Quasi-Exact Tests for the dichotomous Rasch Model

Ingrid Koller, Vienna University
Reinhold Hatzinger, WU-Vienna

Rasch model

Properties

- unidimensionality/homogenous items
- conditional independence (local independence)
- specific objectivity/sample independence
- strictly monotone increasing item characteristic function
- sufficient statistics

Why quasi-exact tests?

Parametric methods need large samples because of

- consistency and unbiasedness of parameter estimates
- assumption of asymptotic distribution of test statistics
- higher power of test-statistics

Small samples in practise

- large samples often not available (e.g., clinical studies)
- complex study designs (e.g., experiments)
- smaller costs and less time-consuming
- possibility to test the quality of items also in small samples (e.g., stepwise test-construction)

Quasi-exact tests?

with quasi-exact tests it is possible to test the Rasch-model (RM) also with small samples

Sampling binary matrices

description of MCMC method: Kathrin Gruber

development of test-statistics (T) for the dichotomous RM

- Ponocny (1996, 2001)
- Chen & Small (2005)
- Verhelst (2008)
- Koller & Hatzinger (in prep.)

General procedure for the T-statistics

- $\mathbf{A}_0$ is the observed matrix with the margins r_v and c_i where $r_v = \sum_i x_{vi}$ (person score) and $c_i = \sum_v x_{vi}$ (item score)
- Σ_{rc} is the set of all matrices with fixed r and c (sample space)

Algorithm

- sample $s = 1, \ldots, S$ matrices $\mathbf{A}_s$ from Σ_{rc}
- calculate T_0 for the observed matrix $\mathbf{A}_0$
- calulate $T_1, \ldots, T_S$ for all sampled matrices $\mathbf{A}_1, \ldots, \mathbf{A}_S$
- determine your p-value by

$$p = \sum_{s=1}^{S} t_s / S \qquad \text{where } t_s = \begin{cases} 1, & T_s(\mathbf{A}_s) \geq T_0(\mathbf{A}_0) \\ 0, & \text{else} \end{cases}$$

Conditional dependence

T_{11}: large inter-item correlations

$$T_{11}(\mathbf{A}) = \sum_{ij} |r_{ij} - \tilde{r}_{ij}| \qquad \text{where} \qquad \tilde{r}_{ij} = \frac{\sum_{s=1}^{S} r_{ij}}{S}$$

r_{ij} . . . the inter-item-correlation for item i and item j

$\tilde{r}_{ij}$ . . . mean of r_{ij} from all simulated matrices

$$p = \sum_{s=1}^{S} t_s / S \qquad \text{where} \qquad t_s = \begin{cases} 1, & T_s(\mathbf{A}_s) \geq T_0(\mathbf{A}_0) \\ 0, & \text{else} \end{cases}$$

if r_{ij} in $\mathbf{A}_0$ is large, then the difference $r_{ij} - \tilde{r}_{ij}$ is also large
only a few T_s show the same or a higher difference than T_0

highly correlated items indicate violation of conditional independence

Multidimensionality

T_{11m}: small inter-item-correlations

same equation as for T_{11}, but modified test:

$$p = \sum_{s=1}^{S} t_s / S \qquad \text{where} \qquad t_s = \begin{cases} 1, & T_s(\mathbf{A}_s) \leq T_0(\mathbf{A}_0) \\ 0, & \text{else} \end{cases}$$

if r_{ij} in $\mathbf{A}_0$ is small, then the difference $r_{ij} - \tilde{r}_{ij}$ is also small
only a few T_s show the same or a smaller difference than T_0

small correlations between items indicate multidimensionality

Conditional dependence

T_1: many equal responses

- count the number of $\{00\}$ and $\{11\}$ patterns in items i and j
- how many T_s have same or a higher value than T_0

$$T_1(\mathbf{A}) = \sum_{v} \delta_{ij} \qquad \text{where} \qquad \delta_{ij} = \begin{cases} 1, & x_{vi} = x_{vj} \\ 0, & x_{vi} \neq x_{vj} \end{cases}$$

many equal responses indicate violation of conditional independence

Multidimensionality:

T_{1m}: few equal responses

- how many T_s have same or a lower value than T_0

few equal responses indicate that the correlation between items is too small, unidimensionality assumption may be violated

Learning

$T_{1\ell}$: many {11} patterns (e.g.,Koller & Hatzinger)

- count only {11} patterns as opposed to T_1

$$T_{1l}(\mathbf{A}) = \sum_v \delta_{ij} \qquad \text{where} \qquad \delta_{ij} = \begin{cases} 1, & x_{vi} = x_{vj} = 1 \\ 0 & \text{else} \end{cases}$$

- how many T_s have same or a higher value than T_0

if person has learned from one item ($x_{vi} = 1$) then the probability $p(x_{vj} = 1)$ is increased for a positive reponse to another item j

Conditional dependence

T_2 : high dispersion of rawscore r_v for a set of items

- if items are dependent, the variance of r_v is large
- because of $var(z) = var(x) + var(y) + 2 * cov(x, y)$
- define a set of items I and calculate $r_v^{(I)}$
- count how many $T_s \geq T_0$

$$T_2(\mathbf{A}) = var_v(r_v^{(I)}) \qquad \text{where} \qquad r_v^{(I)} = \sum_{i \in I} x_{vi}$$

other possibilities: range, mean absolute deviation, median absolute deviation.

Multidimensionality:

T_{2m} : low dispersion of rawscore r_v for a set of items

- count how many $T_s \leq T_0$

Multidimensionality

T_{MU}: correlation of rawscore for item subsets

(Koller & Hatzinger)

- if two sets of items I are unidimensional, r_v^I of set I and r_v^J of set J should be positiv correlated
- with increasing r_v^I also r_v^J should be increasing
- count the number of correlations $T_s \leq T_0$

$$T_{MU}(\mathbf{A}) = cor(r_v^I, r_v^J) \qquad = \qquad r_v^I = \sum_{i \in I} x_{vi}$$

Subgroup-invariance

T_{10}: based on counts on certain item responses

- n_{ij}/n_{ji} is proportional to the ratio of $\exp(\beta_i)/\exp(\beta_j)$
- no parameter differences for focal-group foc and reference-group ref: $n_{ij}^{ref}/n_{ji}^{ref} = n_{ij}^{foc}/n_{ji}^{foc}$
- sum of differences for all pairs of items
- counts of $T_s \geq T_0$

$$T_{10}(\mathbf{A}) = \sum_{ij} |n_{ij}^{ref} n_{ji}^{foc} - n_{ji}^{ref} n_{ij}^{foc}|$$

- if the parameter differ across groups, the difference should be increasing

Note:

- external criterion (e.g., gender): uniform DIF
- internal criterion (e.g., rawscore-median): discrimination, guessing, falsity
- split on specified item: conditional dependence

Subgroup-invariance

T_4:counts of positive responses in person subgroups

- assumption: in one group of persons G one or more items are easier/more difficult as expected in the RM
- count the number of persons who solved these items
- easier: counts of $T_s \geq T_0$
- more difficult:counts of $T_s \leq T_0$

$$T_4(\mathbf{A}) = \sum_{v \in G} x_{vi}$$

Note:

- tests the same assumptions as T_{10}

Subgroup-invariance

T_{DTF}: based on item differences on item sumscores (Koller & Hatzinger)

- similar to T_4, but with the possibility to test DTF (all items in a test shows subgroup-invariance)
- calculate the sumscores (c) for one item (or a group of items) for the reference group c_i^{ref} and for the focal group c_i^{foc}
- calculate the difference of c between focal and reference group.
- easier: counts of $T_s \geq T_0$
- more difficult:counts of $T_s \leq T_0$

$$T_{DTF}(\mathbf{A}) = \sum_{i \in I} (c_i^{ref} - c_i^{foc})^2$$

Note:

- tests the same assumptions as T_{10} & T_4

Unfolding response structure - monotonicity

T_6: responses in three person subgroups

- similar to T_4.
- split the sample in three rawscore groups and count the number of positive responses only in the middle group G_m

