Variational approximations for structural equation models

Luca Maestrini[1,2] and **Khue-Dung Dang**[3]

[1]University of Technology Sydney ✉ *luca.maestrini@uts.edu.au*
[2]Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers
[3]University of Melbourne

Research Objective

We propose fast and accurate variational approximations for fitting structural equation models and making Bayesian inference.

Structural Equation Models

Structural equation models (SEMs) are commonly used in social sciences to study and test the relationship between sets of observed and unobservable variables.

As an example, consider a social study aimed to evaluate human intelligence:
- human intelligence cannot be measured directly (latent variable);
- psychologists can develop a hypothesis of intelligence and design a test with questions (observable variables) to measure intelligence according to the hypothesis;
- a SEM can be used to test the hypothesis based on data from people who took the intelligence test.

SEMs are typically studied through linear models

$$y = \Lambda\eta + \varepsilon,$$

where y and η are vectors of observations and latent variables, respectively, and Λ is a matrix of factor loadings.

Statistical Approaches for SEMs

Frequentist approaches:
- software for maximum likelihood and weighted least squares available;
- the asymptotic properties of the statistics are dependent on the sample covariance matrix being asymptotically normal;
- may suffer from computational and theoretical problems for small sample sizes or non-normal data.

Bayesian approaches:
- allow incorporation of prior information;
- facilitate the adoption of more flexible model structures, such as those with cross-loadings and non-normal errors;
- may suffer from slow convergence and long running times, compared to frequentist approaches.

Remedy: variational approximations.

Mean Field Variational Bayes

For Bayesian statistical models with observed data D and parameter vector θ:
- a mean field variational approximation $q^*(\theta)$ to $p(\theta|D)$ is the minimizer of the Kullback–Leibler divergence

$$\int q(\theta)\log\left\{\frac{q(\theta)}{p(\theta|D)}\right\}d\theta$$

  s.t. a restriction $q(\theta) = \prod_{i=1}^{M} q(\theta_i)$, for a $\{\theta_1,\ldots,\theta_M\}$ partition of θ;
- the optimal q-density functions satisfy

$$q^*(\theta_i) \propto \exp\left\{E_{q(\theta\backslash\theta_i)}\log p(\theta_i|D,\theta\backslash\theta_i)\right\}, \quad 1 \le i \le M,$$

  where $\theta\backslash\theta_i$ denotes the entries of θ with θ_i omitted;
- this gives rise to an iterative optimization scheme known as **mean field variational Bayes (MFVB)**.

Bootstrap

MFVB underestimates the variance of the posterior density and may produce biased parameter estimates.

We adopt an empirical bootstrap strategy to overcome these issues:
- sample with replacement from the original dataset;
- recompute the variational estimator for each bootstrap sample;
- use the distribution of these bootstrapped estimators, or related quantities, to derive uncertainty measures.

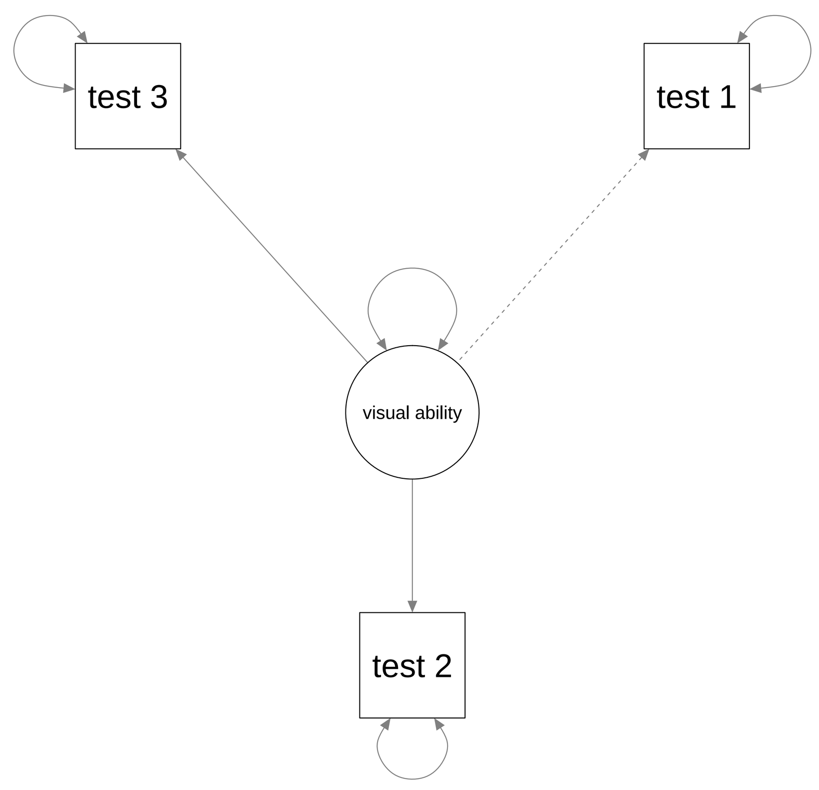
Data Study

We consider the following model (plus priors):

$$y_i\,|\,\nu,\lambda,\eta_i,\psi \overset{\text{ind.}}{\sim} N(\nu+\lambda\eta_i,\text{diag}(\psi)),\quad \eta_i\,|\,\sigma^2 \overset{\text{ind.}}{\sim} N(0,\sigma^2),\quad i=1,\ldots,n,$$

$$\lambda_j\,|\,\psi_j \overset{\text{ind.}}{\sim} N(\mu_\lambda,\sigma^2_\lambda\psi_j),\quad j=1,\ldots,m,$$

where y_i is a vector of m observed outcomes for individual i from a group of n individuals. We use this model to study the Holzinger & Swineford (1939) data, of which we consider $m=3$ outcomes (tests) to assess visual ability.



