Factor Analysis for Hierarchical Models in Psychometrics

Jason Farris
AMS 225
Brief Motivation

• Ed. meas. intuitively leads to factor analysis:
  
  − Achievement is naturally multidimensional in latent factors
  
  − Multiple, possibly correlated, latent constructs valuable source of information on test takers and test items.
  
  − Observed achievement in a particular subject due to one or more underlying latent abilities \(\rightarrow\) reduce scoring array into subscores on those latent abilities.

• Grouping into a hierarchy is common (classrooms into schools, schools into districts, etc.)
“Bayesian Factor Analysis for Multilevel Binary Observations”

- Psychometrika---vol. 65, No. 4, 475-496

- Develops MCMC methods for hierarchical binary data.

- Methods not limited to educational measurement, applications to personality or attitudes tests.

- Hierarchical structure used to model heterogeneity across population groups.

- Methods applied to mathematics achievement test.
Model

- Hierarchical model for underlying continuous variables (abilities):

  \[ y_{ijk} = \begin{cases} 
  1 & \text{if } w_{ijk} > 0 \\ 
  0 & \text{otherwise,} 
  \end{cases} \quad \text{for } k = 1 \text{ to } p. \]

  \[ w_{ij} \sim N_p(\tau_i, V_1), \quad \text{(iid)} \]

  \[ \tau_i \sim N_p(\nu, V_2), \quad \text{(iid)} \]

  \[ V_1 = \Omega_1 \Psi_1 \Omega_1^\prime + \Delta_1 \]

  \[ V_2 = \Omega_2 \Psi_2 \Omega_2^\prime + \Delta_2. \]

- The \( \Omega \) matrices are factor loadings, the \( \Psi \)'s are variance/covariance matrices, \( \Delta \)'s are diagonal variance matrices at both the first level and the group mean level.
Priors

• Due to scaling and other restrictions, the parameters of the model are:

\[ \gamma = \{\mu, \Lambda_2, \Theta_2, \Psi_2, \Lambda_1, \Psi_1\} \]

• Where \( \Lambda_2 = T\Omega_2, \mu = Tv \) and \( \Theta_2 = T\Delta_2T \). And \( \Lambda_1 = T\Omega_1 \)

• \( T \) is fixed to be a diagonal matrix such that \( TV_1T \) reduces to a correlation matrix

• Proper, but diffuse priors were chosen on all the parameters of the model
Priors Continued

• Priors were chosen as independent priors on all the parameters
• Prior on overall mean is MVN with mean $\eta=0$ and diagonal variance matrix $\mathbf{C}$ with large variances to represent vague knowledge.
• Inverse Gamma priors on the diagonal $\Theta_2$ matrix
• Diffuse MVN prior over the free parameters of correlation matrices $\Psi_2$ and $\Psi_1$
• Diffuse MVN prior over the non-zero elements within each row of the matrix $\Delta_2$ leading to a product of independent priors for $\Delta_2$
• Same method for prior on $\Delta_1$
Inference, Model Checking and Comparison

- Gibbs and M-H steps in tandem with data augmentation are used to sample from the posterior distribution for the parameters.

- Group means and underlying ability estimates are sampled in addition to factor scores.

- Posterior predictive p-values are used to detect model fit.

- Pseudo-Bayes factors are used for model comparison and is easily computed from MCMC draws.
Application

- Mathematic achievement data from Second International Mathematics Study: 12 items from four sub areas of mathematics

- Administered to 8th grade students in the U.S.
  - 274 classes and 5601 students: hierarchical structure captures differences between classroom instruction

- Objective is to model covariation in student achievement on questions from four math subareas: arithmetic, algebra, measurement and geometry.

- One underlying latent “math” ability or one latent ability for each subarea?
Application Continued

• MCMC procedure is applied to estimate two models
  – M1 assumes one underlying math ability
  – M2 assumes four, possibly correlated, underlying abilities

• Diffuse prior information used in each model

• Posterior predictive p-values were calculated on the correlations between all the items: M2 does a better job of recovering the correlations between the test items.

