Part 6: Testing the Rasch Model (II)

Martin-Löf Test and some nonparametric (‘exact’) Tests

Martin-Löf Test

evaluate unidimensionality
test whether two sets of items form a Rasch scale

Idea:
similar to Andersen's LR Test – focus now on items
the items are partitioned into two subsets of k_1 and k_2 items

$$LR_{ML} = 2\ln\left(\frac{L_c}{L_c^{(1)} \cdot L_c^{(2)}} \cdot q\right) \qquad q = \frac{\prod_r (\frac{n_r}{N})^{n_r}}{\prod_{\boldsymbol{r}} (\frac{n_{\boldsymbol{r}}}{N})^{n_{\boldsymbol{r}}}}$$

product of likelihoods for subsets the same as for whole set

$\boldsymbol{r} = (r_1, r_2)'$, score patterns on the two subtests ($r_1 = 0, \ldots, k_1$ and $r_2 = 0, \ldots, k_2$), $n_{\boldsymbol{r}}$ is number of persons obtaining pattern $\boldsymbol{r}$
$r = r_1 + r_2$ (sum score), n_r accordingly

Martin-Löf Test

$$LR_{ML} = 2\left(\sum_{\boldsymbol{r}} n_{\boldsymbol{r}} \ln\left(\frac{n_{\boldsymbol{r}}}{N}\right) - \sum_r n_r \ln\left(\frac{n_r}{N}\right) - \ln L_c + \ln L_c^{(1)} + \ln L_c^{(2)}\right)$$

LR_{ML} has an asymptotic χ^2-distribution with $df = k_1 k_2 - 1$

```
> MLoef(rmod, splitcr = "median")
Martin-Loef-Test (split criterion: median)
LR-value: 8.009
Chi-square df: 8
p-value: 0.433

> MLoef(rmod, splitcr = c(1, 1, 1, 1, 0, 0))
Martin-Loef-Test (split criterion: user-defined)
LR-value: 7.934
Chi-square df: 7
p-value: 0.338
```


Nonparametric (‘exact’) Tests

Idea:

Parameter estimates depend only on the marginal totals $\boldsymbol{r}$ and $\boldsymbol{s}$
if the Rasch model fits the data, all binary matrices with the same marginals are equally likely

for any statistic of the data matrix, one can approximate the null distribution (i.e., the distribution if the Rasch model is valid)

take a random sample from the collection of equally likely data matrices
construct the observed distribution of the statistic

one can then simply determine the exceedence probability of the statistic in the observed sample (its p-value)
and thus construct a nonparametric test of the Rasch model

Nonparametric Tests (cont'd)

sample space: $\boldsymbol{\Sigma}_{rs}$ (all possible matrices with fixed $\boldsymbol{r}$ and $\boldsymbol{s}$)

the distribution is uniform over of $\boldsymbol{\Sigma}_{rs}$
for each data matrix X, $X \in \boldsymbol{\Sigma}_{rs}$: $p(X) = \frac{1}{\#\boldsymbol{\Sigma}_{rs}}$

$\#\boldsymbol{\Sigma}_{rs}$ can be huge
for a 12 × 12 table with all $r = s = 2$ this is 21,959,547,410,077,200

algorithmic difficulty is to draw uniform random samples

three approaches:
– Ponocny(2001) (simple Monte Carlo)
– Chen(2005) (importance sampling)
– Verhelst, Hatzinger & Mair (2007) (MCMC, in eRm)

using eRm: rsampler()

(from package RaschSampler)

```
> rmat <- rsampler(stress, rsctrl(burn_in = 100, n_eff = 100, seed = 123))
> summary(rmat)
Status of object rmat after call to RSampler:
        n  =  100
        k =  6
        burn_in =  100
        n_eff =  100
        step =  16
        seed =  123
        tfixed =  FALSE
        n_tot =  101
        outvec contains 10100 elements
```


Nonparametric Tests

Procedure:

- X_0 observed data matrix
- simulate n Rasch-conform matrices $X_1, \ldots, X_n$ with fixed margins as in X_0
- calculate a test-statistics $T^{(0)}$ for X_0
- calculate $T^{(1)}, \ldots, T^{(2)}$ for $X_1, \ldots, X_n$

