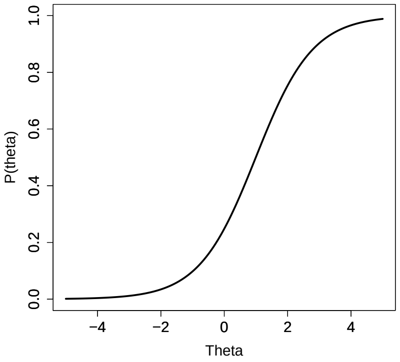
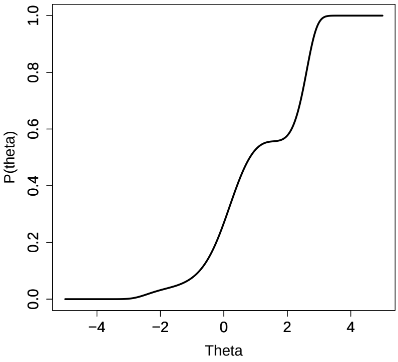
A logistic function of a monotonic polynomial for estimating item response functions

Carl F. Falk and Li Cai

University of California, Los Angeles

IMPS 2013





2PL ??

- Possible consequences[1]
  - Item response function (IRF) does not follow 2PL
  - IRF recovery not good
  - Latent trait estimates not good

[1] (e.g., Liang, 2007; Ramsay & Abrahamowicz, 1989)

Possible Solutions

- Estimate entire latent distribution as non-normal
  - Empirical histogram approach[1]
  - Ramsay Curves[2]
- Non-parametric modelling IRF
  - Kernel smoothing[3]
  - Use of external test with known IRFs[4]
- Semi-parametric modelling of IRF
  - Bayesian approaches [5]
  - Regression splines[6]
  - Monotonic polynomial (Liang, 2007)

[1] (e.g., Bock & Aitkin, 1981)
[2] (e.g., Woods & Thissen, 2006)
[3] (Ramsay, 1991; Rossi, Wang, & Ramsay, 2002)
[4] (Samejima, 1977, 1979)
[5] (Miyazaki & Hoshino, 2009; Qin, 1998)
[6] (Ramsay & Winsberg, 1991)

Possible Solutions

- Estimate entire latent distribution as non-normal
  - Empirical histogram approach[1]
  - Ramsay Curves[2]
- Non-parametric modelling IRF
  - Kernel smoothing[3]
  - Use of external test with known IRFs[4]
- Semi-parametric modelling of IRF
  - Bayesian approaches [5]
  - Regression splines[6]
  - Monotonic polynomial (Liang, 2007)

[1] (e.g., Bock & Aitkin, 1981)
[2] (e.g., Woods & Thissen, 2006)
[3] (Ramsay, 1991; Rossi, Wang, & Ramsay, 2002)
[4] (Samejima, 1977, 1979)
[5] (Miyazaki & Hoshino, 2009; Qin, 1998)
[6] (Ramsay & Winsberg, 1991)

Possible Solutions

- Estimate entire latent distribution as non-normal
  - Empirical histogram approach[1]
  - Ramsay Curves[2]
- Non-parametric modelling IRF
  - Kernel smoothing[3]
  - Use of external test with known IRFs[4]
- Semi-parametric modelling of IRF
  - Bayesian approaches [5]
  - Regression splines[6]
  - Monotonic polynomial (Liang, 2007)

[1] (e.g., Bock & Aitkin, 1981)
[2] (e.g., Woods & Thissen, 2006)
[3] (Ramsay, 1991; Rossi, Wang, & Ramsay, 2002)
[4] (Samejima, 1977, 1979)
[5] (Miyazaki & Hoshino, 2009; Qin, 1998)
[6] (Ramsay & Winsberg, 1991)

Possible Solutions

- Estimate entire latent distribution as non-normal
  - Empirical histogram approach[1]
  - Ramsay Curves[2]
- Non-parametric modelling IRF
  - Kernel smoothing[3]
  - Use of external test with known IRFs[4]
- Semi-parametric modelling of IRF
  - Bayesian approaches [5]
  - Regression splines[6]
  - Monotonic polynomial (Liang, 2007)