Results

Computational times to fit the model:
- MFVB converges in 0.1 seconds;
- MCMC in rstan takes around 5 minutes.

We also construct confidence intervals using bootstrap with:
- the percentile method (based on percentiles of the distribution of the bootstrap variational estimators);
- the (studentized) pivotal method (which requires a consistent estimator of the variance of bootstrap estimator).

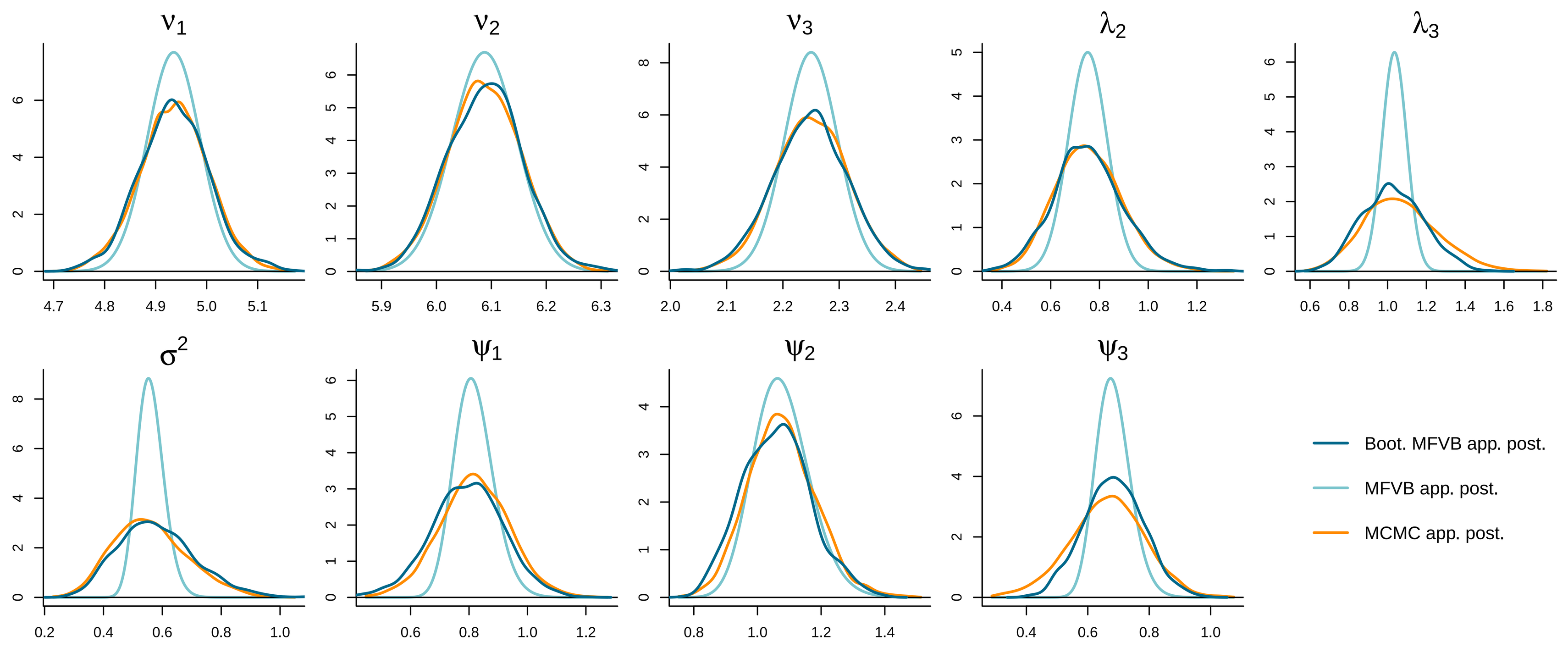


Figure 1:*Approximate marginal posterior densities of the parameters of interest obtained via MFVB (light blue), MCMC (orange) and density estimates produced using MFVB point estimates from 1,000 bootstrap samples (dark blue curves).*

	ν_1	ν_2	ν_3	λ_2	λ_3	σ^2	ψ_1	ψ_2	ψ_3
MFVB	0.86	0.92	0.82	0.76	0.56	0.52	0.72	0.91	0.61
MFVB with jackknife	0.96	0.96	0.92	0.95	0.92	0.95	0.92	0.96	0.92
MFVB with per. boot. (100)	0.92	0.91	0.95	0.95	0.92	0.92	0.86	0.94	0.93
MFVB with piv. boot. (100)	0.97	0.95	0.97	0.98	0.94	0.98	0.97	0.98	0.98
MFVB with per. boot. (500)	0.95	0.94	0.91	0.96	0.92	0.91	0.90	0.95	0.94
MFVB with piv. boot. (500)	0.98	0.98	0.96	0.97	0.93	0.96	0.99	0.97	0.99
MFVB with per. boot. (1,000)	0.95	0.94	0.93	0.97	0.95	0.92	0.89	0.96	0.94
MFVB with piv. boot. (1,000)	0.99	0.97	0.97	0.98	0.94	0.96	1.00	0.98	0.98
MCMC	0.96	0.96	0.93	0.93	0.94	0.96	0.93	0.93	0.97

Table 1:Average empirical coverage percentages for advertized 95% credible intervals of the parameters of interest from a simulation study based on 100 replications. For the percentile and pivotal bootstrap results, 100, 500 and 1,000 bootstrap iterations were used.

Article

Dang, K.D. and Maestrini, L. (2021). Fitting Structural Equation Models via Variational Approximations. arXiv preprint arXiv:2105.15036.