• The pseudo-Bayes factor supports M2 over M1.
Application Continued/Extensions

• The second level factor loadings support accounting for cross-classroom differences

• The second level factors are positively correlated suggesting further unobserved factors common the four modeled factors.

• Extend to include non-binary data, and more than 2 levels and regressors at each level of the hierarchy.
“A Multivariate Multilevel Approach to the Modeling of Accuracy and Speed of Test Takers”

• Accuracy and speed of test takers is modeled on binary responses and continuous response times to items.

• Multivariate multilevel model accommodates covariates to explain the variance in speed and accuracy between individuals and groups of test takers.

• MCMC methods developed to estimate all parameters using Gibbs and M-H steps.

• Bayes factors and DIC are computed as model checking/selection criterion.
Model

• Measurement models used for responses and response times
  – IRT model used to estimate student ability $\theta_{ij}$ of test taker i in group j and the item discrimination, difficulty and guessing parameters.
  – The response time model is $t_{ijk} = -\phi_k \zeta_{ij} + \lambda_k + \epsilon_{ijk}$, where the last term is distributed as $N(0, a_k^2)$ and $\zeta_{ij}$ is the speed at which a person works
• More interested in person parameters ($\theta_{ij}, \zeta_{ij}$) than item parameters
• Model ($\theta_{ij}, \zeta_{ij}$) as a two level hierarchical model with regressors at each level of the hierarchy and non-zero correlations between parameters.
Model Continued

• The item parameters (discrimination, difficulty, time intensity, time discrimination) are assumed to follow a MVN:
  – mean vector $\mu_i$ and covariance $\Sigma_i$

• Covariance structure allows for correlations between item parameters: easy items take less time, and hard items take more time.

• Guessing parameter is dropped since it has no analogous parameter in the response time measurement model.
Inference

• Data augmentation is performed to allow for fully Gibbs MCMC implementation
  – $s_{ijk} = 1$ when person i in group j knows correct answer to question k, and is zero otherwise
  – continuous latent responses ($z$) are defined in terms of the discrimination and difficulty item parameters and ability estimate $\theta$ for person i in group j along with an associate normally distributed error term.

• Normal inverse-Wishart priors are used on the MVN model for the item parameters and the two level hierarchical model for the person parameters and conjugate priors are used elsewhere.
Model Selection

- DIC and Bayes factors were used for model selection criterion

- Simulated data was used to test prior sensitivity and properties of the Bayes factors and DIC criterion.

- Simulations suggest models with both the covariate and group structure for test takers was optimal
Empirical Example

• 388 test takers were administered 65 items on a computer-based version of the *Natural World Assessment Test* (NAW-8)

• Required of sophomore college students at a particular university to assess quantitative and scientific reasoning proficiency

• Covariates: SAT score, gender, self-reported measure of citizenship and self-reported measure of test effort.
Empirical Example Continued

• Vague prior knowledge was specified

• DIC suggested the full model was optimal and posterior predictive checks showed good model fit

• Strong negative correlation between person parameters suggests high ability students are better at time management during test taking.

• Also strong negative correlation suggests that test was not speeded: higher ability students who took their time were not penalized.
Empirical Example Continued

- The hypothesis that gender and citizenship had no effect on ability and speed was confirmed.
- However, SAT scores and test effort explained a significant amount of the variation between the person parameters.
- A positive relationship between test effort and ability is intuitive since students who did not care about their results would normally score lower than students with high test effort measures (and consequently spend less time on each item).
- The positive relationship between SAT scores and ability was expected, but there was no similar relationship with speed.
Extensions

• Extend to non-binary response data using MCMC methods developed for such items.

• Extend to mixture models to model test takers utilizing different test taking strategies.

• Implement hierarchical model for the item parameters to adjust for speed differing across different groups of items.