[1] (e.g., Bock & Aitkin, 1981)
[2] (e.g., Woods & Thissen, 2006)
[3] (Ramsay, 1991; Rossi, Wang, & Ramsay, 2002)
[4] (Samejima, 1977, 1979)
[5] (Miyazaki & Hoshino, 2009; Qin, 1998)
[6] (Ramsay & Winsberg, 1991)

Monotonic Polynomial Approach

- Liang (2007)
  - Applied to only 2PL (i.e., dichotomous items)
  - Uses 2-stage / surrogate-based (SB) estimation approach
    - Use first principal component to estimate θ
    - Use provisional θ to estimate model parameters using complete-data likelihood
- Our Approach
  - Model 2PL, 3PL, and polytomous items
  - Bock-Aitkin Marginal Maximum Likelihood (EM-MML)[1]

[1] (Bock & Aitkin, 1981)

Monotonic Polynomial

A monotonic polynomial:

$$m_i(\theta) = \xi_i + b_{1i}\theta + b_{2i}\theta^2 + \cdots + b_{2k+1,i}\theta^{2k+1}$$

And its derivative:

$$m_i'(\theta) = a_{0i} + a_{1i}\theta + a_{2i}\theta^2 + \cdots + a_{2k,i}\theta^{2k}$$

$k = 0, 1, \ldots$ controls order of polynomial $(2k+1)$.
$i = 1, 2, \ldots n$ is item index.

Monotonic Polynomial

- How to ensure $m_i(\theta)$ is monotonically increasing?
  - Odd order: $2k+1$
  - $m_i'(\theta)$ must be positive across θ
  - Reparameterize and implement constraints[1]

$$m_i'(\theta) = \begin{cases} \lambda_i \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \beta_{ui})\theta^2) & \text{if } k > 0 \\ \lambda_i & \text{if } k = 0 \end{cases}$$

    - All $\beta > 0, \lambda > 0$.
  - Our parameterization

$$m_i'(\theta) = \begin{cases} \exp(\omega_i) \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \exp(\tau_{ui}))\theta^2) & \text{if } k > 0 \\ \exp(\omega_i) & \text{if } k = 0 \end{cases}$$

    - $\ln(\beta) = \tau$, and $\ln(\lambda) = \omega$.

[1] (Elphinstone, 1985)

Monotonic Polynomial

- How to ensure $m_i(\theta)$ is monotonically increasing?
  - Odd order: $2k+1$
  - $m_i'(\theta)$ must be positive across θ
  - Reparameterize and implement constraints[1]

$$m_i'(\theta) = \begin{cases} \lambda_i \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \beta_{ui})\theta^2) & \text{if } k > 0 \\ \lambda_i & \text{if } k = 0 \end{cases}$$

    - All $\beta > 0, \lambda > 0$.
  - Our parameterization

$$m_i'(\theta) = \begin{cases} \exp(\omega_i) \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \exp(\tau_{ui}))\theta^2) & \text{if } k > 0 \\ \exp(\omega_i) & \text{if } k = 0 \end{cases}$$

    - $\ln(\beta) = \tau$, and $\ln(\lambda) = \omega$.

[1] (Elphinstone, 1985)

Monotonic Polynomial

- How to ensure $m_i(\theta)$ is monotonically increasing?
  - Odd order: $2k+1$
  - $m_i'(\theta)$ must be positive across θ
  - Reparameterize and implement constraints[1]

$$m_i'(\theta) = \begin{cases} \lambda_i \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \beta_{ui})\theta^2) & \text{if } k > 0 \\ \lambda_i & \text{if } k = 0 \end{cases}$$

    - All $\beta > 0, \lambda > 0.$
  - Our parameterization

$$m_i'(\theta) = \begin{cases} \exp(\omega_i) \prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \exp(\tau_{ui}))\theta^2) & \text{if } k > 0 \\ \exp(\omega_i) & \text{if } k = 0 \end{cases}$$

    - $\ln(\beta) = \tau$, and $\ln(\lambda) = \omega$.

