Why ‘Random Item’ Response Models?

Sun-Joo Cho

Department of Psychology and Human Development

November 18, 2012
What is ‘random item’ response models?

- ‘Fixed item’ response models:
  Traditional item response models
- ‘Random item’ response models:
  Recent development
  Have been applied in specialized measurement issues
What is ‘random item’ response models?

Why ‘random item’ response models?
  ▶ Answer 1. Improving stability and accuracy of IRT parameter estimation:
    IRT shrinkage estimation
  ▶ Answer 2. Taking the data complexity into account:
    Item modeling for explanation and item generation

Summary
What is ‘random item’ response models?

- ‘Fixed item’ response models:
  Random person effect and fixed item effect (in marginal MLE)

- ‘Random item’ response models:
  Random person effect and random item effect (in hierarchical Bayesian analysis)

<table>
<thead>
<tr>
<th></th>
<th>Fixed item effect</th>
<th>Random item effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed person effect</td>
<td>Joint MLE</td>
<td>Marginal MLE</td>
</tr>
<tr>
<td>Random person effect</td>
<td>Marginal MLE</td>
<td>Hierarchical Bayesian analysis</td>
</tr>
</tbody>
</table>
Marginal maximum likelihood estimation:
Random person effect ($\theta_j$), Fixed item effect ($\beta_i$)

- Random person effect: Different values across persons from the same distribution
  \[ \theta_j \sim N(0, 1) \]
- Fixed item effect: The same values across persons

Example: 3 persons and 3 items

<table>
<thead>
<tr>
<th>person</th>
<th>item</th>
<th>y</th>
<th>$\hat{\theta}_j$</th>
<th>$\hat{\beta}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.5</td>
<td>1.2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0.5</td>
<td>2.2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>2.2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2.5</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2.5</td>
<td>2.2</td>
</tr>
</tbody>
</table>
‘Random item’ response models

- Two types of ‘random item’ response models
  - Random across items
    \[
    \text{logit} \left[ \Pr(y_{ji} = 1|\theta_j, \beta_i) \right] = \mu + \theta_j + \beta_i, \tag{1}
    \]
    where
    \( \mu \) is the average logit of the probability of a correct response, averaging over persons \((j = 1, \ldots, J)\) and items \((i = 1, \ldots, I)\),
    \( \theta_j \sim N(0, \psi_1) \) is a random person parameter [random across persons], and
    \( \beta_i \sim N(0, \psi_2) \) is a random item location parameter [random across items].
  - Random across persons within an item
    \[
    \text{logit} \left[ \Pr(y_{ji} = 1|\theta_j, \beta_{ji}) \right] = \mu + \theta_j + \beta_{ji}, \tag{2}
    \]
    where
    \( \beta_{ji} \sim N(0, \psi_2) \) is a random item location parameter [random across persons].
### Random item effect: Random across items

- Item responses are crossed-classified by persons and items.
- Random item effect and random person effects are *crossed*.

![Diagram showing the crossed classification of items and persons with a table representing the responses.](image-url)
Uses of random item effect [across items]

- Reviews (Cho & Rabe-Hesketh, 2011; De Boeck, 2008)
  - Items may be randomly sampled from an item bank.
    Albers, Does, Imbos, Janssen (1989)
  - We may want to generalize to the “universe of items” as in Generalization Theory, even if items are not randomly sampled.
    Briggs and Wilson (2007)
  - Item family is characterized by the distribution of item parameters.
    Glas & van der Linden (2003); Janssen, Tuerlinckx, Meulders, & De Boeck (2000); Johnson and Sinharay (2005); Sinharay, Johnson, & Williamson (2003)
  - If model includes item covariates, it is natural to allow for a random residual in the “item generation.”
Why ‘random item’ response models?

- **Answer 1. Improving stability and accuracy of IRT parameter estimation:**
  IRT shrinkage estimation

- **Answer 2. Taking the data complexity into account:**
  [1] Item modeling for item generation

  [2] Item modeling for item explanation
Why ‘random item’ response models?

