SELECTION OF ITEMS FITTING A RASCH MODEL

ABSTRACT

In a preceeding paper, the authors propose a procedure based on the Multidimensional Marginaly Sufficient Rasch Model (MMSRM) to select items in scales with good measurement properties. The main drawback of this procedure is the time of computing, because the used models are Generalized Linear Mixed Models (GLLM), a kind of models which the parameters are long to estimate with classical methods in the generalist statistical software. In this paper, we propose a fast way to realize this procedure. A Stata module is proposed to perform the two versions of this procedure. Simulations allows comparing this procedure to two others ones, MSP and HCA/CCPROX.

Key Words : Quality of life, Multidimensional IRT, Rasch model, Items selection, Raschfit.

0.1. Introduction

Item Response Models (IRM) [FIS 95] [LIN 97] are used models in educational testing, psychology or health related quality of life. These models consider that a latent trait (latent variable) explains the responses to the items. The latent trait generally is multidimensional, but full of IRM generally consider only a unidimensional latent trait. More, the relations between the items and the component of the latent trait often

———————

Chapitre rédigé par Jean-Benoit HARDOUIN (Laboratory of Biomathematics and Biostatistics - University of Nantes - France - jean-benoit.hardouin@univ-nantes.fr) et Mounir MESBAH (Laboratory of Theoretical and Applied Statistics - University Pierre and Marie Curie (Paris 6) - France).

are unknown : they are supposed by subjective advices of experts and the work of the statistician consists to validate them with psychometric models. The statistician generally is not integrated in the exploratory analysis to define these links and to help the experts to define links which are coherent with psychometric properties.

Some procedures (Factor analysis, MSP, HCA/CCPROX) allow the links between the items and the latent traits to be defined, but none of them is based on the direct fit of an Item Response Models to the data. The authors propose [HAR 04] a procedure which allows defining these links and obtaining scales with a good fit of a given Item Response Model : the Multidimensional Marginaly Sufficient Rasch Model (MMSRM) which is a multidimensional counterpart of the most famous IRM, the Rasch model.

But the fit of a multidimensional IRM, evaluated by the likelihood of the model, is a long process to run if we consider the model as a Multidimensional Generalized Linear Mixed Model (GLMM) : in this paper, a fast way to realize this procedure is proposed. It allows obtaining correct results in a reasonable time.

0.2. Notations and assumptions

0.2.1. *Notations*

Let Θ_q the qth component of the multidimensional latent trait characterizing the individuals with $q = 1, ..., Q$ and θ_{nq} the realization of this latent trait for the nth individual, $n = 1, ..., N$. $\boldsymbol{\theta}_n$ is the vector of the values on the Q latent traits for the nth individual $(\theta_{n1}, ..., \theta_{nq}, ..., \theta_{nQ})$.

The jth item is characterized by a vector of parameters $\boldsymbol{\nu}_j$, $j = 1, ..., J$ [FIS 95]. The response to this item is represented by the random variable X_j wich the realization for the nth individual is noted x_{nj}.

We consider only dichotomous items, and for each of them, the more favourable answer is named "positive response" and is coded 1 and the other answer is named "negative response" and is coded 0.

The Item Response Function (IRF) of the jth item is the probability that a given individual n positively respond to this item as a function of the value of the latent trait for the nth individual $\boldsymbol{\theta}_n$.

0.2.2. *Fundamental assumptions of the Item Response Theory (IRT)*

The IRT is the set of IRM which verify three fundamental assumptions [LIN 97] :

– Fixed dimension : the dimension Q of the latent trait is known. For the majority of IRM, the unidimensionality ($Q = 1$) is required.

– Local independency : the responses to the items are independent conditionnally to the latent trait.

– Monotonicity : the IRF are non-decreasing functions of each component of the latent trait.

0.3. The Rasch model and the multidimensional marginally sufficient Rasch model

0.3.1. *The Rasch model*

The Rasch model [RAS 60] is an unidimensional IRM : the responses to the items are assumed to depend of an unidimensional latent trait : θ_n is a scalar. More, each item j, $j = 1, ..., J$, is defined by only one parameter δ_j : this parameter is interpreted as the difficulty of the jth item, because the more its value is high, the more the probability to positively respond to the item j is small. The IRF of the jth item is :

$$P(X_{nj} = x_{nj}/\theta_n; \delta_j) = \frac{\exp\left(x_{nj}\left(\theta_n - \delta_j\right)\right)}{1 + \exp\left(\theta_n - \delta_j\right)} \qquad [1]$$

The latent trait can be considered as a set of fixed parameters or as a random variable. In the fixed effects Rasch model, the classical maximum likelihood technic makes the estimations not consistent.

The Rasch model is a famous IRM because this model has a specific property : the score $S_n = \sum_{j=1}^{J} X_{nj}$ is a sufficient statistic of the latent trait, that is to say that all the available information about the latent trait is contained in the score [AND 77]. Consequently, if the latent trait is considered as a set of fixed parameters, the conditional maximum likelihood can be used : the likelihood is maximized conditionnally to the score computed as the number of positive responses to all the items for each individual. These estimations are consistent [AND 70].

