Bayesian Dynamic EFA with Hidden Markov Models

Abstract

A dynamic latent class structural equation model (DLC-SEM) provides a comprehensive framework for intensive longitudinal data by accounting for latent class memberships, intra-individual changes, inter-individual differences, and time-specific effects. In order to correctly estimate a person's latent class membership at each time point and to detect latent class switches, it is necessary to specify state-specific measurement models that best describe the data at hand. However, in some cases, factor structures for psychological constructs (e.g., Hamilton Rating Scale for Depression (HAMD)) are unclear across studies or vary depending on the samples being tested. To address the unstable nature of factor structure for psychological data, exploratory factor analysis (EFA) was used as a measurement model within a latent state that follows the Hidden Markov Model process (HMM). Research on the exploratory factor analytic version of DSEM with Hidden Markov models has not been explored much yet, and Bayesian estimation of the model has also not been extensively applied. Therefore, this study explores models under different conditions through simulation studies in a Bayesian framework and evaluates the performance of the models. As an empirical example, the application of the model to the Hamilton Rating Scale for Depression is illustrated.

Method

To investigate the performance of models under varying conditions, data were randomly generated from two combined factor models which contain a two-factor model in state 1 and a three-factor model in state 2, with the pre-specified model switch probability. Two models have the same factor structure where the latent factors have person-specific random intercepts, and each latent factor follows AR(1) structure with cross-lagged effects set to zero. The data consists of 9 items for 100 samples measured 15 times. Measurement models were defined in each state in the same way as the data was generated. However, cross-loadings were included in the models to benefit from EFA.

Before evaluating the performance of the switching model, the model has to be verified how well it recovers the factor loading patterns pre-specified in the data generation process. That is, the model has to identify which loadings should be zero and which loadings should be non-zero. This can be achieved by specifying a shrinkage prior for cross-loadings. The model was tested for three shrinkage priors (Bayesian lasso, Horseshoe, Spike-and-Slab), of which the Horseshoe prior and Spike-and-Slab prior performed best.

Following the model definition step, the model was tested under different settings varying with respect to (a) sample size (100 or 200), (b) MCMC chain iterations (50,000 or 25,000), and (c) MCMC thinning (1 or 2), resulting in eight models. Finally, model performance on dimensionality changes was evaluated via specificity and sensitivity between the estimated switching time point and the true switching time point.

The empirical Hamilton Rating Scale for Depression data contains 24 items for 220 patients and was collected weekly for 13 weeks. The dynamic EFA with HMM used in the simulation study was applied again. Considering the unclear factor structure of HAMD, EFA with target rotation was performed to obtain the initial factor structure. The result suggested six- and three-factor models for each state, allowing for cross-loadings.

The code for all the models used in this study was written in JAGS which performs Markov Chain Monte Carlo (MCMC) sampling, and was implemented using the "R2jags" package (Version 0.8-9) in R software (Version 4.4.2).