- Why ‘random item’ response models?
- Answer 1: Improving stability and accuracy of IRT parameter estimation
  IRT shrinkage estimation
### Various parameter estimation of item response models

- **Prior for item parameter and/or Marginal over item parameter**

<table>
<thead>
<tr>
<th>Prior for Item Parameter</th>
<th>Marginal over Item Parameter</th>
<th>Marginal over Person Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>NA</td>
<td>No: Joint MLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes: Joint Bayes Modal</td>
</tr>
<tr>
<td>Yes: Single-Stage</td>
<td>No: Bayes Modal</td>
<td>Lord (1968)</td>
</tr>
<tr>
<td></td>
<td>Yes: Bayes Mean (MCMC)</td>
<td>Marginal MLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bock &amp; Lieberman (1970)</td>
</tr>
<tr>
<td>Yes: Hierarchical</td>
<td>No: Bayes Modal</td>
<td>Joint Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>Yes: MLE/ Empirical Bayes</td>
<td>Marginal Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>Yes: Bayes Mean (MCMC)</td>
<td>Mislevy (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hierarchical Joint Bayes Modal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marginal Bayes Modal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mislevy (1986) - float</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random-Item Difficulty Marginal MLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rigdon &amp; Tsutak. (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hierarchical Bayes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Johnson and Albert (1999)</td>
</tr>
</tbody>
</table>

**Cho and Rabe-Hesketh (2012)**
Computational challenges for models with crossed random effect

- Generalized linear and nonlinear mixed models
  Difficulty of numerical integration
    - For nested random effects
      (Rabe-Hesketh, Skrondal, & Pickles, 2005)

The likelihood contribution of a given level-\(l\) unit can be obtained recursively as

\[
\begin{align*}
\text{Level-2: } f^{(2)}(S|\theta^{(3)}, \ldots, \theta^{(L)}) & = \int g(\theta^{(2)}; 0, \Sigma^{(2)}) \prod_{i=1}^{2} f^{(1)}(S|\theta^{(2)}, \ldots, \theta^{(L)}) d\theta^{(2)} \\
\text{Level-3: } f^{(3)}(S|\theta^{(4)}, \ldots, \theta^{(L)}) & = \int g(\theta^{(3)}; 0, \Sigma^{(3)}) \prod_{i=1}^{2} f^{(2)}(S|\theta^{(3)}, \ldots, \theta^{(L)}) d\theta^{(3)} \\
\vdots \\
\text{Level-L: } f^{(L)}(S) & = \int g(\theta^{(L)}; 0, \Sigma^{(L)}) \prod_{i=1}^{L-1} f^{(L-1)}(\theta|\theta^{(L)}) du^{(\theta)}
\end{align*}
\]

- For crossed random effects??
What is ‘Random Item’ response models?  
Why ‘random item’ response models?  
Summary

Different estimation methods for models with crossed random effect [Random intercept]

- **Approximations to MLE: Approximation to integrand**
  - Pairwise likelihood approach: Bello & Varin (2005)

- **Approximations to MLE: Approximation to integral**
  - Cho & Rabe-Hesketh (2011); Cho, Partchev, & De Boeck (2012)

- **Monte Carlo EM (MCEM)**
  - Vaida & Meng (2005)

- **Markov chain Monte Carlo (MCMC)**
  - Karim & Zeger (1992); Rasbash & Browne (2007)
Motivating example: Salamander mating data

- PQL (Breslow & Clayton, 1993; Booth & Hobert, 1999), H-likelihood (Lee et al, 2006), Pairwise likelihood (Bellio & Varin, 2005), MCMC (Karim & Zeger, 1992), Monte Carlo EM (MCEM) (Vaida & Meng, 2005)

Do salamanders from different populations mate successfully?

60 females and 60 males of two species of salamander were paired with a crossed, blocked, and incomplete design. Each male mates with 6 females and each female with 6 males.
Different estimates

Two slice-EM algorithms for fitting GLMM

Table 1 The salamander mating data: RB, Rough Butt; WS, White Side; M, males; F, females

<table>
<thead>
<tr>
<th></th>
<th>RBM</th>
<th>WSM</th>
<th>RBM</th>
<th>WSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

where $\beta_{ij}$ is the fixed effect corresponding to the species combination of the $\{i,j\}$-pair of salamanders with $\beta=(\beta_{RR}, \beta_{RW}, \beta_{WR}, \beta_{WW})$; $u=(u^F, u^M)$ is the vector of female and male random effects, respectively, for which it is assumed that, independently, $u^F_i \sim N(0, \sigma^2_F), u^M_i \sim N(0, \sigma^2_M), i, j = 1, \ldots, 60$. Each animal participates in six matings.
Different estimates

Data structure:

<table>
<thead>
<tr>
<th></th>
<th>Salamander Data</th>
<th>Person × Item Response Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td>Binary responses</td>
<td>Binary responses</td>
</tr>
<tr>
<td>Effect</td>
<td>Male and Female</td>
<td>Person and Item</td>
</tr>
<tr>
<td>Model</td>
<td>Logistic Model</td>
<td>Logistic Model</td>
</tr>
<tr>
<td>Parameters</td>
<td>Random intercept for males</td>
<td>Random person</td>
</tr>
<tr>
<td></td>
<td>Random intercept for females</td>
<td>Random item location (difficulty)</td>
</tr>
</tbody>
</table>

Computational challenges:

- Crossed random effects: Random effect for male salamanders and random effect for female salamanders
- High-dimensionality: Six 20-dimensional integrals
- Small cluster sizes: 6
  (Each male mates with 6 females and each female with 6 males.)
What is ‘Random Item’ response models?