If the latent trait is considered as a random variable, its distribution function $G(\theta)$ is assumed (generally as a centered gaussian distribution of variance σ^2 [FIS 95]), and

consistent estimations of the items parameters δ_j and of the parameters of this distribution (generally, only the variance σ^2) can be obtained by maximizing the marginal likelihood :

$$L_M(\sigma^2, \boldsymbol{\delta}/\boldsymbol{x}) = \prod_{n=1}^{N} \int \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta; \delta_j)G(\theta/\sigma^2)d\theta \qquad [2]$$

with $\boldsymbol{\delta} = (\delta_1, ..., \delta_j, ..., \delta_J)$.

0.3.2. *The multidimensional marginally sufficient Rasch model*

Hardouin and Mesbah [HAR 04] propose a extention of the Rasch model to the multidimensional case. In this model, the responses to the items are governed by a multidimensional latent trait (of dimension Q), but the response to a given item j is governed by only one component of the latent trait indexed q_j. Consequently, marginally to the others items and to the others components of the latent trait, all the items governed by the same component of the latent trait (that is to say each scale) fit a classical Rasch model. This model can be relied to the between items response model defined by Adams and al. [ADA 97].

In this model, the response function of the jth item is :

$$P(X_{nj} = x_{nj}/\boldsymbol{\theta}_n; \delta_j) = \frac{\exp\left(x_{nj}\left(\theta_{nq_j} - \delta_j\right)\right)}{1 + \exp\left(\theta_{nq_j} - \delta_j\right)} = P(X_{nj} = x_{nj}/\theta_{nq_j}; \delta_j) [3]$$

As a consequence, the score $S_{nq} = \sum_{j=1/q_j=q}^{J} x_{nj}$ computed with only the items associated to the qth component of the latent trait Θ_q is a sufficient statistic of Θ_q.

This model is named a Multidimensional Marginaly Sufficient Rasch Model (MM-SRM) for this reason. By considering the latent trait as a multivariate random variable (distributed by a multivariated centered gaussian distribution with a unknown covariance matrix $\boldsymbol{\Sigma}$), the items parameters (and the elements of the $\boldsymbol{\Sigma}$ matrix) can be consistently estimated by marginal maximum likelihood, by maximizing :

$$L_{M1}(\boldsymbol{\Sigma}, \boldsymbol{\delta}/\boldsymbol{x}) = \prod_{n=1}^{N} \int \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta_{q_j}; \delta_j)G(\boldsymbol{\theta}/\boldsymbol{\Sigma})d\boldsymbol{\theta} \qquad [4]$$

This method of estimation is a long process with classical software as SAS, Splus or Stata, because these integrals must be approximated at each step of the algorithm.

We can gain full of time by estimating independently the items parameters and the elements of the Σ matrix : like each marginal scale verifies a Rasch model, the items parameters can be estimated scale by scale, by the maximum marginal likelihood method. These estimations are noted $\hat{\delta}_j$, $j = 1, ..., J$.

Then, the elements of the covariance matrix Σ can be estimated by assuming the items parameters as known parameters and by maximizing the new quantity :

$$L_{M2}(\Sigma/\boldsymbol{x}, \hat{\boldsymbol{\delta}}) = \prod_{n=1}^{N} \int \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta_{q_j}; \hat{\delta}_j)G(\boldsymbol{\theta}/\Sigma)d\boldsymbol{\theta} \qquad [5]$$

which is an approximation of $L_{M1}(\Sigma, \boldsymbol{\delta}/\boldsymbol{x})$.

0.4. The Raschfit procedure

In [HAR 04], a procedure of items selection in Rasch scales based on the fit of the items to a MMSRM is proposed. This procedure is referenced in the present paper by "Raschfit".

At a step k of this procedure, we assume at each step to have a set of $J^{(k)}$ items, named kernel, which verifies a Rasch model, and we search if a Rasch model fits the data by adding a new item (indexed by $j = 0$) to the kernel. We compare the fit of a Rasch model, and the fit of the MMSRM with the items of the kernel relied to one component of the latent trait and the new item relied to another component.

At the first step of the Raschfit procedure, the initial kernel is composed of two or more items chosen by the user, or determined by a specific analysis. The order in wich the others items will be introduced in the procedure can be freely determined but can be have importance in the final result. The authors propose to use the Mokken Scale Procedure (MSP) [HEM 95] to select the initial kernel and to order the others items

The fit of the models is valuated by the Akaike Information Criterion (AIC) [HOI 97] with :

$$AIC_m = -2l_m + 2K_m \qquad [6]$$

Where l_m is the value of the log-likelihood and K_m the number of parameters of the model m. At the kth step of the procedure, we have $K_1^{(k)} = J^{(k)} + 2$ for the Rasch model and $K_2^{(k)} = J^{(k)} + 4$ for the MMSRM.

The new item is selected at the kth step of the procedure in the scale if $AIC_1^{(k)} \leq AIC_2^{(k)}$. The procedure is stopped when there is no more item remaining.

