Introduction

Non-parametric bounds of the Average Causal Effect (ACE) are intriguing for several reasons. Getting bounds on the ACE with no further assumptions other than what is embedded in the IV model is very appealing in particular as such assumptions often limit the structure of unmeasured confounders. By definition, assumptions about unmeasured confounders cannot be checked, and it is therefore desirable to avoid such.

We will see how we can find nonparametric bounds in different practical settings, and how they unfortunately often lead to rather results from which a disappointing amount of insights can be gained.

Motivation

Many have considered this approach in the past. Swanson et al. (2018) provide a nice overview of different bounds derived from different sets of assumptions. Arguably the weakest set of assumptions result in what is known as the Balke-Pearl bounds, first presented in Balke and Pearl (1993).

The same set of bounds can be obtained using a geometric approach as presented by Ramsahai (2012). We will rely on this work as it provides a very general approach to finding bounds. This allows us to obtain bounds in a setting where we do not have observations of (X, Y, Z), but rather observations of (X, Z) and (Y, Z) from two different data sources. (We will refer to the former as the trivariate setting, and the latter as the bivariate setting.) This is relevant today in many mendelian randomization (MR) analyses that have a natural desire to take advantage of the large databases readily available. In that vein, it is also of interest to further consider bounds obtained using IVs with three levels. Outside of Ramsahai (2012), most work has been done using binary IVs in the trivariate setting.

Width of Bivariate Bounds

One nice result for bounds obtained from trivariate data tells us that the width will never exceed one. The same is not the case for bounds from bivariate data. We will see that even under relatively strong assumptions, the width of the bounds can only be guaranteed to be less than one only if the IV is strong, and in particular much, much stronger than what is observed in a typical MR analysis.

Improving Bounds from Bivariate Data

Since relatively large databases with bivariate data are already available, it seems worth it to explore if we can utilize these data in any way to improve the bounds we obtain. We will present two ways of doing so, and further look at what happens when those are combined.

Multiple IVs

The simplest way of utilizing multiple IVs is by simply taking intersections of the individual intervals. We will see that in most scenarios, the width of an intersection interval is mainly driven by the strongest IV. In fact, if monotonicity is assumed, it can be shown that the lower and upper bounds are monotonically increasing and decreasing, respectively, as a function of the strength of the IV, all else being equal. This means that the intersection will be exactly the bounds of the strongest IV. We show through simulations that even when all else is not equal, the gain in the width of the intersection is miniscule compared to the best individual bounds.

Possible Trivariate Distributions

Since we know that the bounds obtained from trivariate data will always have length less than 1, and generally be narrower than the bivariate bounds, one could ask what distributions of $(X, Y|Z)$ would be possible given we know the distributions of $(X|Z)$, $(Y|Z)$, and assuming the IV model holds. Doing so, one can arrive at what can be thought of as a posterior distribution of the trivariate bounds. [**Note: prior on Cov$(X, Y|Z = z) \sim$ Uniform, likelihood on $P(X = 1|Z = z), P(Y = 1|Z = z)$ point masses would imply this interpretation, no????**] The results of doing so enables us to draw a few different conclusions depending on the specific scenario:

- knowing the full trivariate distribution provides no further insights over the bivariate bounds
- the full trivariate distribution might be very likely to tell us the direction of the effect
- the full trivariate distribution might be very unlikely to tell us the direction of the effect

Possible Trivariate Distributions from Multiple IVs

Finally, we can combine the two before mentioned approaches: from multiple IVs, randomly choose trivariate distributions of $(X, Y|Z_j)$ that fit the bill, get the bounds, and find the intersections. [**Preliminary results: some gain, but uncertain if it will help with direction.**]

Results/Conclusions

- Bounds based on bivariate data are not very useful – we need very, very strong IVs to be guaranteed any insights.
- Even when utilizing multiple IVs, bivariate data simply does not provide enough information to be consistently useful.
- It seems the best use of bivariate data is to get a sense of what the underlying trivarite distribution might look like.

Non-parametric bounds from bivariate data sources does not seem to promising. It seems that more assumptions are needed if we wish to use non-parametric bounds from bivariate data, or we have to move toward a probabilistic approach.

References

Balke, Alexander, and Judea Pearl. 1993. "Nonparametric Bounds on Causal Effects from Partial Compliance Data." JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION.

Ramsahai, Roland R. 2012. "Causal Bounds and Observable Constraints for Non-Deterministic Models." *J. Mach. Learn. Res.* 13 (March): 829–48.

Swanson, Sonja A., Miguel A. Hernán, Matthew Miller, James M. Robins, and Thomas S. Richardson. 2018. "Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes." *Journal of the American Statistical Association* 113 (522): 933–47. https://doi.org/10.1080/01621459.2018.1434530.