[1] (Elphinstone, 1985)

Monotonic Polynomial - Matrix Form[1]

$$m'(\theta) = a_0 + a_1\theta + a_2\theta^2 + \cdots + a_{2k}\theta^{2k}$$

$$m_i'(\theta) = \begin{cases} \exp(\omega_i)\prod_{u=1}^{k}(1 - 2\alpha_{ui}\theta + (\alpha_{ui}^2 + \exp(\tau_{ui}))\theta^2) & \text{if } k > 0 \\ \exp(\omega_i) & \text{if } k = 0 \end{cases}$$

$$\mathbf{a}_k = \mathbf{T_k T_{k-1}} \ldots \mathbf{T_2 T_1} \exp(\omega)$$

- $\mathbf{a_k}$ can then be converted to $\mathbf{b_k}$ used to compute $m(\theta)$.
- Each $\mathbf{T_k}$ contains only τ_k and α_k parameters.
    - Useful for taking derivatives of τ and α parameters

[1](Due mostly to Browne, 1997 as cited in Liang, 2007)

Monotonic Polynomial - Matrix Form[1]

$$\mathbf{T_2} = \begin{bmatrix} 1 & 0 & 0 \\ -2\alpha_2 & 1 & 0 \\ \alpha_2^2 + \exp(\tau_2) & -2\alpha_2 & 1 \\ 0 & \alpha_2^2 + \exp(\tau_2) & -2\alpha_2 \\ 0 & 0 & \alpha_2^2 + \exp(\tau_2) \end{bmatrix}$$

[1] (Due mostly to Browne, 1997 as cited in Liang, 2007)

Logistic Function of Monotonic Polynomial (LMP)

- 2PL

$$P(y_{ij} = 1|\theta_j, \delta_i, \gamma_i) = \frac{1}{1 + exp(-(\delta_i + \gamma_i\theta_j))}$$

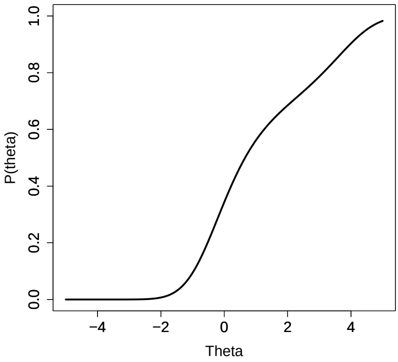
- LMP: Logistic Function of Monotonic Polynomial

$$P(y_{ij} = 1|\theta_j, \xi_i, \omega_i, \boldsymbol{\alpha}_i, \boldsymbol{\tau}_i) = \frac{1}{1 + \exp[-m_i(\theta_j)]}$$

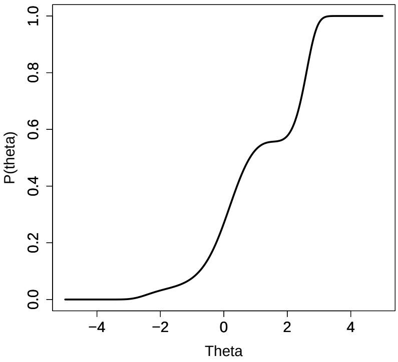
Where $m_i(\theta)$ is the monotonic polynomial (order $2k + 1$)

- $j = 1, 2, \ldots N$ is subject index

Example IRFs (from PISA 2000 Read Book 8)



LMP



LMP

LMP with Asymptote (LMPA)

- 3PL

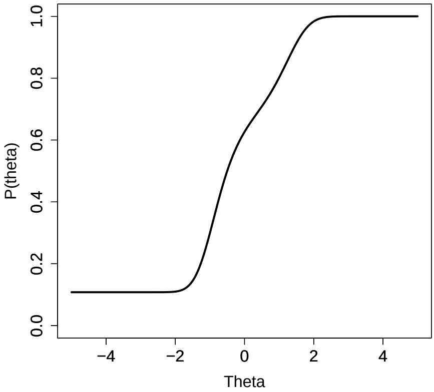
$$P(y_{ij} = 1|\theta_i, \kappa_j, \delta_j, \gamma_j) = c(\kappa_j) + \frac{1 - c(\kappa_j)}{1 + \exp(-(\delta_j + \gamma_j\theta))}$$