This procedure is a very long process, because the log-likelihood of a MMSRM is long to approximate. The estimation can be implemented with the GLLAMM program of Stata, the NLMIXED procedure of SAS or the NLME library of Splus. This three programs approximate the multivariate integrals with (adaptive) gaussian quadratures and necessitate a great amount of computer ressources.

0.5. A fast version of Raschfit

We propose here an adaptation of the Raschfit procedure, referenced as "Raschfit-Fast". This adaptation is based on the fixed effects Rasch model.

0.5.1. *Estimation of the parameters under the fixed effects Rasch model*

In the Rasch model, by considering the latent trait as a set of fixed parameters (θ_n, $n = 1, ..., N$), the individuals values of the latent trait θ_n, $n = 1, ..., N$ cannot be constistently estimated by classical technic [FIS 95].

Indeed, the Rasch model verifies the specific property of sufficiency of the unweighted score on the latent trait : this property signifies that, conditionnally to the score $s_n = \sum_{j=1}^{J} x_{nj}$, the likelihood of the nth individual is independent of the latent trait θ_n.

$$L_{Cn}(\theta_n, \boldsymbol{\delta}/\boldsymbol{x_n}, s_n) = L_{Cn}(\boldsymbol{\delta}/\boldsymbol{x_n}, s_n) \tag{7}$$

In maximizing the quantity

$$L_C(\boldsymbol{\delta}/\boldsymbol{x}, \boldsymbol{s}) = \prod_{n=1}^{N} L_{Cn}(\boldsymbol{\delta}/\boldsymbol{x_n}, s_n), \tag{8}$$

we obtain a consistent estimation of the $\boldsymbol{\delta}$ vector of parameters.

The weighted maximum likelihood technics [FIS 95] allows consistently estimating the $\theta_s, \ s = 0, ..., J$ parameters by maximizing :

$$L_{WC}(\boldsymbol{\theta_s}/\hat{\boldsymbol{\delta}}, \boldsymbol{x}) = \prod_{n=1}^{N} L_n(\theta_{s_n}/\hat{\boldsymbol{\delta}}, \boldsymbol{x_n})g(\theta_{s_n}) \tag{9}$$

where $\boldsymbol{\theta_s} = (\theta_0, ..., \theta_s, ..., \theta_J)'$ and

$$g(x) = \prod_{j=1}^{J} \frac{\exp(x - \hat{\delta}_j)}{\left(1 + \exp(x - \hat{\delta}_j)\right)^2} \tag{10}$$

0.5.2. *Principle of Raschfit-Fast*

The principle of Raschfit-Fast is globally the same than for Raschfit : at each step, a kernel fits a Rasch model, and a new item is added to this kernel if the new scale has a good fit to a Rasch model. The mixed Rasch model is replaced by a fixed effects Rasch model, and the MMSRM is replaced by a specific model built from the following consideration : if the set of items composed of the items of the kernel and the new item does not follow a Rasch model, we consider than the responses to the new item are independent of the latent trait. The likelihood associated to these responses is estimated by a logistic form (as in the Rasch model) but with only an unknow parameter characterizing the item in the linear composant (which can be interpreted as a difficulty parameter).

At a step k of the algorithm, let 0 the index of the new item, the J_k items of the kernel are indexed from 1 to J_k.

0.5.3. *A model where the new item is explained by the same latent trait than the kernel*

If the new item is explained by the same latent trait than the kernel, a Rasch model can be used. At the kth step of the algorithm, the score $S_n^{(k*)}$ is computed with the J_k

items of the kernel and the new item.

The likelihood used to compute the Akaike Information Criterion is in this case similar to this one presented in the equation 9, and the number of parameters is $K_1^{(k)} = 2J_k + 3$.

0.5.4. *A model where the new item is not explained by the same latent trait than the kernel*

The likelihood of the responses to the J_k items of the kernel of the nth individual is similar to this one presented in the equation 9 and the corresponding log-likelihood is noted $l_{kernel}^{(k)}$ at the kth step of the algorithm.

The likelihood of the response to the new item of the nth individual is estimated by :

$$P(X_{n0} = x_{n0}/\delta_0) = \frac{\exp\left[x_{n0}\left(-\delta_0\right)\right]}{1 + \exp\left(-\delta_0\right)} \qquad [11]$$

The estimation of the δ_0 parameter can be obtained by maximizing :

$$l_{C0}(\delta_0/\boldsymbol{x_0}) = \log\prod_{n=1}^{N} P(X_{n0} = x_{n0}/\delta_0) \qquad [12]$$

We easily obtain $\hat{\delta}_0 = -\log\left(\frac{t_0}{N-t_0}\right)$ where $t_0 = \sum_{n=1}^{N} x_{n0}$.

The log-likelihood of the model at the kth step of the algorithm is valuated by

$$l_2^{(k)} = l_{kernel}^{(k)} + l_{C0}(\delta_0/\boldsymbol{x_0}) \qquad [13]$$

The number of parameters is in this case $K_2^{(k)} = 2J_k + 2$.

