

Verification of SE Formulas for Mediation Analysis

April 26, 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 or logistic depending on whether the corresponding response variable is continuous or binary¹. 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).

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 100. 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 $N(1, 1)$. I generate 1000 datasets, each with different values for X and W .

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

¹In 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
0.0289	0.0288	0.0286

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

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.