Why ‘random item’ response models?

Summary

Different estimates

- The standard deviation of a random intercept [random item difficulty] for female salamanders
  Cho and Rabe-Hesketh (2011)
Random item discrimination?

Different terms

- Random item discrimination in item response models
- Random factor loading in item factor models
- Random coefficient in generalized nonlinear mixed-effect models
Item discrimination parameter estimation

- Estimation of discrimination parameters of item response models can be problematic, sometimes yielding extremely large estimates.
  - Joint MLE (Swaminathan & Gifford, 1985)
  - Marginal MLE (Bock & Aitkin, 1981)

- Solutions?
  - Imposing an upper bound on the discrimination parameters (Lord & Novick, 1968; Wingersky, 1983)
  - Specify Bayesian prior distributions for the discrimination parameters (Johnson & Albert, 1999; Mislevy, 1986; Swaminathan & Gifford, 1985)
Item discrimination parameter estimation

- Dominant solution in IRT: Hierarchical Bayesian estimation
  - No consensus on the use of prior and hyper-prior (Cho & Suh, 2011)
  - Sensitive to the distributional form of priors for the small number of items (less than 20-item) (Cho, 2010)
- Alternative?
Motivation and purpose of the study (Cho & Rabe-Hesketh, 2012)

- **Motivation:**
  - Difficulty with estimating item discrimination parameters of item response models in joint and marginal maximum likelihood estimation (MLE) methods
  - No consensus on the use of prior and hyper-prior in hierarchical Bayesian solution

- **Purpose:** Improving stability and accuracy of item discrimination parameter estimation in 2-parameter item response models

- **Method:** Random item discrimination marginal MLE alternating imputation posterior (AIP) estimation
## Prior for item parameter and/or Marginal over item parameter

<table>
<thead>
<tr>
<th>Prior for Item Parameter</th>
<th>Marginal over Person Parameters</th>
<th>Marginal over Item Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>●Joint MLE</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>●Marginal MLE</td>
</tr>
<tr>
<td>Yes: Single-Stage</td>
<td>No: Bayes Modal</td>
<td>●Joint Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>Yes: Bayes Mean (MCMC)</td>
<td>●Marginal Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>●Bayes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Albert (1992)</td>
</tr>
<tr>
<td>Yes: Hierarchical</td>
<td>No: Bayes Modal</td>
<td>●Hier. Joint Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>Yes: MLE/Empirical Bayes</td>
<td>●Hierarchical Marginal Bayes Modal</td>
</tr>
<tr>
<td></td>
<td>Yes: Bayes Mean (MCMC)</td>
<td>●Random-Item Discrimination Marginal MLE</td>
</tr>
</tbody>
</table>

Lord (1968)

Bock & Lieberman (1970)

Swam. & Gifford (1985)

Mislevy (1986)

Johnson and Albert (1999)

Cho & Rabe-Hesketh (2012)
Random item discrimination marginal MLE

2-parameter item response model with a random item discrimination

\[
\text{logit} \left[ \Pr(y_{ji} = 1|\theta_j, \alpha_i) \right] = \alpha_i \cdot \theta_j - \beta_i = (\gamma + a_i) \cdot \theta_j - \beta_i, \quad (3)
\]

- Random item discrimination: \( a_i \sim N(0, \psi); \ \alpha_i \sim N(\gamma, \psi) \)
- Random person: \( \theta_j \sim N(0, 1) \) [Model identification constraint]
- Fixed location (difficulty): \( \beta_i \)
Random item discrimination marginal MLE

- Treating a discrimination parameter as a latent random variable to be integrated out, along with a person parameter, to obtain the marginal likelihood

\[
L(y; \beta, \gamma, \psi) = \int_\theta \int_\alpha \left\{ \prod_{j=1}^{J} P(\theta_j) \prod_{i=1}^{I} f(y_{ij}|\theta_j, \alpha_i; \beta, \gamma, \psi) \right\} \prod_i P(\alpha_i; \gamma, \psi) \, d\alpha \, d\theta
\]  

(4)

- Comparison: Likelihood function for marginal MLE

\[
L(y; \beta, \alpha) = \prod_{j=1}^{J} \int_\theta f(y_j|\theta_j; \beta, \alpha)P(\theta_j)d\theta_j
\]  