0.5.5. *Selection of the new item in the scale*

The new item 0 is selected in the scale at the end of the step if $AIC_1^{(k)} \leq AIC_2^{(k)}$.

0.6. A small set of simulations in order to compare Raschfit and Raschfit-Fast

0.6.1. *Parameters of the simulation study*

To test the Raschfit procedure, Hardouin and Mesbah [HAR 04] have realized simulations based on a design proposed in [ABS 04]. Simulated data are unidimensional or bidimensional, and 2000 individuals are used. The parameters used in the simulations are :

    – the model used to simulate the data,

    – the structure of the data,

    – the correlation between the two latent traits,

    – the number of items in each dimension,

    – the dicriminating power of the items of each dimension.

0.6.1.1. *Models*

The model used to simulate the data is a multidimensional counterpart of the Five Parameters Accelerating Model (5-PAM) which considers five parameters for each item : the difficulty (δ_j^*), the discriminating power (α_j^*), the random response to the item (γ_j^{low}), the maximal probability to respond to the item (γ_j^{up}), and an accelerating coefficient (ξ_j) which is a coefficient of dissymmetry of the IRF. The IRF of this model is :

$$P(X_{nj} = x_{nj}/\boldsymbol{\theta}_n; \delta_j^*, \boldsymbol{\alpha}_j^*, \gamma_j^{low}, \gamma_j^{up}, \xi_j)$$
$$= \gamma_j^{low} + \left(\gamma_j^{up} - \gamma_j^{low}\right) \left[\frac{\exp\left(1.7x_{nj}(\sum_{q=1}^{Q}(\alpha_{jq}^*\theta_{nq})-\delta_j^*)\right)}{1+\exp\left(1.7\left(\sum_{q=1}^{Q}(\alpha_{jq}^*\theta_{nq})-\delta_j^*\right)\right)}\right]^{\xi_j} \quad [14]$$

with $0 \leq \gamma_j^{low} < \gamma_j^{up} \leq 1$, $\alpha_{jq}^* > 0$ and $\xi_j > 0$. We name this model a M5-PAM (for Multidimensional 5-PAM). If we have $\gamma_j^{low} = 0$, $\gamma_j^{up} = 1$ and $\xi_j = 1$ $\forall j$, the model is a multidimensional couterpart of the 2 parameters logistic model (noted M2-PLM).

The parameters $(\alpha_{jq}^*, \delta_j*)$ are computed in order to obtain Item Characteristics Curves with the same maximal slope ($\frac{1.7\alpha_{jq}}{4}$) and the same localisation of this maximal slope ($\frac{\delta_j}{\alpha_{jq}}$) on the qth component of the latent trait whatever a given value for

(α_{jq}, δ_j).

In the simulations, we consider a bidimensional latent trait, so $Q = 2$. The component of the latent which influence the more the response to the item j is indexed by q_j and the other one by $\overline{q_j}$.

The responses of a given item j can be influenced only by θ_{q_j}. In this case, $\alpha_{j\overline{q_j}} = 0$. This case corresponds to a simple structure (SS) [J. 99].

But the responses of a given item j can be mainly influenced by one main component of the latent trait and weakly by the other ($0 < \alpha_{j\overline{q_j}} << \alpha_{jq}$). This case corresponds to an approximate simple structure (ASS) [J. 99]. In the simulations, we use $\alpha_{j\overline{q_j}} = 0.2$.

We consider four cases described in the table 1.

Case	$\alpha_{j\overline{q_j}}$	Structure	$(\gamma_j^{low}, \gamma_j^{up})$	ξ_j	Model
I	0	SS	(0,1)	1	MMSRM
II	0.2	ASS	(0,1)	1	M2-PLM
III	0	SS	(0.1,0.9)	2	M5-PAM
IV	0.2	ASS	(0.1,0.9)	2	M5-PAM

Tableau 1. *Values of the parameters used in the simulations*

We note that in the case I and II, we have $\alpha_{jq}^* = \alpha_{jq}$, $\forall q$ and $\delta_j^* = \delta_j$.

0.6.1.2. *Simulation of the multidimensional latent trait*

The two latent traits are simulated by a centered standardized multinormal distribution. The correlation coefficient between the two latent traits is noted ρ and can take six different values : 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0.

0.6.1.3. *The number of items in the two dimensions*

We use two different sizes for each dimension. These values correspond to a mean value of the number of items by dimension in Quality of Life questionnaires (7 items) and to a big value of this number (14). Three designs are used : 7 items in each dimension, or 7 items in one dimension and 14 in the other. The used values of the difficulty parameters in the simulations are chosen in the 2-PLM and in each dimension q as the $l/(J_q + 1)$, $l = 1, ..., J_q$ percentiles of a standardized centered gaussian distribution.

0.6.1.4. *The discriminating powers of the items*

The same value α_q is used for the discriminating power of all the items mainly relied the same dimension q. Three different values are used in the simulations for the parameters α_q, $q = 1, ..., Q$: a low value (0.4), a medium value (0.7) and a high value (1.7). The six designs crossing two by two this three values are used.

