Bayesian Nonparametric Rasch Modeling: Methods and Software

George Karabatsos
University of Illinois-Chicago

Keynote talk
Friday May 2, 2014 (9:15-10am)
Ohio River Valley Objective
Measurement Seminar (ORVOMS)

Supported by NSF-MMS Research Grant SES-1156372
Outline of Presentation

I. Review Rasch model
   – Key feature of model: Interpretability and simplicity.

II. Issues with Rasch modeling in practice
    – Simple models often misfit real data.
    – Difficulties with Rasch model fit analysis.

III. A Bayesian nonparametric (BNP) Rasch model as a solution
     – New model provides “Rasch analysis without fit statistics,”
       and “automatic Rasch analysis.”

IV. Illustrate the BNP Rasch model on real NAEP data.
    – Compare with ordinary Rasch, 1PL, 2PL, and 3PL models.
    – Demonstrate my free software with new BNP Rasch model.

V. Conclusions
Rasch (1960) Model

- \( \Pr[Y_{ij} = y \mid \theta_i, \beta_j] = \Phi(\theta_i - \beta_j)^y[1 - \Phi(\theta_i - \beta_j)]^{1-y} \)

- \( Y_{ij} \in \{y = 0, 1\} \) response by examinee \( i \) on test item \( j \) for \( i = 1, \ldots, n; \ j = 1, \ldots, J \).
- \( \theta_i \) examinee ability; \( \beta_j \) item difficulty (real-valued)

- \( \Phi(\cdot) = \Pr[Y^* \leq \cdot] \) continuous c.d.f.; for **Inverse Link**.

**Usually:** Logistic(0,1) c.d.f.: \( \Phi(\cdot) = \exp(\cdot)/[1+\exp(\cdot)] \)

**Alternatively:** Normal ogive: \( \Phi(\cdot) \): Normal(0,1) c.d.f. 
\( (\Phi(\cdot/1.7) \) approximates the Logistic(0,1))

Or parameterize scale \( \sigma \), for normal \( \Phi(\cdot/\sigma) \).

Etc.
Rasch (1960) Model

- \( \Pr[Y_{ij} = y | \theta_i, \beta_j] = \Phi(\theta_i - \beta_j)^y[1 - \Phi(\theta_i - \beta_j)]^{1-y} \)

- Model is very attractive for its interpretability/simplicity:
  - \( Y_{ij} \) a function of examinee ability & item difficulty.
  - Under model, total test score, \( \sum_j y_{ij} \) sufficient statistic for \( \theta_i \), and total item score \( \sum_i y_{ij} \) sufficient statistic for \( \beta_j \).
  - Binary (e.g., logistic) GLM regression model with examinee indicator \((0,1)\) and item indicator \((0,-1)\) predictors/covariates.
  - Additional covariates can be added easily.
  - Easily extended to handle ratings \( Y_{ij} \), using GLM ideas.
  - Interpretable/simple models, because they are understandable, are preferred for high-stakes decisions involving examinee measurement, and for making policy decisions.
Rasch (1960) Model

- \( \Pr[Y_{ij} = y \mid \theta_i, \beta_j] = \Phi(\theta_i - \beta_j)^y[1 - \Phi(\theta_i - \beta_j)]^{1-y} \)

- Rasch model is attractive for its interpretability/simplicity.
- Unfortunately, real data often poorly fit with simple models, and fit better with more complex/less interpretable models.

- Data misfit makes Rasch model less interpretable/meaningful.
- Data misfit destroys all sufficiency properties of Rasch models.

- For example, for a 6-item test, with items ordered by difficulty, the \( \theta \) estimate of item responses 111000 is the same as the \( \theta \) estimate of item responses 000111.
- In such a situation, the Rasch model is no longer interpretable.
Rasch (1960) Model

• Rasch model is attractive for its interpretability/simplicity.
• Unfortunately, real data often poorly fit with simple models, and fit better with more complex/less interpretable models.

• Fit statistics are often used to identify/remove “problematic” items from Rasch model (and “problematic” persons).
  – This practice is not uncontroversial: leads to information loss.
  – Identifying mis-fitting items is time consuming and difficult: An item may or may not appear to misfit, depending on what other test items happen to be included in the model (!)
  – Many large data sets indicate items misfit the Rasch model.
  – Incoherency: Coherency requires that the model represents the analyst’s beliefs about the data. But the act of model fit checking actually indicates her/his lack of belief in the model.
Rasch (1960) Model

- In real psychometric practice, is it possible to exploit the interpretability and simplicity of the Rasch model, while avoiding the pitfalls of model misfit and fit analysis?

