Monte Carlo tests of the Rasch model based on scalability coefficients

KB Christensen
S Kreiner

Dep. of Biostatistics, Univ. of Copenhagen
Item response theory (IRT)
Several items measuring a single unidimensional latent variable (trait, ability, ...). Items are categorical, often dichotomous.

IRT models can be parametric or nonparametric. Used for measurement and for validation.
The Rasch model\textsuperscript{1,2} simplest parametric IRT model (1PL).

\begin{equation}
Pr(X_i = 1|\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)}
\end{equation}

Implies requirements: unidimensionality, absence of differential item functioning, local independence, and equal logistic item discrimination.

Used for validation.

Is equal discrimination necessary

1PL

2PL
Rasch advocated, but never implemented, exact conditional inference: Study distribution of test statistic independent of person and item parameters.

Computing \(p\)-values feasible using MCMC sampling techniques.
Distribution of data matrix $X = (X_{vi})_{v=1,...,N,i=1,...,I}$ determined by $\theta = (\theta_1, \ldots, \theta_N)$ and $\beta = (\beta_1, \ldots, \beta_I)$.

$R_v = X_{v.} = \sum_j X_{vi}$ and $S_i = X_{.i} = \sum_i X_{vi}$ sufficient.

$P(X|R, S)$ uniform distribution over (very large) set $\mathcal{N}(R, S)$ of all item response matrices with these margins.
Exact tests

$X_0$ observed data, $R_0, S_0$ observed margins, $T = T(X)$ any test statistic, $p$-value defined as conditional probability

$$Pr(T(X) \geq T(X_0)|R_0, S_0) = \frac{\sum_{X \in \mathcal{N}(R,S)} 1(T(X) \geq T(X_0))}{\left[\frac{R}{S}\right]}.$$  \hspace{1cm} (2)

calculate $T(Y)$ for all $Y \in \mathcal{N}(R, S)$ count how many are larger than $T(Y_0)$. This is not feasible, but ...
... MCMC estimates of $p$-values can be computed.

Simple technique based swapping of item responses\textsuperscript{9–12}

\[
\begin{bmatrix}
\vdots & \vdots & & \\
\vdots & 1 & \cdots & 0 & \cdots \\
\vdots & & \ddots & & \vdots \\
\vdots & \cdots & \cdots & 0 & \cdots \\
\vdots & & & 1 & \cdots \\
\vdots & & & & \ddots
\end{bmatrix}
\mapsto
\begin{bmatrix}
\vdots & \vdots & & \\
\vdots & 0 & \cdots & 1 & \cdots \\
\vdots & & \ddots & & \vdots \\
\vdots & \cdots & \cdots & 1 & \cdots \\
\vdots & & & 0 & \cdots \\
\vdots & & & & \ddots
\end{bmatrix}.
\]

Nonparametric IRT

Mokken model of (double) monotonicity\(^3-6\): unidimensionality, local independence, nondecreasing (and nonintersecting) item response functions.

Rasch model is a special case.

Idea: Use nonparametric IRT to test Rasch model.

Scalability coefficients summarize number of Guttman errors\(^7\):

wrong answer to an easy item, correct answer to difficult item.

Guttman errors. Ordering persons and items.

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
\vdots & \vdots \\
1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]
Marginal probabilities

\[ \pi_i(x) = \int Pr(X_i = x|\theta) dP(\theta) \]

ordering \( \pi_1(1) \geq \pi_2(1) \geq \ldots \geq \pi_I(1) \). For item pairs

\[ \pi_{ik}(x_i, x_k) = \int Pr(X_i = x_i, X_k = x_k|\theta) dP(\theta). \]

Probability of Guttman error \( e_{ik} = \begin{cases} 
\pi_{ik}(1, 0) & \text{if } i < k \\
\pi_{ik}(0, 1) & \text{if } i > k 
\end{cases} \)

Expected (marginal independence) \( e_{ik}^{(0)} = \begin{cases} 
\pi_i(1 - \pi_k) & \text{if } i < k \\
(1 - \pi_i)\pi_k & \text{if } i > k 
\end{cases} \).
Item coefficient $H_i$ (based on Loevingers $H^{13,14}$)

$$H_i = 1 - \frac{\sum_{k \neq i} e_{ik}}{\sum_{k \neq i} e_{ik}^{(0)}}$$

Expected values, CI’s and $p$-values from the exact conditional distribution of $H_i$ given observed margins $R_0, S_0$.

Note: Too few Guttman errors also violation of Rasch model.

$H_i$ large $\leftrightarrow$ few Guttman errors $\leftrightarrow$ item discrimination?

Scalability coefficient applied to transposed data matrix yields test of intersecting item response functions.\textsuperscript{15} Observed percentage of persons with negative $H_v^T$ values compared to expected.

Total scalability coefficients $H$ and $H^T$ cannot be used for exact test of the Rasch model (because the total number of Guttman errors is invariant under switches).

Danish study of mobility in elderly, 731 70-year old. $H_i$ values

<table>
<thead>
<tr>
<th>Item</th>
<th>Obs.</th>
<th>Exp. (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking Indoors</td>
<td>0.856</td>
<td>0.867 (0.848, 0.881)</td>
</tr>
<tr>
<td>Walking outside nice weather</td>
<td>0.874</td>
<td>0.835 (0.822, 0.849)</td>
</tr>
<tr>
<td>Walking outside poor weather</td>
<td>0.871</td>
<td>0.821 (0.806, 0.835)</td>
</tr>
<tr>
<td>Walking on stairs</td>
<td>0.751</td>
<td>0.833 (0.819, 0.847)</td>
</tr>
<tr>
<td>Getting outside</td>
<td>0.870</td>
<td>0.853 (0.832, 0.870)</td>
</tr>
<tr>
<td>Getting up</td>
<td>0.864</td>
<td>0.892 (0.864, 0.816)</td>
</tr>
</tbody>
</table>
MCMC estimates of $P(H_i \leq H_{i,obs}|R_0, S_0)$ and $P(H_i \geq H_{i,obs}|R_0, S_0)$, cf. (2), constitute strong evidence against the model.

Note: Multiple testing - false detection rate controlled.\textsuperscript{16}

No evidence against double monotony: 0.137\% of persons have negative $H^T_v$ (expected percentage is 0.065\%, $p = 0.396$), Birnbaum model should not uncritically be chosen. Local dependence likely to be cause of misfit.

Testing Rasch assumption of equal item discrimination.

Sample correlation $d_i = V(X_i, R)/\sqrt{V(X_i)V(R)}$ traditional measure of item discrimination.\(^{17}\) Chen & Small (2005) propose

$$Y = \sum_{i=1}^{I} \frac{[d_i - E(d_i|R, S)]^2}{V(d_i|R, S)}$$

(3)

as intuitive test statistic - more powerful than $Q_1^{18}$ and $R_{1c}^{19}$ tests statistics.

Simulation study.

$\theta$'s for 250 persons (standard normal distribution), 100 tables Birnbaum items (difficulty zero, discriminations 1.0, 1.1, 1.2, 1.3, 1.4, 1.5).

Comparing $H_i$ and $H^T$ to (3)

$H_i$  Rejection rate 31%
$H^T$  Rejection rate 7%
$Y$    Rejection rate 16% [Chen & Small, 2005, Table 6]