0.6.1.5. *Description of the four main cases*

Data simulated in the case I are equivalent to data simulated with a MMSRM. In this case, we can write the IRF of the jth item :

$$
\begin{aligned}
P(X_{nj} = x_{nj}/\boldsymbol{\theta}_n; \delta_j, \boldsymbol{\alpha}_j) &= \frac{\exp\left(1.7 x_{nj}(\alpha_{jq_j}\theta_{nq_j} - \delta_j)\right)}{1 + \exp\left(1.7\left(\alpha_{jq_j}\theta_{nq_j} - \delta_j\right)\right)} \\
&= \frac{\exp\left(x_{nj}(\tilde{\theta}_{nq_j} - \tilde{\delta}_j)\right)}{1 + \exp\left(\tilde{\theta}_{nq_j} - \tilde{\delta}_j\right)}
\end{aligned}
\tag{15}
$$

with $\tilde{\theta}_{nq_j} = 1.7\alpha_{jq_j}\theta_{nq_j}$ and $\tilde{\delta}_j = 1.7\delta_j$. This expression is equivalent to this one present in the equation (3). $\tilde{\boldsymbol{\theta}}_n = (\tilde{\theta}_{n1}, \tilde{\theta}_{n2})'$ is in this case distributed as a centered multinormal distribution with a covariance matrix

$$
\tilde{\Sigma} = \begin{pmatrix} (1.7\alpha_1)^2 & 1.7^2\alpha_1\alpha_2\rho \\ 1.7^2\alpha_1\alpha_2\rho & (1.7\alpha_2)^2 \end{pmatrix}
\tag{16}
$$

This case is interesting when we search to study the results of the procedure when the model underlying the data is a MMSRM, that is to say the model used by the procedure.

The case II allows knowning the behavior of the procedures when the IRF of the items have the same general form than in the MMSRM but when the structure is less particular (the SS is a very rare structure in practice, so the ASS is a likelier structure of real data). These cases allow seeing if the introduction of a minor latent trait strongly affects in practice the notion of sufficiency of the score on the (main) latent trait.

The cases III and IV allow studying the results of the procedure when the IRF are different of the supposed IRF. The results are more difficult to analyse because the underlying notion of sufficiency of the score on the latent trait is not verified in this model. These simulations considers the cases where the unidimensionality is the main link between the items. latent trait.

0.6.1.6. *The number of simulations*

By crossing these five factors, we obtain 360 designs. Each of them is simulated one time.

0.6.2. *Results and time of computing*

0.6.2.1. *Tested procedures*

The simulated data are been treated by four procedures :

– the Raschfit procedure,

– the Raschfit-Fast procedure,

– a Mokken scale procedure (MSP) [HEM 95] which builds scales of items which verify the properties of the IRT unless searching the fit to a parametric IRM,

– a Hierarchical Cluster Analysis on conditional measures of proximity (HCACC-PROX), wich clusters together the items having the greater proximity (based on the conditional covariance between the items).

0.6.2.2. *MSP*

MSP is a procedure described in [HEM 95]. This procedure search to build scales which verifies a Mokken scale, that is to say a scale verifying the fundamental assumptions of the IRT (unidimensionality, local independency, and monotonicity).

The Mokken scale are non parametric models, and necessitate to fix a threshold (c) as the minimum acceptable value for the used indices (The Loevinger H indices [LOE 48]). The authors of this procedure suggest to choose $c \geq 0.3$, this minimum value is used in the simulations.

0.6.2.3. *HCACCPROX*

HCACCPROX is defined in [ROU 98]. This method is based on the same methods than the classical HCA : a proximity matrix is defined and a each step, the two closer elements among all these ones defined at the preeceding step are clustered together, until obtaining only one cluster. The authors defined, in the field of the IRT, three specific proximity matrix based a weighted sum of the covariance, correlation or odds-ration computed for each value of the score. They shows with simulations that the use of this method gives better results than classical measures of proximity with IRT items. The DETECT indice is used to chose the number of clusters of items (The partition which presents the minimal value for this indice is chosen). In the simulations, the distance based on the conditional covariances between the items and the WPGMA method of aggregation are used.

Procedure	Number of items	Time of computing in seconds			
		Average	Standard error	Minimum	Maximum
Raschfit	7 ;7	18190	3384	13740	39540
	7 ;14	38829	9291	23460	64800
Raschfit-Fast	7 ;7	320	110	120	720
	14 ;7	1216	530	240	2640
MSP	7 ;7	20	5	5	187
	14 ;7	89	18	8	382
HCA/CCPROX	7 ;7	89	5	79	123
	14 ;7	353	81	287	905

Tableau 2. *Average, minimum and maximum time of computing for one simulation for each procedure*

0.6.2.4. *Clustering of the results*

Let a major error of classement defined as two items which have been simulated from two disctinct dimensions and which are classified together. When we simulate two dimensions with a perfect correlation ($\rho = 1$: unidimensional case), a major error of classement is a classement which allows findind the original dimension of each item.