(5)
Random item discrimination marginal MLE

- **ML estimates** of $\beta$ [item locations], $\gamma$ [mean of item discriminations], and $\psi$ [variance of item discriminations]
- **Empirical Bayes prediction** of $\theta_j$ [person score] and $a_i$ [item discrimination]

\[
\text{logit } [\Pr(y_{ji} = 1|\theta_j)] = \alpha_i \cdot \theta_j - \beta_i = (\gamma + a_i) \cdot \theta_j - \beta_i, \quad (6)
\]
Random item discrimination marginal MLE

Population assumption (random item effect across items) actually reduces the number of parameters because it acts as a constraint on the item-specific parameters.

\[
\text{logit} \left[ \Pr(y_{ji} = 1 | \theta_j, \alpha_i) \right] = \alpha_i \cdot \theta_j - \beta_i = (\gamma + a_i) \cdot \theta_j - \beta_i, \quad (7)
\]

\[a_i \sim N(0, \psi); \; \alpha_i \sim N(\gamma, \psi):\]
The same distribution with the same variance for each item (exchangeability)

40-item example

- **Fixed** item discrimination \(\alpha_i\):
  - 40 item discrimination parameters
- **Random** item discrimination \(\alpha_i \sim N(\gamma, \psi)\):
  - 2 population parameters
Shrinkage estimation

- Hierarchical Bayesian estimation and alternating imputation posterior (AIP) estimation:
  Shrinkage estimates to the mean of item discriminations ($\gamma$)
  $\alpha_i \sim N(\gamma, \psi)$
Shrinkage estimation

- Mean square error of an estimator (MSE)
  \[ \text{MSE} = \text{Variance} + \text{Squared bias of the estimator} \]

- In different estimation methods,
  - MLE: larger variance, small bias
  - Bayesian, AIP: small variance, large bias
  - Total MSE in MLE > Total MSE in Bayesian and AIP estimation
Random item discrimination marginal MLE

- Alternating imputation posterior (AIP) algorithm with adaptive quadrature (AQ) 

<table>
<thead>
<tr>
<th>Person Wing</th>
<th>Item Wing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known discrimination parameters</td>
<td>Known person parameters</td>
</tr>
<tr>
<td>MLE of $\vartheta_1 = {\gamma, \beta_i}$</td>
<td>MLE of $\vartheta_2 = {\gamma, \beta_i, \log\psi}$</td>
</tr>
<tr>
<td>Sample $\vartheta_1 \sim MN(\hat{\vartheta}<em>1, \hat{\Sigma}</em>{\vartheta_1})$</td>
<td>Sample $\vartheta_2 \sim MN(\hat{\vartheta}<em>2, \hat{\Sigma}</em>{\vartheta_2})$</td>
</tr>
<tr>
<td>Sample $\theta$</td>
<td>Sample $a$</td>
</tr>
</tbody>
</table>

$$\logit [\Pr(y_{ji} = 1|\theta_j, \alpha_i)] = \alpha_i \cdot \theta_j - \beta_i = (\gamma + a_i) \cdot \theta_j - \beta_i,$$

- Random item discrimination: $a_i \sim N(0, \psi); \alpha_i \sim N(\gamma, \psi)$
- Random person: $\theta_j \sim N(0, 1)$ [Model identification constraint]
- Fixed location (difficulty): $\beta_i$
AIP with adaptive quadrature: Step 2 [Posterior]

- Sampling parameters from a multivariate normal distribution with mean given by MLEs and covariance matrix by the inverse of the estimated information matrix

  - Person Wing for $\vartheta_1 = \{\gamma, \beta_i\}$
    
    $$\vartheta_1^k | a^{k-1} \sim MN(\hat{\vartheta}_1^k, \hat{\Sigma}_\vartheta_1^k)$$  (9)

  - Item Wing for $\vartheta_2 = \{\gamma, \beta_i, \log \psi\}$
    
    $$\vartheta_2^k | \theta^k \sim MN(\hat{\vartheta}_2^k, \hat{\Sigma}_\vartheta_2^k)$$  (10)

- AIP vs. Bayesian analysis (Gibbs Sampling)

  - Not necessary to assume hyperprior distributions in AIP
  - Vector of nodes sampled in AIP so that AIP will converge more rapidly than Gibbs sampling (Rao-Blackwellization)
AIP with adaptive quadrature: Step 3 [Imputation]

