.

Empirical	Mean Analytical	Median Analytical
0.1060	0.1064	0.1062

Table 3: Standard errors for estimated mediation effect with binary response and mediation, fixed-effects.

total effect estimator, as well as the mean and median analytical SE.

3 Binary Response, Binary Mediator, Fixed-Effects

For my own future reference, I'm going to set out some notation here. a and b are the vectors of regression coefficients in a model for M and Y respectively, whether linear or logistic. We use η and ζ for the linear predictors of M and Y respectively. In my code for Section 2, I used η, but didn't really need ζ. Here I will. In the previous section, I used γ for the total mediation effect and δ for $\mathbb{E}(M|X = x + 1) - \mathbb{E}(M|X = x)$, although the latter I only used in my code. Here, I can just talk about odds and odds ratios.

I had to update Bruno's formulas. It looks like he forgot to include the $\mathbb{P}(Y = 1|X = x, \mathbf{W} = \mathbf{w})$ term in the odds, and thus in the odds ratio. This has been changed now in the Overleaf document. I also worked out all the partial derivatives using Maple, which can produce *Julia* (or *R*, *Python*) code to evaluate these horrendous expressions. All partials were validated against a simple finite difference approximation.

Table 3 gives empirical and analytical SEs when $n = 1000$.

4 Binary Response, Binary Mediator, Random-Effects

Before doing any computing, a few notes about Bruno's derivation. The calculation for $n^{1/2}(\hat{\sigma}_U(x) - \sigma_U(x))$ contains a few steps which weren't immediately obvious to me. I'm going to clarify some of them here. First, what is the point of the first step? Well, note that $\hat{\sigma}_U(x) + \sigma_U(x) = 2\sigma_U(x) + o_p(1)$, so the denominator is well-controlled. The numerator is now structured so we can use our formula for $\sigma_U^2(x)$ in terms of the components of Σ_U from the previous line. While not strictly necessary, this saves us from the first step of every derivative calculation being applying the chain rule to deal with a square root. Being a little bit more formal, what we're doing here is applying the δ-method to the numerator (with the $n^{1/2}$), then using Slutsky's Theorem/The Continuous Mapping Theorem to handle the ratio. No problems.

The second point that tripped me up briefly is how the formula is organized. This is actually chosen to match our expression for $\sigma_U^2(x)$ in terms of the components of Σ_U from the previous line. He then re-arranges to collect like terms in the parameter estimates. Third, there are a lot of $\sigma_U(x)$s in the denominators, but no 2s. This is because every term in our expression for $\sigma_U^2(x)$ is either a quadratic or multiplied by 2, so every term in the gradient of $\sigma_U^2(x)$ contains a 2.

Fourth, I point out that the definition of $\phi(\mu, \sigma)$ is pretty clever. It reduces the problem from one with multiple random effects of a univariate problem with only a single Gaussian. This simplification arises from the observation that the random effects only enter into $\mathbb{P}(Y = 1)$ through a weighted sum. Since weighted sums of Gaussians are Gaussian, we can replace the whole sum with a new single Gaussian, setting the mean and variance of this new RV appropriately.

Now onto some computation.

The approximate covariance matrix of the GLMM parameter estimates can be obtained directly from a fitted `lme4` object using the `merDeriv` package in `R`. This saves all the work I was starting in the following subsection. Please disregard.

4.1 UQ for GLMM Parameters (Old)

Immediately I've encountered a problem. The standard software reports standard errors for the fixed effects' estimators, but not for the random effects' variance parameters[2]. I guess I'm going to have to work this out myself.

To this end, I need the Hessian of the observed data (log) likelihood. There are tricks from the EM algorithm literature for doing this, but I think our problem is simple enough that I can just reduce it to a small-ish number of one-dimensional integrals. To start, recall the formula Bruno gives for $\mathbb{P}(Y = 1|X = x, \mathbf{W} = \mathbf{w}) =: P_1$ (similar for P_0). Specifically, he expresses this quantity in terms of the ϕ function evaluated at four different arguments. Thus, if I have the Hessian of ϕ, I can compute the Hessian of P_1 (and P_0), which in-turn gives me the Hessian of the log-likelihood based on a single point.