Each result is affected to a class among these five ones :

– Class 1 : The true classification of the items is found,

– Class 2 : Less than 2 items (for dimensions with $J_q = 7$) or 3 items (for dimensions with $J_q = 14$) are not classified in the two main dimensions,

– Class 3 : The true classification of the items is not found but there is no major error of classement,

– Class 4 : There is one or several major error(s) of classement,

– Class 5 : Unspecified results : at least 2 items (for dimensions with $J_q = 7$) or 3 items (for dimensions with $J_q = 14$) are unselected by the procedure [Only for MSP].

0.6.2.5. *Time of computing*

The average, the minimum and the maximum times of computing for each procedure are presented in the table 2. The values depends of the number of used items. The used computer is cadenced at 950MHz with 512Mo of RAM.

The table 2 shows that Raschfit, in its original version, is a very long process and is unadapted in practice (in average, 5 hours to run the procedure with seven items in each dimension and 11 hours to run it with seven and fourteen items in the two dimensions). Compared to Raschfit, Raschfit- Fast reduces the time of computing by

a factor 60 for 7 items in each dimension, and by a factor 30 for 7 and 14 items in the two dimensions.

0.6.2.6. *Results of the data simulated by a MMSRM (case I)*

The results are influenced by two main factors : the correlation coefficient between the two components of the latent traits and the fact that the discriminating powers of the items in the two dimensions are equal ($\alpha_1 = \alpha_2$) or not.

Concerning the correlation coefficient between the two component of the latent traits, we consider 3 main cases :
– the correlation is low ($\rho \leq 0.4$),
– the correlation is high ($0.6 \leq \rho \leq 0.8$),
– the two simulated latent traits are counfound ($\rho = 1$).

The table 3 presents the results of the four tested procedures in all these cases.

Procedure	Results	Equal discriminating power among the two dimensions			Different discriminating power among the two dimensions		
		$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Number of simulations		18	12	6	27	18	9
Raschfit	good	14(78%)	7(58%)	4(67%)	24(89%)	13(72%)	0
	bad	3(17%)	5(42%)	0	2(7%)	4(22%)	7(78%)
Raschfit	good	15(83%)	0	6(100%)	26(96%)	15(83%)	1(11%)
Fast	bad	3(17%)	12(100%)	0	1(4%)	3(17%)	3(33%)
HCA/CCPROX	good	15(83%)	8(67%)	6(100%)	26(96%)	10(56%)	0
	bad	1(6%)	3(25%)	0	1(4%)	2(11%)	1(11%)
MSP	good	12(67%)	3(25%)	4(67%)	8(30%)	0	4(44%)
($c = 0.3$)	bad	0	6(50%)	0	0	9(50%)	0
	unspecified	6(33%)	2(17%)	2(33%)	18(67%)	9(50%)	5(56%)

Tableau 3. *Results with data simulated by a MMSRM*

When the discriminating powers used in the simulations are the same in the two sets of items, results are good when the correlation coefficient is low ($\rho \leq 0.4$), and continue to be correct when it is medium ($0.6 \leq \rho \leq 0.8$) for Raschfit and HCA/CCPROX.

When the latent trait underlying the two sets of the items is the same ($\rho = 1$), the results are good for Raschfit and Raschfit-Fast if the two sets of items have the same

Procedure	Results	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Number of simulations		45	30	15
Raschfit	good	27(60%)	12(40%)	4(27%)
	bad	8(18%)	18(60%)	2(13%)
Raschfit-Fast	good	24(53%)	8(27%)	8(53%)
	bad	21(47%)	22(73%)	1(7%)
HCA/CCPROX	good	29(64%)	9(30%)	0
	bad	9(20%)	10(33%)	0
MSP	good	6(13%)	2(7%)	2(13%)
($c = 0.3$)	bad	0	2(7%)	0
	unspecified	39(87%)	26(87%)	8(53%)

Tableau 4. *Results with data simulated by a Multidimensional 2PLM*

discriminating powers (and so, if the global set of items can be considered as a only one Rasch scale), and if the two sets of items have different discriminating powers, these two procedure tend to consider that the latent trait can be measured with two Rasch scales. HCA/CCPROX seems to be sensible to the discriminating powers of the items, but produces few errors. MSP have success rates in these cases similar to these ones obtained with a lower value of the correlation coefficient.

0.6.2.7. *Results of the data simulated by a multidimensional 2PLM (case II)*

When the data are simulated by a multidimensional 2PLM with an ASS (see table 4), Raschfit, Raschfit-Fast and HCA/CCPROX have similar rate of success when the correlation coefficient is different to 1 : a high rate (53% to 64%) when the correlation is low, and a medium rate (27% to 40%) when the correlation is high). MSP have poor rate of success and an important rate of unspecified results (87%).

When the two components of the latent traits are counfounded, Raschfit-Fast is the procedure which produces the best rate of success (53%), but HCA/CCPROX has the advantage to not produce bad results.

0.6.2.8. *Results of the data simulated by a multidimensional 5PAM (Cases III and IV)*

The table 5 presents the results obtained with data simulated by a multidimensional 5PAM.

