Old and new approaches for the analysis of categorical data in a SEM framework

Yves Rosseel¹ & Myrsini Katsikatsou²

¹Ghent University, ²London School of Economics

In traditional software for structural equation modeling (SEM), there are two options to deal with binary and ordinal observed variables. The first option is called the three-stage limited information approach (in the Mplus world, this known as estimator WLSMV). The second option is based on the IRT tradition where full information maximum likelihood estimation is the golden standard. Recently, a new approach called pairwise likelihood (PL) estimation has been introduced in the literature. Pairwise likelihood keeps the computational complexity low regardless of the model size (i.e. the number of observed or latent variables) and at the same time shares some of the desired properties of likelihood methods (i.e. the derived estimator is asymptotically unbiased, consistent and normally distributed). In addition, many familiar inference tools can be extended to the case of PL: Wald test for the parameter estimates, pairwise likelihood ratio test (PLRT) for overall fit, nested models, and equality constraints, and PL versions of AIC and BIC criteria. In this talk, I will briefly discuss the three approaches, and their implementation in SEM software. The PL estimation approach is currently only implemented in the R package lavaan, and serves as an example of how open-source software can foster the development of new statistical ideas.