Practical Guidelines for the Estimation of Multidimensional Ordered Polytomous Models

Presentation at NCME

Lian Hortensius and Chun Wang

University of Minnesota

April 4th, 2014

Outline

Outline

Multidimensional polytomous IRT

- Item response theory (IRT): model responses of people to items as a function of (an) underlying trait(s)
- Multidimensional IRT: several underlying traits
- Polytomous IRT: more than two possible response categories
- Examples: partial credit items, rating scale items, for surveys/educational tests

Multidimensional polytomous IRT

Common models:

- multidimensional generalized partial credit model (MGPCM)
- multidimensional graded response model (MGRM)

In unidimensional case: very similar

MGPCM

Item response function:

$$P(u_{ij} = k|\boldsymbol{\theta}_j) = \frac{e^{k\mathbf{a}_i^\intercal \boldsymbol{\theta}_j - \sum_{u=0}^{k} \beta_{iu}}}{\sum_{v=0}^{K_i} e^{v\mathbf{a}_i^\intercal \boldsymbol{\theta}_j - \sum_{u=0}^{k} \beta_{iu}}}$$

- $k = 0, 1, \ldots, K_i$
- Slope parameter for each dimension ($\mathbf{a}_i$)
- $K_i + 1$ threshold parameters (first threshold constrained to be 0)

MGRM

Item response function:

$$P(u_{ij} = k|\boldsymbol{\theta}_j) = P(u_{ij} \geq k) - P(u_{ij} \geq k+1)$$
$$= \frac{1}{1 + e^{-\sum a_{ih}(\theta_h - d_{k,i})}} - \frac{1}{1 + e^{-\sum a_{ih}(\theta_h - d_{k+1,i})}}$$

- $k = 0, 1, \ldots, K_i$
- $d_{k,i}$: ease with which a person will reach the kth step
- $P(u_{ij} \geq 0) = 1$

Problem statement

Several possible software packages:

- flexMIRT
- bmirtII
- EQSIRT
- IRTPRO
- etc.

Research questions:

- When do MGPCM and MGRM lead to different results?
- Do different software packages lead to different results?

Outline

flexMIRT

- Uses Expectation-Maximization (EM) algorithm (Bock & Aitken, 1981)
- Able to estimate both MGRM and MGPCM
- 3 academic licenses for $125 / year
- Available θ estimation methods are MLE, EAP, and MAP
- Allows for non-normal latent trait distributions

bmirtII

- Uses Markov Chain Monte Carlo Metropolis Hastings (MCMC MH) algorithm (Metropolis et al., 1953)
- Able to estimate both MGRM and MGPCM
- Software is free
- Available θ estimation methods are MLE, EAP, MAP, and the θs from the MCMC output

Some other available software

- Package mirt in R uses Metropolis-Hastings Robbins-Monroe (MH-RM; Cai, 2010) for MGRM; free
- Software IRTPRO uses EM-algorithm, adaptive quadrature, or MH-RM for MGRM and MGPCM; $495 for unlimited use, $130/year, $75 / 6 months
- Software EQSIRT can do MGRM and MGPCM; $695 corporate license / $595 academic license for unlimited use, $300 / year (slightly cheaper for Mac and Linux)

Outline

Design

- Two pieces of software: flexMIRT and bmirtII
- Two true models: MGRM and MGPCM
- Two estimating models: MGRM and MGPCM
- Two sample sizes (N): 250 and 1000
- Two item pool sizes (J): 30 and 60

Design

- $\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}\right)$
- Factor 1: First third & last third of items
- Factor 2: Second third & last third of items
- 10 iterations / condition $\rightarrow$ 80 datasets $\rightarrow$ 360 analyses, results averaged across iterations

Output

- $\boldsymbol{\theta}_{true}$ vs. $\boldsymbol{\theta}_{est}$ correlations and Mean Square Error (MSE)
- $\theta_{est,1}$ and $\theta_{est,2}$ correlation
- True vs. estimated item parameter correlations and MSE (for matching models)