When data are simulated from a Multidimensional 5PAM, results are comparable to these ones obtained with the multidimensional 2-PLM with more important rate of bad results for Raschfit, Raschfit-Fast and HCA/CCPROX. The rate of unspecified

Procedure	Results	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Number of simulations		90	60	30
Raschfit	good	49(54%)	23(38%)	13(43%)
	bad	35(39%)	35(58%)	6(20%)
Raschfit-Fast	good	59(66%)	6(10%)	12(40%)
	bad	31(34%)	54(90%)	0
HCA/CCPROX	good	45(50%)	17(28%)	0
	bad	24(27%)	28(47%)	2(7%)
MSP	good	16(18%)	0	17(57%)
($c = 0.3$)	bad	27(30%)	45(75%)	0
	unspecified	44(49%)	14(23%)	4(13%)

Tableau 5. *Results with data simulated by a Multidimensional 5PAM*

results for MSP is lower but unspecified results are "becomed" bad results. The relatively good results obtained with Raschfit and a model very different of the MMSRM can be imputed to the fact that, in the simulations, the parameters are fixed in order to obtain Items Characteritics Curves (ICC) with the same location of the maximum slope and the same value of the maximum slop, so this model can not enought far of the MMSRM to create a large disturbance of the algorithm.

0.7. A large set of simulations in order to compare Raschfit-Fast, MSP and HCA/CCPROX

Raschtest-Fast seems to give similar results to Raschfit, but the slowness of Raschfit avoids to realize a large set of simulations with this procedure. In this part, we propose a simulations study with a large set of simulations, to compare Raschfit-Fast, MSP and HCA/CCPROX. The simulated cases concerns unidimensional scale (are the procedure able to detect a unidimensional scale ? - case A) and perturbance create by the add of a perturbant item (are the procedures able to detect a bad item ? - cases B-E).

0.7.1. *Parameters of the simulations*

We simulate one or bi-dimensional data.

We define four cases indexed from A to E. In each case, the correlation between the two components of the latent trait is fixed ($\rho = 0.0$ for B, $\rho = 0.2$ for C, $\rho = 0.4$ for D, $\rho = 0.6$ for E). The case A is the unidimensional case : none item is relied to the second component of the latent trait.

In each case, 7 items are simulated relatively to the first component of the latent trait, and one item is simulated relatively to the second component (except in the case A where there is none such item). The aim is to determine if the procedures can detect one bad item in a set of items (in the case A, the aim is to see if the procedures detects bad items in a set of unidimensional items).

We define too four scenarii, numbered I to IV, following the average of the discriminating power of the items : $\mu_{\alpha_I} = .5$, $\mu_{\alpha_{II}} = 1$, $\mu_{\alpha_{III}} = 2$, and $\mu_{\alpha_{IV}} = 3$. The discriminating power of each item is randomly defined among a normal distribution of standard error .2.

The difficulty of the items are taken as the percentiles of the standardized gaussian distribution. The latent trait is simulated by a multinormal centered distribution, with the identity matrix as covariance matrix.

800 replications are simulated of each case.

Case/ Scenario	Raschfit-Fast			MSP				HCA/CCPROX		
	Correct	Medium	Bad	Correct	Medium	Bad	Unspecified	Good	Medium	Bad
A I	781	19	-	0	0	-	800	0	800	-
II	800	0	-	26	221	-	553	0	800	-
III	800	0	-	800	0	-	0	800	0	-
IV	800	0	-	800	0	-	0	800	0	-
B I	625	0	175	0	0	0	800	0	775	25
II	800	0	0	23	73	0	704	2	789	9
III	800	0	0	800	0	0	0	799	1	0
IV	800	0	0	800	0	0	0	800	0	0
C I	562	0	238	0	0	0	800	0	767	33
II	799	0	1	29	97	0	674	1	786	13
III	800	0	0	800	0	0	0	800	0	0
IV	800	0	0	800	0	0	0	800	0	0
D I	469	0	331	0	0	0	800	0	755	45
II	712	0	88	22	107	0	671	0	770	30
III	800	0	0	800	0	0	0	795	5	0
IV	800	0	0	780	0	20	0	800	0	0
E I	259	0	541	0	0	0	800	0	733	67
II	186	0	614	30	63	0	0	0	753	47
III	187	0	613	495	0	305	0	767	32	1
IV	702	0	98	1	0	799	0	800	0	0

Tableau 6. *Results of the simulations concerning the comparison of Raschfit-Fast, MSP and HCA/CCPROX*

0.7.2. *Discussion*

Raschfit-Fast allows obtaining good results as soon as the correlation between the two component of the latent trait is low, and/or the discriminating power of the items is high. When the set of items is unidimensional, Raschfit-Fast allows detecting that there is no perturbant item. Indeed, the quality of the results strongly decreases with the increasing of the correlation between the two components of the latent trait.

MSP produces full of unspecified results, notably when the discriminating powers of the items is low (inferior or equal to 1). In the others cases, the results are correct, except if the discriminating powers of the items are too high (superior to 1) and if the correlation between the two components of the latent trait are too correlated (coefficient superior to 0.4).

HCA/CCPROX produces good results when the discriminating powers of the items are high (superior to 1) and else, medium results.