- **Discrimination parameters**
  
  **Normal approximation** with item-specific posterior mean $\mu_{a_i}^k$ and posterior standard deviation $\tau_{a_i}^k$ for parameters $\vartheta_2^k$,

  \[
P(a_i|y_i, \theta^k; \gamma^k, \beta^k) \sim N(a_i; \mu_{a_i}^k, \tau_{a_i}^k).
  \]  

- **Person parameters**
  
  **Normal approximation** with person-specific posterior mean $\mu_{\theta_j}^k$ and posterior standard deviation $\tau_{\theta_j}^k$ for parameters $\vartheta_1^k$,

  \[
P(\theta_j|y_j, a^{k-1}; \gamma^k, \beta^k) \sim N(\theta_j; \mu_{\theta_j}^k, \tau_{\theta_j}^k).
  \]
### Estimation results

<table>
<thead>
<tr>
<th>iteration $k$</th>
<th>Person Wing</th>
<th>Item Wing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\gamma}$[MLE]</td>
<td>$\hat{\beta}_i$[MLE]</td>
</tr>
</tbody>
</table>
Posterior means and variances

Poster Mean

\[ \hat{E}(\hat{\vartheta} | y) \approx \overline{\hat{\vartheta}} \equiv \frac{1}{n-p} \sum_{k=p+1}^{n} \hat{\vartheta}^k. \]  

(13)

The covariance matrix is estimated by both within variance [the first term] and between variance [the second term](Uncertainty of parameter).

\[ \text{Var}(\hat{\vartheta} | y) \approx \frac{1}{n-p} \sum_{k=p+1}^{n} \hat{\Sigma} \hat{\vartheta} + \frac{1}{n-p-1} \sum_{k=p+1}^{n} (\hat{\vartheta}^k - \overline{\vartheta})(\hat{\vartheta}^k - \overline{\vartheta})'. \]  

(14)

MSE = Variance + Square bias of the estimator
Variance = Within variance in MLE
Variance = Within variance + between variance (Uncertainty of parameter) in AIP
Predictions and associated variances: Item discriminations

- **Prediction mean**

\[
E(a_i | y_i) \approx \bar{\mu}_{ai} = \frac{1}{n - p} \sum_{k=p+1}^{n} \hat{\mu}_{ai}^k
\]  

- The covariance matrix is estimated by both within variance [the first term] and between variance [the second term].

\[
\text{Var}(a_i | y_i) \approx \frac{1}{n - p} \sum_{k=p+1}^{n} \hat{\tau}_{ai}^k + \frac{1}{n - p - 1} \sum_{k=p+1}^{n} (\hat{\mu}_{ai}^k - \bar{\mu}_{ai})(\hat{\mu}_{ai}^k - \bar{\mu}_{ai})
\]  

(15)
‘Uncertainty” in estimates

- “Uncertainty” reflects lack of complete knowledge about a parameter.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fixed/Random on parameters</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal MLE</td>
<td>Random</td>
<td>No</td>
</tr>
<tr>
<td>Hierarchical Bayesian analysis</td>
<td>Random</td>
<td>Yes</td>
</tr>
<tr>
<td>AIP</td>
<td>Random</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Comparisons

- Different estimation methods for item discriminations

<table>
<thead>
<tr>
<th>Method</th>
<th>Fixed/Random</th>
<th>Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Marginal MLE (MMLE)</td>
<td>Fixed</td>
<td>No</td>
</tr>
<tr>
<td>2. Marginal Bayes Modal (MBM)</td>
<td>Fixed</td>
<td>logN(0,0.25)</td>
</tr>
<tr>
<td>3. Hierarchical Bayes (MCMC)</td>
<td>Random</td>
<td>normal on $\gamma$ ($N(0,1000)$); half-Cauchy on $\sqrt{\psi}$</td>
</tr>
<tr>
<td>4. AIP</td>
<td>Random</td>
<td>normal priors on $\gamma$ and $\log \psi$ uniform hyper-priors as MLE</td>
</tr>
</tbody>
</table>

- In all four methods
  - Same person parameter distribution: $\theta_j \sim N(0,1)$
    - MMLE and MBE: Uncertainty of item discrimination parameters is ignored.
    - AIP and MCMC: Uncertainty of item discrimination parameters is taken into account.
  - Same *fixed* item intercept parameter: $\beta_i$
Empirical comparisons

- College-level English placement test with large sample sizes and the number of items: 1,000-examinee and 31-item
- Root mean square difference for item parameters

<table>
<thead>
<tr>
<th></th>
<th>MMLE</th>
<th>MBM</th>
<th>MCMC</th>
<th>AIP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBM</td>
<td>0.097</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>0.227</td>
<td>0.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIP</td>
<td>0.269</td>
<td>0.173</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td><strong>Item Discrimination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBM</td>
<td>0.099</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>0.208</td>
<td>0.126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIP</td>
<td>0.261</td>
<td>0.171</td>
<td>0.069</td>
<td></td>
</tr>
</tbody>
</table>
Empirical comparisons