- I propose a Bayesian nonparametric (BNP) Rasch model, which:
  - Can provide accurate estimates of examinee ability and item difficulty parameters, that are robust to (control for) all observed and unobserved covariates/predictors/factors that are excluded from the model;
  - Then, the BNP Rasch model is not “wrong” or “incorrect” for the data, and therefore this model:
    - doesn’t require data fit analysis/model checking; there is no point in performing a model fit analysis.
    - provides a “Rasch analysis without fit statistics.”
Bayesian Nonparametric Rasch Model

- We enlarge the Rasch model: \( \Pr[Y_{ij} = 1 | \theta_i, \beta_j] = \Phi(\theta_i - \beta_j) \)
to the Bayesian nonparametric (BNP) Rasch model, an *infinite mixture* Rasch model defined by:

\[
\Pr[Y_{ij} = 1 | \zeta] = \int \Phi(\{\theta_i - \beta_j + \beta_0 + \mu_0\}/\sigma) dG_{ij}(\mu_0)
= \sum_{k=-\infty}^{\infty} \Phi(\{\theta_i - \beta_j + \beta_0 + \mu_{0k}\}/\sigma) \omega_{ijk}
\]
with \( \sum_{k=-\infty}^{\infty} \omega_{ijk} = 1. \)

where for examinee \( i \) and item \( j \):
- \( \mu_0 \) is the effect of all unobserved covariates/predictors/factors that are excluded from the model;
- \( G_{ij} \) is the mixing distribution, with mixture weights \( \omega_{ijk} = \Phi(\{j - (\theta_{\omega i} - \beta_{\omega j} + \beta_{\omega 0})\}/\sigma_{\omega}) - \Phi(\{j - 1 - (\theta_{\omega i} - \beta_{\omega j} + \beta_{\omega 0})\}/\sigma_{\omega}) \);
- \( \Phi(\cdot) \) is the Normal c.d.f.
BNP Rasch Model

- The BNP Rasch model:

\[
\Pr[Y_{ij} = 1 \mid \zeta] = \int \Phi(\{\theta_i - \beta_j + \beta_0 + \mu_0\}/\sigma)dG_{ij}(\mu_0)
= \sum_{k=-\infty:}\Phi(\{\theta_i - \beta_j + \beta_0 + \mu_{0k}\}/\sigma)\omega_{ijk};
\]

\[
\omega_{ijk} = \Phi(\{j - (\theta_{\omega i} - \beta_{\omega j} + \beta_{\omega 0})\}/\sigma_{\omega}) - \Phi(\{j - 1 - (\theta_{\omega i} - \beta_{\omega j} + \beta_{\omega 0})\}/\sigma_{\omega})
\]

is completed by the specification of prior distributions:

\[
\mu_{0k} \mid \sigma_\mu \sim i.i.d. \text{ Normal}(0, \sigma_\mu^2)
\]
\[
\sigma_\mu \sim \text{ Uniform}(0, b_{\sigma_\mu})
\]
\[
\beta_0 \mid \sigma^2 \sim \text{ Normal}(0, \sigma^2\nu_{\beta 0} \rightarrow \infty)
\]
\[
(\theta, \beta_1:J) \mid \sigma^2 \sim \text{ Normal}_{nJ}(0, \sigma^2\nu I_{nJ})
\]
\[
\sigma^2 \sim \text{ InverseGamma}(a_0/2, a_0/2)
\]
\[
(\theta_{\omega}, \beta_{\omega}) \mid \sigma_{\omega}^2 \sim \text{ Normal}_{nJ}(0, \sigma_{\omega}^2\nu \omega I_{nJ+1})
\]
\[
\sigma_{\omega}^2 \sim \text{ InverseGamma}(a_{0\omega}/2, a_{0\omega}/2)
\]
BNP Rasch Model

- The Bayesian infinite mixture Rasch model has parameters:
  \[ \zeta = (\{\mu_{0k}\}_{k=-\infty}^{\infty}, \sigma_\mu, \theta, \beta, \sigma, \beta_\omega, \theta_\omega, \sigma_\omega). \]

- It is a Bayesian nonparametric model because it has an infinite number of parameters, allowing for high model flexibility.

- Following Bayes’ theorem, the data, \( D = (y_{ij})_{n \times J} \), updates the joint prior density \( \pi(\zeta) \) to a posterior density:
  \[ \pi(\zeta | D) \propto \prod_i \prod_j \Pr[Y_{ij} = y_{ij} | \zeta] \pi(\zeta), \]
  with \( \Pr[Y_{ij} = 1 | \zeta] \) given by the mixture model (previous slide), and \( \Pr[Y_{ij} = 0 | \zeta] = 1 - \Pr[Y_{ij} = 1 | \zeta] \).