This actually isn't quite what I need to do. In order to incorporate grouping structure, I need a generalization of the ϕ function, whose integrand contains a product of terms with the form $(1 + \exp(-\mu_i - \sigma_i z))^{-1}$, with i ranging over all points in a single group. More formally, write

$$\Phi(\boldsymbol{\mu}, \boldsymbol{\sigma}) := \int \left[\prod_i \frac{1}{(1 + e^{-\mu_i - \sigma_i z})} \right] \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \tag{1}$$

[2] I'm going to be a bit sloppy here and refer to the standard error of a parameter. By this, I mean the standard error of our estimator for that parameter. It's just getting to be too much typing to always write the latter.

for the generalization of ϕ to multiple μ, σ from the same group. Write $P(z, \boldsymbol{\mu}, \boldsymbol{\sigma})$ for the product term in the integrand or, when no confusion can arise, $P(z)$ or even P. The likelihood of a dataset can now be obtained by first evaluating Φ on each group, then multiplying the results.

In order to compute the Hessian of the log-likelihood, and by extension the asymptotic variance of our MLEs, we will require the gradient and Hessian of Φ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. To this end, observe that $\nabla_\mu P = P\boldsymbol{g}$, where $\boldsymbol{g}$ is a vector with ith component $-\nabla_{\mu_i} \log[1 + \exp(-\mu_i - \sigma_i z)] = [1 + \exp(\mu_i + \sigma_i z)]^{-1}$

$$\begin{aligned}
\nabla_\mu \Phi(\boldsymbol{\mu}, \boldsymbol{\sigma}) &= \int \nabla_\mu P \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz && (2)\\
&= \int P\boldsymbol{g} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz && (3)\\
&= \int \frac{P}{1 + e^{\boldsymbol{\mu} + \boldsymbol{\sigma} z}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz && (4)
\end{aligned}$$

where arithmetic operations on vectors are interpreted componentwise. Similarly, $\nabla_\sigma P = zP\boldsymbol{g}$ and

$$\nabla_\sigma \Phi(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \int zP\boldsymbol{g} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = \int z \frac{P}{1 + e^{\boldsymbol{\mu} + \boldsymbol{\sigma} z}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \tag{5}$$

Next, we need the second-derivatives. These are obtained simply using Maple or Wolfram Alpha. To be continued...

Before working harder on the derivatives of Φ, I want to say exactly what I plan to do with this function. First, observe that the likelihood for the observed data, Y and M, trivially factors as $\mathcal{L}(Y, M|X, W) = \mathcal{L}(Y|M, X, W) \cdot \mathcal{L}(M|X, W)$. Thus, we can treat our regression models for M and $Y|M$ separately. Next observe that, within a single group,

$$\begin{aligned}
\mathcal{L}(M|X, W) &= \mathbb{P}(M|X, W) && (6)\\
&= \mathbb{E}_V \mathbb{P}(M|V, X, W) && (7)\\
&= \mathbb{E}_V \left(\left[\prod_{m_i=1} \mathbb{P}(M = 1|V, X = x_i, W = w_i) \right] \right.\\
&\quad \left. \left[\prod_{m_i=0} \mathbb{P}(M = 0|V, X = x_i, W = w_i) \right] \right) && (8)\\
&= && (9)
\end{aligned}$$

4.2 UQ for Mediation Effects

Following Section 2.2 of B & B's document, we have an expression for the mediation effect, Ψ, in terms of a small number of integrals, ϕ. Each of these integrals is univariate, and can thus be computed accurately using quadrature.

As for uncertainty quantification, we still need to compute the partial derivative of Ψ wrt each of its arguments. This shouldn't be too hard to do in Maple (hopefully). I will probably need to leave ϕ undefined, and get Maple to return an answer in terms of the partial derivatives of ϕ.

Note that Ψ depends explicitly on a combination of model parameters and transformations thereof. In particular, the functions σ_U and σ_V depend implicitly on the mixed effects' covariance matrices. Conveniently, B & B have worked out the gradients of σ_U and σ_V wrt the model parameters. They're messy, but ultimately code-able.

As of Friday, May 17, I have coded the gradient of Φ. Next I need to do UQ for each argument of Φ (e.g., $\sigma_V(x+1)$), then I can apply the δ-method.

5 Open Questions

How does the standard error decompose between multiple responses with the same fixed covariates (i.e. X and W), versus across replicated covariates? I can imagine how to do this, at least naively: do a bunch of analyses on each of a bunch of sets of covariates. Compute the average "within-covariates" variance and the average total variance. Compute the fraction of total variance due to response replication and, interestingly, one minus this fraction as a naive representation of the variance fraction due to covariate variability. **Every time I've checked this, the difference has been very small. This isn't a priority for me right now.**