- Person ability scoring: Expected a Posteriori (EAP)

  Kendall’s rank order correlation

<table>
<thead>
<tr>
<th></th>
<th>MMLE</th>
<th>MBM</th>
<th>MCMC</th>
<th>AIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBM</td>
<td>0.877</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>0.742</td>
<td>0.831</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIP</td>
<td>0.696</td>
<td>0.803</td>
<td>0.830</td>
<td></td>
</tr>
</tbody>
</table>

Sun-Joo Cho

Why ‘Random Item’ Response Models?
Empirical comparisons

- Standard errors of EAPs with uncertainty of item discrimination parameters in MCMC and AIP

- MSE = Variance + Square bias of the estimator
  Variance = Within variance in MLE
  Variance = Within variance + between variance (Uncertainty of parameter) in AIP

\[ P(\theta_j | y_j, a^{k-1}, \gamma^k, \beta^k, 1) \sim N(\theta_j; \mu^k_{\theta_j}, \tau^k_{\theta_j}) \quad (17) \]
Simulation study

- Simulation design
  - 100-person and 15-item
    This condition was designed to investigate the performance of MMLE, MBM, MCMC, and AIP with small sample size and short test.
  - 250-person and 25-item
    This condition was chosen to study the performance of MMLE, MBM, MCMC, and AIP when the size of the prior variance may not matter in estimating item discrimination parameters (Harwell & Janosky, 1991).
Mean and variance of a random item discrimination parameter:

- Smaller RMSE and Bias in AIP than in MCMC
- Underestimated variance in AIP; Overestimated variance in MCMC
- No significant bias for variance in AIP

<table>
<thead>
<tr>
<th>Method</th>
<th>J = 100 and I = 15</th>
<th>J = 250 and I = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>0.357</td>
<td>0.322*</td>
</tr>
<tr>
<td>AIP</td>
<td>0.197</td>
<td>0.162*</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>0.272</td>
<td>0.215*</td>
</tr>
<tr>
<td>AIP</td>
<td>0.124</td>
<td>-0.051</td>
</tr>
</tbody>
</table>

* Significantly different from 0 at the 5% level.

J: Number of persons; I: Number of items
Simulation results

- Item discriminations
  - The smallest RMSE and Bias in AIP
  - No significant Bias in AIP

<table>
<thead>
<tr>
<th>Method</th>
<th>$J = 100$ and $I = 15$</th>
<th>$J = 250$ and $I = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corr.</td>
<td>RMSE</td>
</tr>
<tr>
<td>MMLE</td>
<td>0.731</td>
<td>0.700</td>
</tr>
<tr>
<td>MBM</td>
<td>0.756</td>
<td>0.487</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.739</td>
<td>0.432</td>
</tr>
<tr>
<td>AIP</td>
<td>0.731</td>
<td>0.408</td>
</tr>
</tbody>
</table>

* Significantly different from 0 at the 5% level.

$J$: Number of persons; $I$: Number of items
Corr.: Correlation(true parameters, estimates)
Simulation results

- **Item intercepts:**
  Similar results in Bias and RMSE across four estimation methods

- **Person parameters:**
  Similar results in Bias and RMSE across four estimation methods
  
  Kendall's rank order correlation between true parameters and estimates

<table>
<thead>
<tr>
<th></th>
<th>$J = 100$ and $I = 15$</th>
<th>$J = 250$ and $I = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MMLE</strong></td>
<td>0.703</td>
<td>0.844</td>
</tr>
<tr>
<td><strong>MBM</strong></td>
<td>0.796</td>
<td>0.844</td>
</tr>
<tr>
<td><strong>MCMC</strong></td>
<td>0.796</td>
<td>0.844</td>
</tr>
<tr>
<td><strong>AIP</strong></td>
<td>0.796</td>
<td>0.954</td>
</tr>
</tbody>
</table>

*J*: Number of persons; *I*: Number of items
Summary

- Comparisons for item discriminations, given simulation conditions we considered:
  \[ \text{MMLE} < \text{MBE} < \text{MCMC} < \text{AIP} \]

- Alternative to hierarchical Bayesian analysis, esp. for the small number of items and small sample data

- Extensions
  (Cho, 2012; Cho, De Boeck, Embretson, & Rabe-Hesketh, 2012)
  - Shrinkage item estimates for both item difficulties and discriminations
  - Shrinkage item estimates with item attributes
  - Shrinkage item estimates for multilevel item structure
  - Shrinkage item estimates for multidimensional data
Why ‘random item’ response models?