Outline

MGRM true model, estimating as MGRM

Table: Flexmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.96	.96	.48	.07	.08
	60	.98	.98	.49	.04	.04
1000	30	.96	.96	.50	.08	.08
	60	.98	.98	.50	.04	.04

Table: Bmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.96	.96	.44	.08	.09
	60	.98	.98	.48	.06	.06
1000	30	.96	.96	.45	.08	.08
	60	.98	.98	.48	.06	.06

MGRM true model, estimating as MGPCM

Table: Flexmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.96	.96	.48	.08	.09
	60	.98	.98	.49	.04	.04
1000	30	.96	.96	.50	.08	.08
	60	.98	.98	.50	.04	.04

Table: Bmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.96	.96	.43	.08	.09
	60	.98	.98	.47	.05	.05
1000	30	.96	.96	.44	.08	.08
	60	.98	.98	.46	.05	.04

MGPCM true model, estimating as MGPCM

Table: Flexmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.97	.97	.51	.06	.06
	60	.98	.99	.51	.03	.03
1000	30	.97	.97	.50	.06	.06
	60	.98	.98	.50	.04	.03

Table: Bmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.97	.97	.47	.07	.07
	60	.98	.99	.49	.04	.04
1000	30	.97	.97	.46	.06	.06
	60	.98	.98	.48	.04	.04

MGPCM true model, estimating as MGRM

Table: Flexmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.97	.97	.50	.06	.06
	60	.98	.99	.51	.03	.03
1000	30	.97	.97	.50	.06	.06
	60	.98	.98	.50	.03	.03

Table: Bmirt results: θ

N	J	cor(θ_1)	$cor(\theta_2)$	cov(θ_1, θ_2)	θ_1 MSE	θ_2 MSE
250	30	.97	.97	.48	.09	.08
	60	.98	.99	.51	.06	.07
1000	30	.97	.97	.47	.07	.07
	60	.98	.98	.49	.06	.06

MGRM true model, estimating as MGRM

Table: Flexmirt results: item parameter correlations

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.978	.980	.969	.991	.969
	60	NA	NA	NA	NA	NA
1000	30	.994	.996	.994	.998	.993
	60	.996	.996	.993	.997	.992

Table: Bmirt results: item parameter correlations

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.974	.975	.959	.987	.962
	60	.973	.975	.965	.986	.967
1000	30	.991	.991	.992	.997	.991
	60	.992	.993	.991	.996	.990

MGPCM true model, estimating as MGPCM

Table: Flexmirt results: item parameter correlations

N	J	a_1	a_2
250	30	.977	.974
	60	NA	NA
1000	30	.994	.995
	60	.995	.995

Table: Bmirt results: item parameter correlations

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.978	.976	.961	.982	.955
	60	.977	.980	.961	.983	.956
1000	30	.993	.993	.987	.995	.989
	60	.994	.994	.990	.994	.987

Outline

MGRM vs. MGPCM

- Both estimating models performed well regardless of true model
- Results were slightly better for MGPCM as true model, which practitioners have no control over
- In both cases a larger item bank led to better results

flexMIRT vs. bmirtII in terms of results

- $\boldsymbol{\theta}$ correlations: equal performance
- $\boldsymbol{\theta}$ MSE: flexMIRT somewhat better
- MGRM item parameters: flexMIRT somewhat better both in terms of correlations and MSE
- MGPCM a-parameters: flexMIRT somewhat better for high N, bmirtII somewhat better for low N