- The posterior \( \pi(\zeta | D) \), and all posterior functionals of interest, can be estimated via MCMC methods developed by Karabatsos & Walker (2012, *Elec J Stat*).
Easy Extensions of BNP Rasch Model

- **Ordinal** item responses \( k = 0, 1, \ldots, m \), by model specification:
  \[
  \Pr[Y_{ij} = k | \zeta] = \int_{\{w(k-1) < y^* < w(k)\}} n(y^* | \theta_i - \beta_j + \beta_0 + \mu_0, \sigma) dG_{ij}(\mu_0) dy^*
  \]
given ordinal thresholds:
  \(-\infty \equiv w_{-1} < w_0 = 0 < w_1 = 1 < w_2 = 2 < \cdots < w_m = \infty\),
and \( n(\cdot | \mu, \sigma) \) the density function of the Normal(\( \mu, \sigma \)) distribution.

- **Continuous-valued** item response \( Y_{ij} \), by model specification:
  \[
  f(y_{ij} | \zeta) = \int n(y_{ij} | \theta_i - \beta_j + \beta_0 + \mu_0, \sigma) dG_{ij}(\mu_0).
  \]

- **Extra covariates/predictors** (x) can be added to the model, beyond examinee and item indicators. E.g., judge indicators, and covariates describing the examinees, items, and/or judges (e.g., SES, item time), to provide a “**BNP FACETS model**.” Then we have a more general model for \( G_x(\mu_0) \).
Application of BNP Rasch Model to NAEP data

- Data: 1990 NAEP 6-item reading exam of 4th and 6th graders.
- 75 examinees (randomly). Dichotomous (0,1) item scores.

- Standardized residual:
  \[ \hat{z}_{ij} = \frac{y_{ij} - \Pr(Y_{ij} = 1| \hat{\theta})}{\sqrt{\Pr(Y_{ij} = 1| \hat{\theta})[1 - \Pr(Y_{ij} = 1| \hat{\theta})]}}, \]

- Other Rasch and IRT Models
  Number of Outliers (i.e., item responses where \(|\hat{z}_{ij}| > 2\))
  out of the total of 450 (=75*6) item responses.
  
  - Rasch model (JMLE/WINSTEPS) 17 (4%)
  - Rasch/1PL (MMLE/irtoys) 33 (7%)
  - 2PL and 3PL (MMLE/irtoys) 48 (11%)
Bayesian Regression: Nonparametric and Parametric Models

\[ y|x; \theta \sim f(y|x; \theta) \]
\[ \theta \sim \pi(\theta) \]

U.S. Copyright to George Karabatsos
Professor, University of Illinois-Chicago
Supported by NSF grant SES-1156372

Data File: C:\Users\George\Desktop\ORVOMSWAEP75.DAT

| OrigRowID1 | y | ID | OrigRowID1=1 | OrigRowID1=2 | OrigRowID1=3 | OrigRowID1=4 | OrigRowID1=5 | OrigRowID1=6 | OrigRowID1=7 | OrigRowID1=8 | OrigRowID1=9 | OrigRowID1=10 | OrigRowID1=11 | OrigRowID1=12 | OrigRowID1=13 | OrigRowID1=14 | OrigRowID1=15 | OrigRowID1=16 | OrigRowID1=17 | OrigRowID1=18 | OrigRowID1=19 | OrigRowID1=20 | OrigRowID1=21 | OrigRowID1=22 |
|------------|---|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1          | 1 | 0  | 1           | 0           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 2          | 1 | 0  | 4           | 1           | 0           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 3          | 1 | 0  | 4           | 1           | 0           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 4          | 1 | 1  | 4           | 1           | 0           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 5          | 1 | 0  | 4           | 1           | 0           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 6          | 1 | 0  | 4           | 1           | 1           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 7          | 1 | 0  | 4           | 1           | 1           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 8          | 1 | 0  | 4           | 1           | 1           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 9          | 1 | 0  | 4           | 1           | 1           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| 10         | 1 | 0  | 4           | 1           | 1           |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |

- **Specify New Model**
- **Open Model**
- **Clear**
- **Run Posterior Analysis**
  - MC Samples: 20000
  - Burn In: 2000
  - Thin: 5
- **Posterior Summaries**
- **Posterior Predictive**
- **Dependent Variable (Y):**
- **Covariates (x):**
Bayesian Regression: Nonparametric and Parametric Models

\[ y|x; \theta \sim f(y|x; \theta) \]
\[ \theta \sim \pi(\theta) \]

U.S. Copyright to George Karabatsos
Professor, University of Illinois-Chicago
Supported by NSF grant SES-1156372
Select model for data analysis
(one of 43 possible choices of models)
Select dependent variable (y). Item response data.
Select examinee indicator (0,1) and item indicator (-1,1) covariates/predictors.
Enter prior parameters of the model.
Software now displays actual model with chosen prior parameters, and chosen covariates.
Click to Run Analysis: Generate 50K MCMC samples from model’s posterior distribution.
After end of analysis run (after ~2 min), this text output file is generated and opened automatically.