- easier (reversed U-shape): counts of $T_s \geq T_0$
- more difficult (U-shape): counts of $T_s \leq T_0$

$$T_6(\mathbf{A}) = \sum_{v \in G_m} x_{vi}$$

Item discrimination

T_5: rawscore for persons with $x_{vi} = 0$ for a certain item

- include persons who answer with 0 to a certain item
- sum r_v of the remaining items for group $x_{vi} = 0$
- counts of $T_s \geq T_0$

$$T_5 = (\mathbf{A}) = \sum_{v | x_{vi} = 0} r_v$$

- if persons with high ability (r_v = high) fail to solve a certain item, this item may show too low discrimination, falsity, or indicate multidimensionality

Constructing new test statistics

- based on substantive considerations.

- based on statistics, where the approximation to the asymptotic distribution is questionable.
  - monotone transformations
    example: point-biserial correlation
  - simplification
    example: Mantel-Haenszel statistic

Example: point-biserial correlation

$$r_{pbis} = \frac{\overline{r}_0 - \overline{r}_1}{s_r}\sqrt{\frac{n_0 n_1}{n(n-1)}} \quad \propto \quad \left(\frac{\sum r_0}{n_0} - \frac{\sum r_1}{n_1}\right) n_0 n_1$$

remove s_r and n (constant)
remove $\sqrt{}$ is a monotone function $(n_0, n_1 \geq 0)$

$$T_{pbis}(\mathbf{A}) = \frac{n_1 \sum r_0 - n_0 \sum r_1}{\cancel{n_0 n_1}} \cancel{n_0 n_1} \quad = \quad n_1 \sum r_0 - n_0 \sum r_1$$

- counts $T_s \geq T_0$
- if persons with high ability (r_v = high) fail to solve a certain item, this item may show too low discrimination, falsity, or indicate multidimensionality

Example: Mantel-Haenszel statistic

tests for conditional independence of two nominal variables across several strata (e.g., $2 \times 2 \times C$ tables)

$$MH = \frac{\left(\sum_c N_{11c} - \sum_c E(N_{11c}|n_c)\right)^2}{Var\left(\sum_c N_{11c}|n_c\right)} \quad \overset{as.}{\sim} \quad \chi^2_{df=1}$$

can be used to test various RM violations.

Verguts & DeBoeck (2001):
sufficiency, unidimensionality, item dependence

Mantel-Haenszel statistic may be simplified to

$$T_{MH} = \left(\sum_c n_{11c} - \sum_c \tilde{n}_{11c}\right)^2 \qquad \text{where } \tilde{n}_{11c} = \sum_{s=1}^{S} n_{11c}/S$$

Implementation in R: eRm & RaschSampler

some statistics in eRm

- subgroup-invariance
  - T_{10} based on counts on certain item responses (global test)
  - T_4 counts of positive responses in person subgroups (test on item level)
- conditional independence
  - T_{11} large inter-item correlations (global test)
  - T_1 many equal responses (test on item level)
  - T_2 high dispersion of rawscore r_v for a set of items (test on item level)

RaschSampler

- supply user defined function for arbitrary T-statistics
  Verhelst, Hatzinger, & Mair (2007)

Example: conditional dependence

T_1: many equal responses (test on item level)

```
> library(eRm)
> library(RaschSampler)

> t1 <- NPtest(raschdat1,n=500,burn_in=500,step=32,seed=123,method="T1")
> print(t1,alpha=0.05)

Nonparametric RM model test: T1 (local dependence -
    increased inter-item correlations)
    (counting cases with equal responses on both items)
Number of sampled matrices: 500
Number of Item-Pairs tested: 435
Item-Pairs with one-sided p < 0.05
  (1,9)  (1,26)  (4,12)  (4,26)  (4,28)  (8,13)  (9,16) (10,24)
  0.030   0.042   0.036   0.036   0.018   0.014   0.010   0.032
(18,28) (21,22) (25,29)
  0.032   0.004   0.002
```