Toward Large-Scale Item Response Theory Factor Analysis of Polytomous Data

Eugene Geis & Gregory Camilli

A Little History on Gibbs IRT

- Albert (1992) suggests augmented Gibbs approach to 2PNO
- Meng & Schilling (1996); Beguin & Glas (2001) extend Albert's approach to MIRT
- Gu & Kong (1998) & Cai (2010) implement stochastic NR on MH-RM

The Probit MIRT Model

- Simplifying the tails of the probability distribution of the 2PL

$$P_{ij}\left(Y_{ij}=1\right)=\Phi\left(Y_{ij}=1 \mid \eta_{ij}\right)$$

$$\eta_{ij}=\mathbf{A}_j\boldsymbol{\theta}_i-b_j,$$

Augmented Data

$$z_{ij} = \mathbf{A}_j \boldsymbol{\theta}_i - b_j + \varepsilon_{ij}$$
$$= \eta_{ij} + \varepsilon_{ij}.$$

- Reducing Multivariate Latent Factors to Univariate Random Variables Conditional on model parameters
- Dichotomous:
  - single truncated normal
  - $\mathbf{b} = -\bar{\mathbf{z}}, \text{ given } \mathrm{E}\left[\sum_i \boldsymbol{\theta}_i\right] = 0$
- Polytomous:
  - double truncated normal
  - augmented data sampled at each threshold

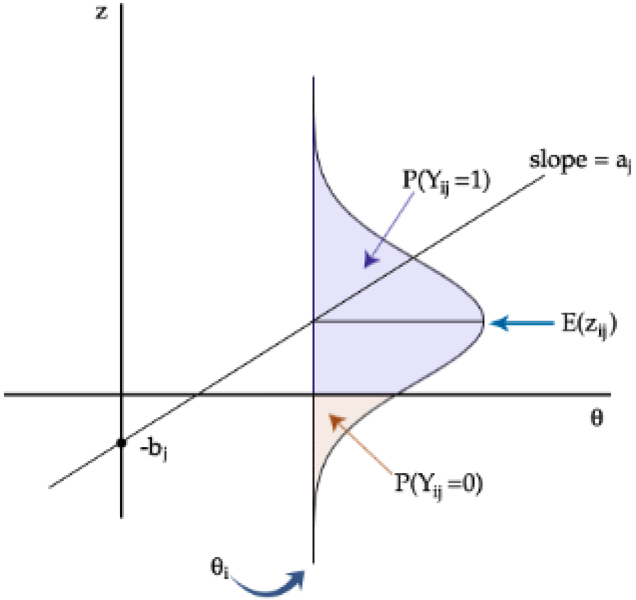

FIGURE 1. *Marginal normal distribution at θ_i.*

Rao-Blackwellization

- Sufficient Statistics simplify the maximization steps
- Dimensionality is embedded within covariance of latent factors and augmented data

$$\sum_i \boldsymbol{\theta}_i \,, \boxed{\sum_i \mathbf{z}_i \,, \sum_i \mathbf{z}_i \mathbf{z}_i^T} \,, \sum_i \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T \,, \; \sum_i \mathbf{z}_i \boldsymbol{\theta}_i^T \,.$$

One Gibbs Cycle

1)

$$x_{ijk} \mid \eta_{ij}, \tau_{jk} \sim \begin{cases} N_{(-\infty,0)}\left(\eta_{ij} - \tau_{jk}, 1\right) & \text{if } y_{ij} \le k \\ N_{(0,+\infty)}\left(\eta_{ij} - \tau_{jk}, 1\right) & \text{if } y_{ij} > k. \end{cases}$$

2)

$$z_{ij} \mid \eta_{ijk} \sim \begin{cases} TN_{(d_{jL},\infty)}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = L = \kappa - 1 \\ \dots & \\ TN_{(d_{j1},d_{j2})}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = 2 \\ TN_{(-\infty,d_{j1})}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = 1 \end{cases}$$

3)

$$\boxed{\begin{aligned} \boldsymbol{\mu}_{x,jk} &= \boldsymbol{\tau}_{jk}^{(t-1)} + \gamma_t \left\{ \bar{\mathbf{x}}_{jk} - \boldsymbol{\mu}_{x,jk}^{(t-1)} \right\} \\ \boldsymbol{\Sigma}_{zz} &= \boldsymbol{\Sigma}_{zz}^{(t-1)} + \gamma_t \left(\mathbf{S}_{zz} - \boldsymbol{\Sigma}_{zz}^{(t-1)} \right) \end{aligned}}$$

4)

$$\hat{\boldsymbol{\theta}} \sim \left(\mathbf{I} + \mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \left(\mathbf{z}_i + \mathbf{b}\right)$$

$$\boldsymbol{\theta} \sim N\left(\hat{\boldsymbol{\theta}}, \left(\mathbf{I} + \mathbf{A}^T \mathbf{A}\right)^{-1}\right),$$

Robbins-Monro "Squeeze"

$$s^{(t)} = s^{(t-1)} + \gamma_t \left\{ S\left(y, \psi^{(t)} \right) - s^{(t-1)} \right\}$$

- γ_t is a positive sequence such that $\gamma_t = 1, \sum_{t=1}^{\infty} \gamma_t \to \infty$, and $\sum_{t=1}^{\infty} \gamma_t^2$ is finite.

- Typically: $\gamma = \{1, 1/2, 1/3, 1/4, \ldots\}^{\alpha}$ where $0.5 < \alpha \leq 1$

Computational Efficiencies using R

- Vectorization
- Compiled C++
  - Example: **apply(M, 2, “mean”)** vs. ***colMeans(M)***
- Parallelization
  - Embarrassingly
    - Windows (SNOW), Linux (parallel)
    - Conditional Independence allows choice, *J* or *N*

Demo of Github Code

- https://github.com/genobobeno/SAEM_IRT