- LMPA: Logistic Function of Monotonic Polynomial with Asymptote

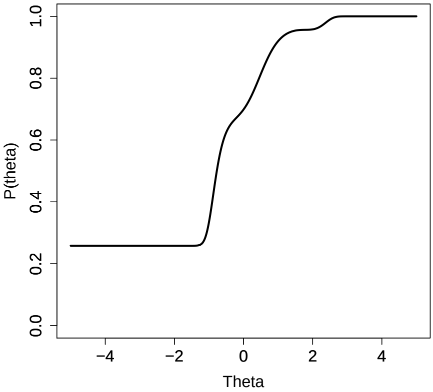
$$P(y_{ij} = 1|\theta_j, \kappa_i, \xi_i, \omega_i, \boldsymbol{\alpha}_i, \boldsymbol{\tau}_i) = c(\kappa_i) + \frac{1 - c(\kappa_i)}{1 + \exp[-m_i(\theta_j)]}$$

$$c(\kappa_i) = \frac{1}{1 + \exp(-\kappa_i)}$$

Example IRFs (from PISA 2000 Read Book 8)



LMPA



LMPA

Generalized Partial Credit - MP (GPCMP)

- GPC

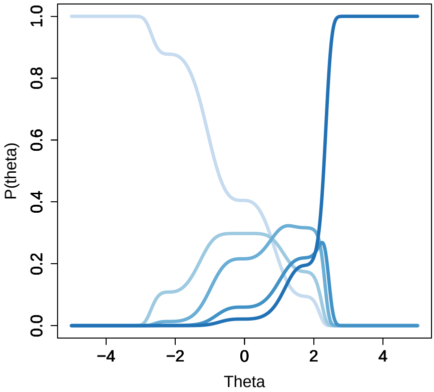
$$P(y_{ij} = q|\theta_j, \boldsymbol{\delta}_i, \gamma_i) = \frac{\exp\left[\sum_{v=0}^{q}(\delta_{iv} + \gamma_i\theta_j)\right]}{\sum_{h=0}^{C_i-1}\exp\left[\sum_{v=0}^{h}(\delta_{iv} + \gamma_i\theta_j)\right]} \quad (1)$$

- GPCMP

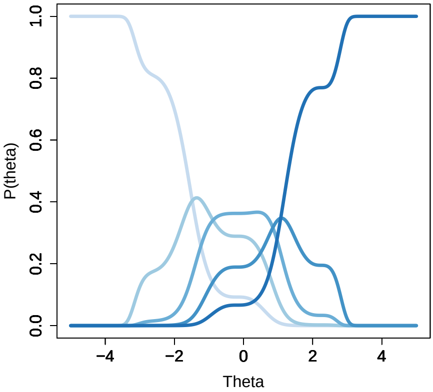
$$P(y_{ij} = q|\theta_j, \boldsymbol{\xi}_i, \omega_i, \boldsymbol{\alpha}_i, \boldsymbol{\tau}_i) = \frac{\exp\left[\sum_{v=0}^{q}(\xi_{jv} + m_i^*(\theta_i))\right]}{\sum_{h=0}^{C_i-1}\exp\left[\sum_{v=0}^{h}(\xi_{jv} + m_i^*(\theta_j))\right]} \quad (2)$$

- $m_i^*(\theta_j)$ is the monotonic polynomial without the intercept term, ξ_i.
- $\xi_{i0} = 0$
- C_i is number of categories

Example IRFs (from PROMIS® Smoking Module)



GPCMP



GPCMP

Estimation

- Misc. Estimation Details
  - EM MML
  - Soft priors often required for τ and α
  - Prior also needed for $c(\kappa)$
  - Identity matrix added to non-positive definite Hessian
- Model Selection
  - All items as $k = 0, 1, 2$
  - AIC selected:
    - Use AIC to select among above 3 models
  - AIC step-wise
    - Start at $k = 0$, loop over items as $k = 1$
    - Select item with best improvement in AIC
    - Repeat (and continue with $k = 2$)

Simulation Study 1

- Purpose:
  - MP versus 2PL, 3PL, GPC
- Manipulated:
  - N: 500, 3000
  - # items: 10, 20
  - Item Type: LMP, LMPA, GPCMP (5 categories)
  - Order: All true items $k = 1$ or 2
- Items:
  - Half based on PISA 2000 Read Book 8 (LMP and LMPA) or PROMIS® Smoking Module (GPCMP)
  - Half randomly generated across 100 replications; based in part on Liang (2007)
- Estimated models
  - All data: $k = 0, 1, 2,$ AIC selected
  - Subset of data: AIC step-wise
  - Scoring: Expected a posteriori (EAP) unless noted