Text display of model.
Prior Parameters of Model

b_s_mu  = 5
v_1    = 100
v_0    = 0.1
w_gamma = 1
a_0    = 1
v      = 1
a_w    = 0.001

Text display of chosen prior parameters of model.
Binary infinite homoscedastic probits regression model with SSVS

\[ y_i \mid x_i \sim f(y \mid x_i), \quad i = 1, \ldots, n \]

\[ f(y \mid x) = \theta_x (1 - \theta_x)^{1 - y} \]

\[ \theta_x = \int \sum_{j=1}^{\infty} n(y^x \mid \mu_j + x^T \beta, \sigma^2) \omega_j(x) dy^x \]

\[ \omega_j(x) = \Phi \left( \frac{j - x^T \beta_\omega}{\exp(x^T \lambda_\omega)^{1/2}} \right) - \Phi \left( \frac{j - 1 - x^T \beta_\omega}{\exp(x^T \lambda_\omega)^{1/2}} \right) \]

\[ \gamma \sim \text{Ber}(w_k), \quad k = 1, \ldots, p; \quad \sigma^2 \sim \text{IG}(a_0/2, a_0/2) \]

\[ \beta_\omega, \sigma^2 \sim \text{N}(0, \sigma^2 \sigma_\omega^2) \]

Click to generate additional plots and output tables.
Convergence Analysis Step #1:
Click table to verify that 95% sizes Monte Carlo Confidence Intervals (MCCI) half-widths, for parameter estimates of interest, are sufficiently small (e.g., around .01).

Convergence Analysis Step #2:
Click to verify that Trace Plots of model parameters of interest display good mixing, i.e., look sufficiently “hairy”.
Convergence Analysis Step #1: Click table to verify that 95% sizes Monte Carlo Confidence Intervals (MCCI) half-widths, for parameter estimates of interest, are sufficiently small (e.g., around .01).

ANSWER: As seen below, the 95% MCCI half widths are small. You can make them even smaller by running additional MCMC sampling iterations, over and beyond the 50K MCMC iterations already run.
Convergence Analysis Step #2: Click to verify that Trace Plots of model parameters of interest display good mixing, i.e., look sufficiently "hairy".

ANSWER: Trace plots look hairy and stable, for the ability parameters of the first 2 examinees (top 2 panels), and for all 6 test items (bottom 6 panels). Recall that we have chosen to remove the first 2K burn-in samples, for posterior parameter estimation.
After verifying MCMC convergence, you may then generate plots of marginal posterior summaries of model parameters. Such summaries are already provided in greater detail, in the text output files mentioned earlier.
Box plot: marginal posterior distribution of ability parameter, for each of the 75 examinees.
Box plot: marginal posterior distributions of difficulty parameter, for each of the 6 test items of the NAEP exam.
Plot the standardized fit residuals of the BNP Rasch model.
Posterior Predictive Model Fit Statistics (from text output file)

<table>
<thead>
<tr>
<th>Stat.</th>
<th>95% MCCIhw</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ squared</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Standardized Residuals of Dependent Variable Responses:

<table>
<thead>
<tr>
<th>Min</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.53</td>
<td>-0.34</td>
<td>-0.25</td>
<td>-0.08</td>
<td>0</td>
<td>0.09</td>
<td>0.33</td>
<td>0.25</td>
<td>0.63</td>
</tr>
</tbody>
</table>

(Range of 95% MCCI sizes of residuals: [.00 , .00]).

Zero outliers under the BNP Rasch model.

But then again, as mentioned, model fit analysis is virtually unnecessary for this Rasch model.
Conclusions

• We propose a BNP Rasch model, which retains the interpretability of the ordinary Rasch model, while providing examinee ability and item difficulty estimates that are robust to outliers of the ordinary model.

• Model accounts for all covariates not included in the model.

• Therefore, the BNP Rasch model is not “wrong.”

• Then, there is virtually no point in performing fit analysis with the model.

• BNP model provides a “Rasch analysis without fit statistics.”

• BNP Rasch model can be easily extended to ordinal or continuous item responses, and can handle extra covariates as in FACETS.

• Free, user-friendly software is available for the BNP model http://www.uic.edu/~georgek/HomePage/BayesSoftware.html
References