Test:

- count number x of the $T^{(\cdot)}$'s which exceed T_0 giving $p = x/n$
- if $p \leq \alpha$ reject H_0

Advantage:
– arbitrary specific tests statistics can be constructed
– valid and powerful, even in small samples
– easily to realise in eRm

Local Dependence

$\mathbf{T_1}$: checks for local dependence via increased inter-item correlations for item pairs: cases are counted with equal responses on both items.

$$T_1(X) = \sum_v \delta_{ij}(x_{vi}, x_{vj}), \quad \delta() = 1 \text{ if } x_{vi} = x_{vj}, 0 \text{ otherwise}$$

$\mathbf{T_2}$: checks for local dependence to detect model deviating subscales via increased dispersion of subscale person rawscores.

$$T_2(X) = \text{Var}_v(r_v^{(\mathcal{S})}), \text{ where } r_v^{(\mathcal{S})} = \sum_{i \in \mathcal{S}} x_{vi}$$

pattern	r
10000	1
11000	2
11100	3
11110	4
11110	4

$Var(r) > Var(r')$

pattern	r'
10100	2
10011	3
01100	2
01101	3
10101	3

using eRm: NPtest()

$\mathbf{T_1}$:

```
> t1 <- NPtest(stress, n = 100, method = "T1")
> t1
Nonparametric RM model test: T1 (local dependence - increased
     inter-item correlations)
    (counting cases with equal responses on both items)

Number of sampled matrices: 100

Number of Item-Pairs tested: 15

Item-Pairs with one-sided p < 0.05

(1,3) (3,5)
 0.01  0.04
```


$\mathbf{T_2}$:

```
> t21 <- NPtest(stress, n = 100, method = "T2", idx = 1:3)
> t21
Nonparametric RM model test: T2 (local dependence - model deviating
     subscales)
    (dispersion of subscale person rawscores)

Number of sampled matrices: 100
Items in subscale: 1 2 3
Statistic: variance
one-sided p-value: 0.09
> t22 <- NPtest(stress, n = 100, method = "T2", idx = c(1, 2, 6),
+     stat = "mad1")
> t22
Nonparametric RM model test: T2 (local dependence - model deviating
     subscales)
    (dispersion of subscale person rawscores)

Number of sampled matrices: 100
Items in subscale: 1 2 6
Statistic: mean absolute deviation
one-sided p-value: 0.77
```


DIF

$\mathbf{T_4}$: checks for group anomalies (DIF) via too high (low) raw scores on item(s) for specified group

$$T_4(X) = \sum_{v \in \mathcal{G}} x_{vi} \text{ where } \mathcal{G} \text{ is any subgroup}$$

Unequal Discrimination

$\mathbf{T_7}$: checks for lower discrimination (2PL) in item subscale via counting cases with response 1 on more difficult and 0 on easier items.
The test is global (a single statistic) for the subscale

$$T_7(X) = -\sum_{i,j \in \mathcal{S}} \delta(x_i < x_j \wedge s_i > s_j)$$

using eRm: NPtest()

$\mathbf{T_4}$:

```
> age <- sample(20:90, 100, replace = TRUE)
> age <- age < 30
> t41 <- NPtest(stress, n = 100, method = "T4", idx = 1:3, group = age)
> t41
Nonparametric RM model test: T4 (Group anomalies - DIF)
    (counting high raw scores on item(s) for specified group)

Number of sampled matrices: 100
Items in Subscale: 1 2 3
Group: age   n = 10
one-sided p-value: 0.72
```

$\mathbf{T_7}$:

```
> t7 <- NPtest(stress, n = 100, method = "T7", idx = 1:3)
> t7
Nonparametric RM model test: T7 (different discrimination - 2PL)
    (counting cases with response 1 on more difficult and 0 on easier item)

Number of sampled matrices: 100

Item Scores:
 1  2  3
47 42 41
one-sided p-value: 0.95
```

these are just a few examples (we will see others later)

Ponocny (2001) gives a systematic and various more examples