Simulation Study 1

- Purpose:
  - MP versus 2PL, 3PL, GPC
- Manipulated:
  - N: 500, 3000
  - # items: 10, 20
  - Item Type: LMP, LMPA, GPCMP (5 categories)
  - Order: All true items $k = 1$ or 2
- Items:
  - Half based on PISA 2000 Read Book 8 (LMP and LMPA) or PROMIS® Smoking Module (GPCMP)
  - Half randomly generated across 100 replications; based in part on Liang (2007)
- Estimated models
  - All data: $k = 0, 1, 2,$ AIC selected
  - Subset of data: AIC step-wise
  - Scoring: Expected a posteriori (EAP) unless noted

Simulation Study 1

- Purpose:
  - MP versus 2PL, 3PL, GPC
- Manipulated:
  - N: 500, 3000
  - # items: 10, 20
  - Item Type: LMP, LMPA, GPCMP (5 categories)
  - Order: All true items $k = 1$ or 2
- Items:
  - Half based on PISA 2000 Read Book 8 (LMP and LMPA) or PROMIS® Smoking Module (GPCMP)
  - Half randomly generated across 100 replications; based in part on Liang (2007)
- Estimated models
  - All data: $k = 0, 1, 2$, AIC selected
  - Subset of data: AIC step-wise
  - Scoring: Expected a posteriori (EAP) unless noted

Simulation Study 1

- Purpose:
  - MP versus 2PL, 3PL, GPC
- Manipulated:
  - N: 500, 3000
  - # items: 10, 20
  - Item Type: LMP, LMPA, GPCMP (5 categories)
  - Order: All true items $k = 1$ or 2
- Items:
  - Half based on PISA 2000 Read Book 8 (LMP and LMPA) or PROMIS® Smoking Module (GPCMP)
  - Half randomly generated across 100 replications; based in part on Liang (2007)
- Estimated models
  - All data: $k = 0, 1, 2$, AIC selected
  - Subset of data: AIC step-wise
  - Scoring: Expected a posteriori (EAP) unless noted

Root Integrated Mean Square Error

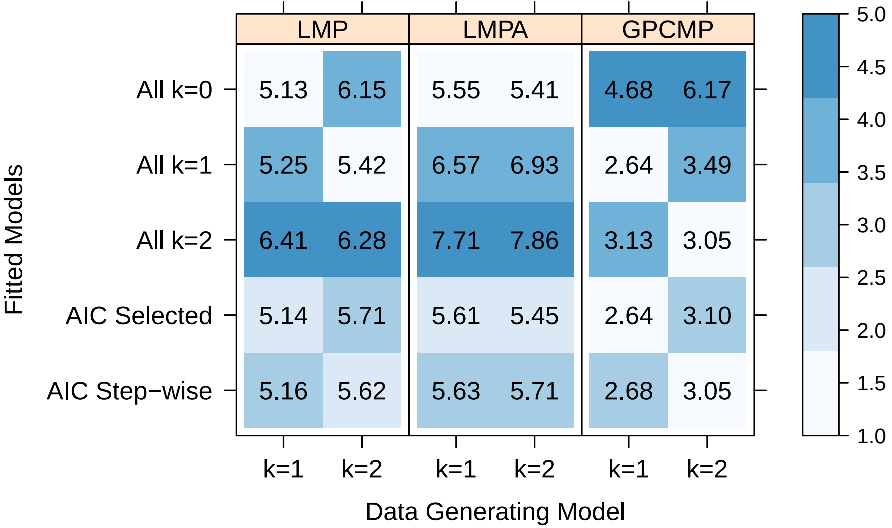
- IRF recovery:[1]

$$RIMSE = \left[\frac{\sum_{r=1}^{R} (\hat{P}(\theta_r) - P(\theta_r))^2 \phi(\theta_r)}{\sum_{r=1}^{R} \phi(\theta_r)} \right]^{1/2} \times 100$$