Raschfit-Fast, in these cases where the real model is close of the MMSRM, the more powerfull procedures among these three ones, notably when the conditions are less favourable : high correlation between the components of the latent traits and/or low discriminating power of the items.

0.8. The Stata module "Raschfit"

We propose a Stata module named -raschfit- to realize the Raschfit procedure. By default, this module run Raschfit-Fast. The Stata module -raschfit- can be downloaded from the FreeIRT Project at http ://freeirt.free.fr.

The syntax of -raschfit- is simple. The user indicates the names of the used items. By default, MSP is run under the items to order them among a negative order, and the two first items selected by MSP are considered as the initial kernel of the scale. The others items are ordered with MSP from the last item selected by MSP to the first one (except the kernel).

It is possible to modify the method to order the items with the "itemsorder" option which can be "msp" (by default), "mspinv" (the kernel is selected from the same way than by default, but the others items are taken in the inverse order) and "order" which orders the items in the same order than this one defined by the user.

The number of sub-scales to build is defined by the "nbscales" option (1 by default). The size of the kernel of the first sub-scale can be defined by the "kernel" option (2 by default). Last, it is possible to run the original version of Raschfit in adding the "nofast" option.

With the syntax of the Stata manual, the syntax of the -raschfit- module is :

. **raschfit** *varlist* [, <u>**kernel**</u>(*#*) <u>**nbscales**</u>(*#*) <u>**items**</u>**order**(*keyword*) **nofast**]

0.9. Conclusion

A new procedure named Raschfit had been proposed in a preceeding paper It allows selecting the items which fit a Rasch model. This procedure is based on the fit of the data to a multidimensional Item Response Model, instead of on the correlations between items (as in the factor analysis) or on the properties of the items (as in the Mokken Scales Procedure - MSP).

Raschfit is more performant than the existing procedures which are based on unidimensionality of the items, especially when the multidimensional model underlying the data is close of this one used in the procedure. The main drawback of Raschfit is the time of computing (until several hours, even the number of items is small). A new version of this procedure, named Raschfit-Fast, is proposed in this paper. Raschfit-Fast allows estimating more fastly the likelihood of the models, and considerably reduces the time of computing, even if already existing procedures (MSP for example) still are faster. This adaptation of Raschfit give similar results compared to the former version, if the latent traits underlying to each set of items have a low correlation.

This is encouraging results for this type of procedure, based on the fit to IRM, even new improvements are necessary to reduce the rates of bad results.

0.10. Bibliographie

[ABS 04] VAN ABSWOUDE A. A. H., VAN DEN ARK L. A., SITJSMA K., « A comparative study on test dimensionality assessment procedures under non parametric IRT models », *Applied Psychological Measurement*, vol. 28, p. 3–24, 2004.

[ADA 97] ADAMS R. J., WILSON M. R., WANG W., « The multidimensional random coefficient multinomial logit model », *Applied Psychological Measurement*, vol. 21, p. 1-23, 1997.

[AND 70] ANDERSEN E. B., « Asymptotic properties of conditional maximum likelihood estimators », *Journal of the Royal Statistical Society, Series B*, vol. 32, p. 283–301, 1970.

[AND 77] ANDERSEN E. B., « Sufficient statistics and latent trait models », *Psychometrika*, vol. 42, n°1, p. 69–81, 1977.

[FIS 95] FISHER G. H., MOLENAAR I. W., *Rasch models. Foundations, recent developments, and applications*, New york : Springer édition, 1995.

[HAR 04] HARDOUIN J.-B., MESBAH M., « Clustering binary variables in subscales using an extended Rasch model and Akaike Information Criterion », *Communications in Statistics - Theory and Methods*, vol. 33, n°6, p. 1277–1294, 2004.

[HEM 95] HEMKER B. T., SITJSMA K., MOLENAAR I. W., « Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model », *Applied Psychological Measurement*, vol. 19, p. 337–352, 1995.

[HOI 97] HOIJTINK H., MOLENAAR I. W., « A multidimensional item response model : constrained latent class analysis using the Gibbs sampler and posterior predictive checks », *Psychometrika*, vol. 62, n°2, p. 171–189, 1997.

[J. 99] J. Z., STOUT W. F., « The theorical DETECT index of dimensionality and its application to approximate simple structure », *Psychometrika*, vol. 64, n°2, p. 213-249, 1999.

[LIN 97] VAN DEN LINDEN W. J., HAMBLETON R. K., *Handbook of Modern Item response Theory*, New york : Springer-verlag édition, 1997.

[LOE 48] LOEVINGER J., « The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis », *Psychological Bulletin*, vol. 45, p. 507–530, 1948.

[RAS 60] RASCH G., *Probabilistic models for some intelligence and attainment tests*, Copenhagen : Nielsen & lydiche édition, 1960.

[ROU 98] ROUSSOS L. A., STOUT . F., MARDEN J. I., « Using new proximity measures with hierarchical cluster analysis to detect unidimensionality », *Journal of Educational Measurement*, vol. 35, n°1, p. 1-30, 1998.