Answer 1. Improving stability and accuracy of IRT parameter estimation:
IRT shrinkage estimation

Answer 2. Taking the data complexity into account:
[1] Item modeling for item generation

[2] Item modeling for item explanation
Practical motivations of random item modeling

- Item generation
- Item explanation
Item generation

- Item generation is becoming important for building an item bank, for adaptive testing, or even to generate items on the fly during adaptive testing.
- Two approaches
  - Item cloning approach (Glas & van der Linden, 2003)
  - Structural modeling approach (Embretson, 1998)
    - Item attribute approach (Freund, Hofer, & Holling, 2008)
- Combined approach
  (Cho, De Boeck, Embretson, & Rabe-Hesketh, revised & resubmitted)
Item generation

- Example: An inductive test with number series of the type “\(x_1, x_2, x_3, x_4, \ldots?\)” where the next number has to be filled out.

- Item cloning approach (Glas & van der Linden, 2003)
  - A parent item with a given rule: \(x_t = x_{t-1} + 1 + t\) (\(t = 1, \ldots\))
  - Item families:
    - (12, 15, 19, 24, \ldots?)
    - (34, 37, 41, 46, \ldots?)
    - (26, 29, 33, 38, \ldots?)

- Structural modeling approach (Embretson, 1998);
- Item attribute approach (Freund, Hofer, & Holling, 2008)
  - Item attribute: Low vs. high numbers for series (1, 4, 8, 13, \ldots?) vs. (1123, 1126, 1130, 1135, \ldots?)
  - Item category [parent item] attribute: The inductive rule requires addition vs. multiplication.
Item model

- **Question for item generation data:** Can we explain and predict item parameters using the item generation rules and attributes?
- **Item model:** Psychometric model to explain and predict item parameters based on the item generation rules and attributes.
- **Why an item model?:** One wants it to be optimally informative in an adaptive procedure.
What is ‘Random Item’ response models?

Why ‘random item’ response models?

Summary

Item model

- IRT measurement model: 2-parameter item response model

$$\text{logit} [\Pr(y_{pi} = 1|\theta_p)] = \eta_{pi} = a_i \cdot \theta_p - b_i, \quad (18)$$

where $p$ is a person index ($p = 1, \ldots, P$), $i$ is an item index ($i = 1, \ldots, I$), $\eta_{pi}$ is the logit of the conditional probability $\Pr(y_{pi} = 1|\theta_p)$ that $y_{pi} = 1$, $\theta_p$ is the normally distributed underlying person trait with a mean of 0 and a variance of 1 for model identification, $a_i$ is a slope or discrimination parameter, and $b_i$ is a location or difficulty parameter.

- Item model
  
  $a_i = ??$
  $b_i = ??$

  Reasonable predictive value based on the item generation rules and attributes??
What is ‘Random Item’ response models?

Why ‘random item’ response models?

Summary

Item structure in item generation (Combined approach)

- Multilevel item structure and item & item category attributes

- Item categories
  - random effect: **RC**

- Individual items
  - random effect: **RI**

- Category attributes
  - fixed effect: **FCA**

- Item attributes
  - fixed effect: **FIA**

Sun-Joo Cho

Why ‘Random Item’ Response Models?
Item response model for the combined approach

- Additive multilevel item structure models with random residuals
  (Cho, De Boeck, Embretson, & Rabe-Hesketh, revised & resubmitted)

\[
\text{logit} \left[ \Pr(y_{pi} = 1|\theta_p) \right] = \eta_{pi} = a_i \cdot \theta_p - b_i
\]  \hspace{1cm} (19)

\[
a_i = \mu_\alpha + \sum_d \gamma_{\alpha d} Q_{id} + \varepsilon_{\alpha i}^{(1)} + \sum_t \delta_{\alpha t} R_{ct} + \varepsilon_{\alpha c}^{(2)}
\]  \hspace{1cm} (20)

\[
b_i = \mu_\beta + \sum_d \gamma_{\beta d} Q_{id} + \varepsilon_{\beta i}^{(1)} + \sum_t \delta_{\beta t} R_{ct} + \varepsilon_{\beta c}^{(2)}
\]  \hspace{1cm} (21)

- Item attribute and its weight: \( Q_{id} \) and \( \gamma \)
- Item category attribute: \( R_{ct} \) and \( \delta \)
- Mean: \( \mu \)
- Random item residual at Level 1: \( \varepsilon_i^{(1)} \)
- Random category residual at Level 2: \( \varepsilon_c^{(1)} \)
Consequence of ignoring unexplained variance