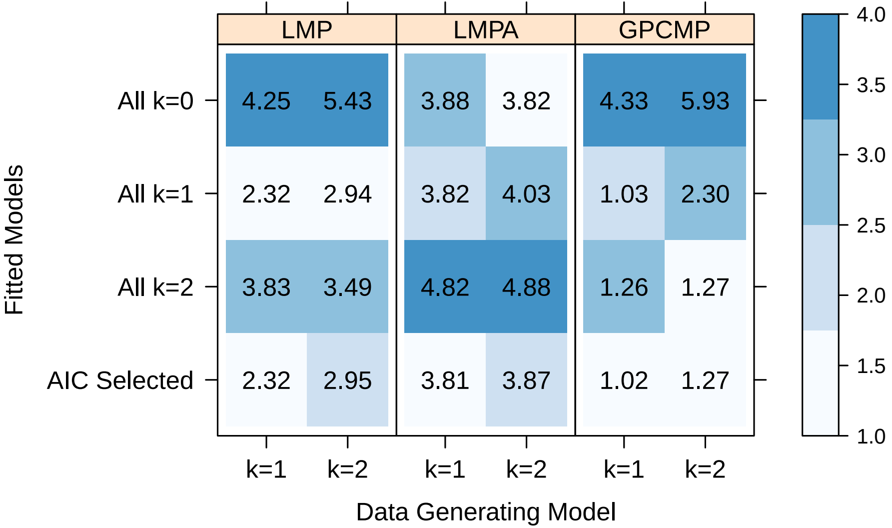
- Replace $P()$ by expected score for polytomous items, $\sum_{h=0}^{C_j-1} hP(y = h|\theta)$
- Replace $P()$ by θ and sum across true θ for evaluation of latent trait recovery

[1] (e.g., Liang, 2007; Ramsay, 1991)

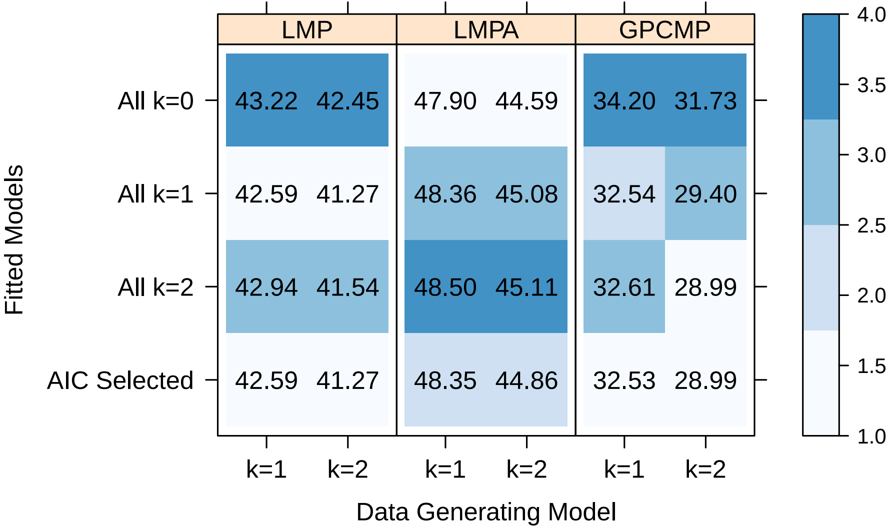
Study 1 Results - IRF Recovery ($N = 500$; 10 items)



Study 1 Results - IRF Recovery ($N = 3000$; 10 items)



Study 1 Results - θ Recovery ($N = 3000$; 10 items)



Simulation Study 2

- Purpose:
  - MP with EM MML versus other approaches
- Manipulated:
  - N: 500, 3000
  - Item Type: Dichotomous (LMP), dichotomous w/ asymptote (LMPA), polytomous (GPCMP; 3 categories)
- Items:
  - 20 items
  - True IRF based on mixture of 2-3 normal CDFs, not MP items
  - CDFs randomly generated across 100 replications
- Estimated models
  - MP as in previous study using EM MML (EM)
  - MP using surrogate-based (SB) approach of Liang (2007)
  - Kernel Smoothing (KS) using KernSmoothIRT package in R