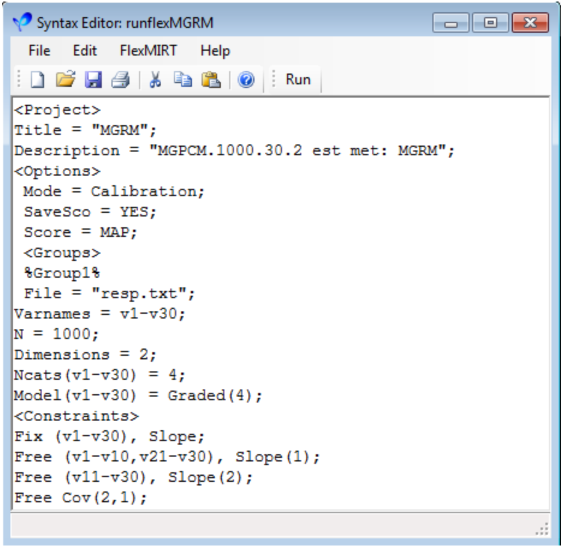
flexMIRT vs. bmirtII

- flexMIRT costs $125 / year for an academic license, bmirtII is free
- Support for flexMIRT was great, I did not contact bmirtII for support
- Organization of bmirtII is much less intuitive, installation and figuring out which files to use is confusing
- bmirtII output files were much easier to read into R, and parametrization mostly corresponded to mine

Guidelines for using flexMIRT

- If there is a problem flexMIRT does not always report clear errors
- e.g. when a category of an item is not selected by any person, the code will not run fully
- flexMIRT support team is very helpful and fast
- Two input files:
  - A response/data file, each line is a person, each column an item (can be .dat, .txt, ...)
  - A .flexmirt file with clear parts to fill out, e.g. Varnames, N, Dimensions, etc.

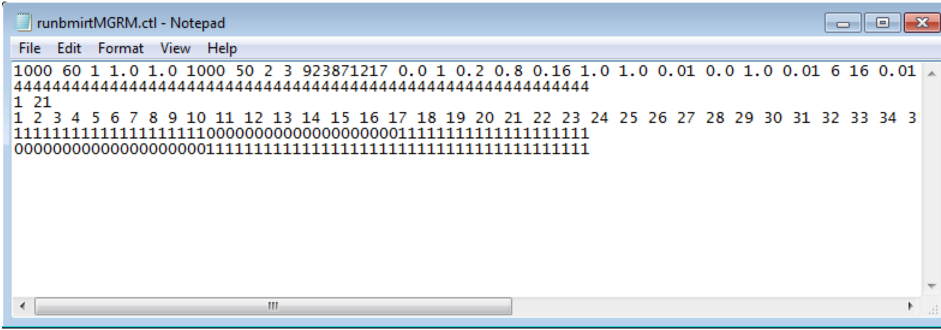
Guidelines for using flexMIRT



Guidelines for using bmirtII

- Pay attention to number of iterations - check convergence criteria
- Use BMIRT28.bat for MGPCM and BMIRTGradeResponse.bat for MGRM
- Three input files:
  - .rwo
  - .ctl
  - .bat is simply a file to call .rwo, .ctl, and the correct .bat file, and to specify where to print output
- Remember to always put the entire lib folder in the folder where you are running bmirt!

Guidelines for using bmirtII



Thank you

For questions, contact me at horte005@umn.edu.

MGRM true model, estimating as MGRM

Table: Flexmirt results: item parameter MSE

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.038	.037	.021	.008	.023
	60	NA	NA	NA	NA	NA
1000	30	.009	.007	.004	.002	.004
	60	.007	.007	.005	.002	.005

Table: Bmirt results: item parameter MSE

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.114	.116	.039	.015	.045
	60	.127	.129	.052	.020	.044
1000	30	.043	.046	.014	.005	.011
	60	.063	.061	.034	.011	.024

MGPCM true model, estimating as MGPCM

Table: Flexmirt results: item parameter MSE

N	J	a_1	a_2
250	30	.052	.047
	60	NA	NA
1000	30	.011	.009
	60	.008	.009

Table: Bmirt results: item parameter MSE

N	J	a_1	a_2	d_1	d_2	d_3
250	30	.044	.041	.282	.149	.308
	60	.041	.035	.239	.128	.327
1000	30	.031	.023	.108	.046	.084
	60	.028	.028	.089	.058	.105