Why random item effect? ($\varepsilon_{\alpha i}^{(1)}$):
Since an explanation is almost never perfect, an error term should be included, as in regression models, for the item variation that cannot be attributed to the item attributes.

$$a_i = \mu_{\alpha} + \sum_d \gamma_{\alpha d} Q_{id} + \varepsilon_{\alpha i}^{(1)} \quad (22)$$

Consequence of ignoring unexplained variance
Omitting the random residuals can lead to underestimated standard errors for the covariate effects on the IRT model parameters.
Consequence of ignoring multilevel structure

- Why category random effect? \( (\varepsilon_{\alpha c}^{(2)}) \)

\[
a_i = \mu_\alpha + \sum_d \gamma_{\alpha d} Q_{id} + \varepsilon_{\alpha i}^{(1)} + \sum_t \delta_{\alpha t} R_{ct} + \varepsilon_{\alpha c}^{(2)} \tag{23}
\]

- Consequence of ignoring multilevel structure
  - Bias in parameter estimation and an increase in the mean absolute estimation error
  - Inflated variance of random item effect
Nested & crossed random effects:
AIP with adaptive quadrature was developed for the multilevel item structure.

- Nested random effects: Random item effect and random item category effect
- Crossed random effects: Random item effect/random item category effect and Random person effect
Why ‘random item’ response models?

- **Answer 1.** Improving *stability and accuracy of IRT parameter estimation*:
  IRT shrinkage estimation

- **Answer 2.** Taking the data complexity into account:
  [1] Item modeling for item generation
  [2] Item modeling for item explanation
Empirical research questions raised by my co-authors (Amanda Goodwin at T & L; Jennifer Gilbert at SPED)

1. How do the various sources of information (i.e., decoding, spelling, and meaning) that comprise lexical representations relate to one another?

[Measurement Model]
2. Do item level characteristics (i.e., frequency, consistency, transparency, and number of morphemes) $[X]$ and person level characteristics (i.e., vocabulary knowledge, reading comprehension, and morphological awareness) $[Z]$ affect lexical representations? [Explanation Model]
Two approaches

- Two-step approach?:
  [Measurement Model] first, then [Explanatory Model]
  Measurement error of the estimated parameters is not taken into account when the effects of covariates on parameters are estimated.

- Simultaneous approach?:
  [Measurement Model] + [Explanatory Model]
word *perceive*

- “Please read the following word aloud: *perceive*”
- “Please spell the word *perceive* on your paper.”
- “Do you know the meaning of the word *perceive*?”
Multidimensional and multilevel item structures

- Multidimensional with respect to item groups
  Multidimensional lexical quality: decoding, spelling, and meaning
- Multilevel on the item side

Decoding Items: Please read the following word along: “Word”.
Spelling Items: Please spell “Word” on your paper.
Meaning Items: Do you know the meaning of the word “Word”? 
Multidimensional multilevel random item response model

- Random item effect:
  To take unexplained variance into account

- Random item category [item group] effect
  To take dependency in item responses due to item group into account

  - Intra-class correlation for item groups:
    32.1 % of total item difficulty variances are explained by item groups.
What is ‘Random Item’ response models?

Why ‘random item’ response models?

Summary

Multidimensional multilevel random item response model

- Estimation: Laplace estimation
- “The (final) model fitted is rather complex, with a substantial number of parameters and a complex structure of the measurement model. I have some doubts about the sample size requirements for fitting such a model, and whether 172 participants is enough.”
  - Measurement model
    - Fixed item approach: 156
    - Random item approach: 4
  - Measurement model + Explanatory model
    - Fixed item approach: 156 + 51
    - Random item approach: 4 + 51
Why ‘random item’ response models?

- Answer 1. Improving stability and accuracy of IRT parameter estimation: IRT shrinkage estimation
- Answer 2. Taking the data complexity into account: Item modeling for explanation and item generation
Future work

- Methodological work
  - Evaluation of likelihood function
  - Evaluation of model selection method
  - Testing of variance parameters
  - Code release to the public (Stata Programming Language)
    Stata command:
    - Rasch model (2011, CSDA): `AIP, iter() nip() nii() sav()`
    - Multidimensional Rasch model (2012, BJMSP): `MAIP, iter() nip() nii() sav()`
    - 2-parameter model (under review, PMET): `2AIP, iter() nip() nii() sav()`
    - Graded response model (in prep.)
    - Posterior moments (2011, CSDA): `PM`

- Empirical work
  - Probing interactions between persons and items in reading education
  - Intervention design in special education
  - Response time-accuracy data analysis in neuropsychology