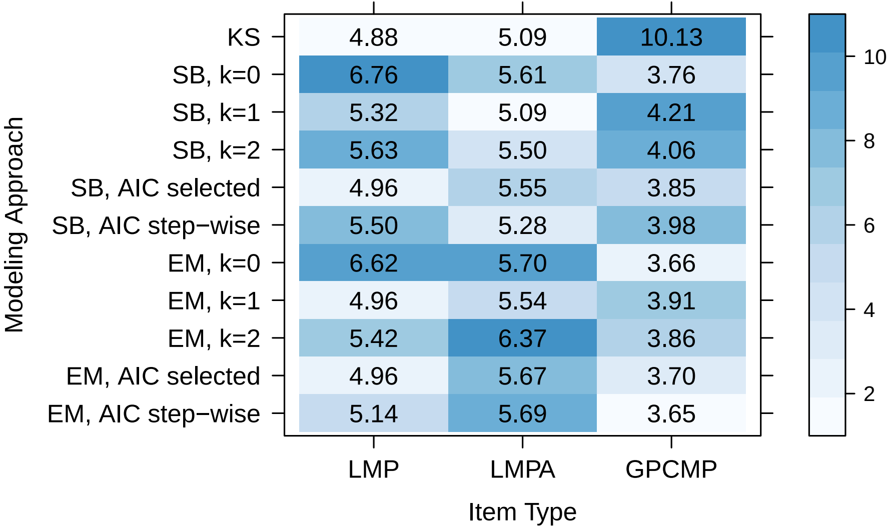
Simulation Study 2

- Purpose:
  - MP with EM MML versus other approaches
- Manipulated:
  - N: 500, 3000
  - Item Type: Dichotomous (LMP), dichotomous w/ asymptote (LMPA), polytomous (GPCMP; 3 categories)
- Items:
  - 20 items
  - True IRF based on mixture of 2-3 normal CDFs, not MP items
  - CDFs randomly generated across 100 replications
- Estimated models
  - MP as in previous study using EM MML (EM)
  - MP using surrogate-based (SB) approach of Liang (2007)
  - Kernel Smoothing (KS) using KernSmoothIRT package in R

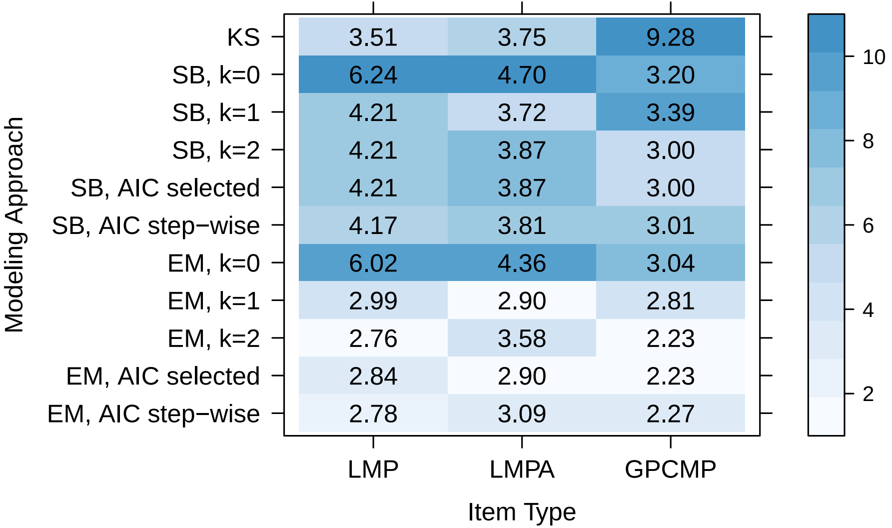
Simulation Study 2

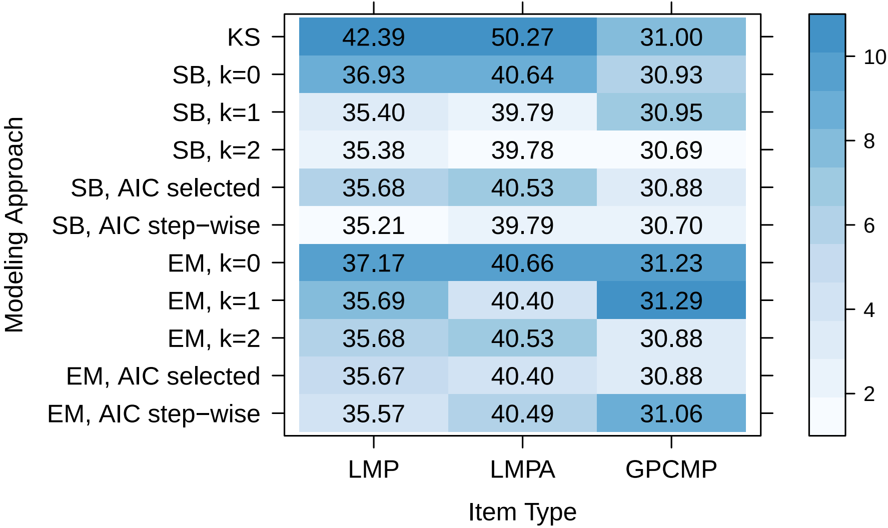
- Purpose:
  - MP with EM MML versus other approaches
- Manipulated:
  - N: 500, 3000
  - Item Type: Dichotomous (LMP), dichotomous w/ asymptote (LMPA), polytomous (GPCMP; 3 categories)
- Items:
  - 20 items
  - True IRF based on mixture of 2-3 normal CDFs, not MP items
  - CDFs randomly generated across 100 replications
- Estimated models
  - MP as in previous study using EM MML (EM)
  - MP using surrogate-based (SB) approach of Liang (2007)
  - Kernel Smoothing (KS) using KernSmoothIRT package in R

Study 2 Results - IRF Recovery ($N = 500$)



Study 2 Results - IRF Recovery ($N = 3000$)




Modeling Approach
KS
SB, k=0
SB, k=1
SB, k=2
SB, AIC selected
SB, AIC step−wise
EM, k=0
EM, k=1
EM, k=2
EM, AIC selected
EM, AIC step−wise
42.39 50.27 31.00
36.93 40.64 30.93
35.40 39.79 30.95
35.38 39.78 30.69
35.68 40.53 30.88
35.21 39.79 30.70
37.17 40.66 31.23
35.69 40.40 31.29
35.68 40.53 30.88
35.67 40.40 30.88
35.57 40.49 31.06
LMP
LMPA
GPCMP
Item Type
10
8
6
4
2

Summary/Conclusion

- MP models can perform better than standard 2PL, 3PL, GPC, and kernel smoothing approaches
  - IRF recovery and EAP estimates
  - Truer at higher sample sizes, more items
  - LMPA requires a good prior for guessing parameter
- AIC works ok for selecting degree of polynomial
- Fitting true MP model not always best
  - e.g., $k = 2$ true model might be best recovered by $k = 1$
- SB vs. EM MML difference in IRF / trait recovery

Summary/Conclusion

- MP models can perform better than standard 2PL, 3PL, GPC, and kernel smoothing approaches
  - IRF recovery and EAP estimates
  - Truer at higher sample sizes, more items
  - LMPA requires a good prior for guessing parameter
- AIC works ok for selecting degree of polynomial
- Fitting true MP model not always best
  - e.g., $k = 2$ true model might be best recovered by $k = 1$
- SB vs. EM MML difference in IRF / trait recovery

Summary/Conclusion

- MP models can perform better than standard 2PL, 3PL, GPC, and kernel smoothing approaches
  - IRF recovery and EAP estimates
  - Truer at higher sample sizes, more items
  - LMPA requires a good prior for guessing parameter
- AIC works ok for selecting degree of polynomial
- Fitting true MP model not always best
  - e.g., $k = 2$ true model might be best recovered by $k = 1$
- SB vs. EM MML difference in IRF / trait recovery

Summary/Conclusion

- MP models can perform better than standard 2PL, 3PL, GPC, and kernel smoothing approaches
  - IRF recovery and EAP estimates
  - Truer at higher sample sizes, more items
  - LMPA requires a good prior for guessing parameter
- AIC works ok for selecting degree of polynomial
- Fitting true MP model not always best
  - e.g., $k = 2$ true model might be best recovered by $k = 1$
- SB vs. EM MML difference in IRF / trait recovery

Thank you!

Part of this research is supported by the Institute of Education Sciences (R305B080016 and R305D100039) and the National Institute on Drug Abuse (R01DA026943 and R01DA030466). Carl Falk is supported by a post-doctoral fellowship from the Social Sciences and Humanities Research Council of Canada. The views expressed here belong to the authors and do not reflect the views or policies of the funding agencies.