SPARSE EXPLORATORY MULTIDIMENSIONAL IRT MODELS*

L. Fontanella [1], S. Fontanella[2], P. Valentini [1] , N. Trendafilov[2]

[1] Department of Economics, University of Chieti-Pescara, Italy (e-mail: lfontan@unich.it, p.valent@unich.it)

[2] Department of Mathematics and Statistics, The Open University, UK , (e-mail: Sara.Fontanella@open.ac.uk, Nickolay.Trendafilov@open.ac.uk)

KEYWORDS: "multidimensional IRT", "exploratory item factor analysis", "spike and slab priors".

1 Introduction

In the framework of Item Response Theory (IRT), great attention has been recently paid to the multidimensional formulation of IRT models, which accounts for distinct underlying latent traits, involved in producing the manifest responses to the selected items. IRT can be applied in a way that is analogous to exploratory and confirmatory factor analysis for continuous variables (Reckase, 1997). In an exploratory perspective, few works have been proposed in literature and this is probably due to the fact that the rotational invariance of the final solution has to be considered. As seen in classical exploratory factor analysis framework (West, 2003), the introduction of sparsity in the model allows to eliminate the rotation indeterminacy. In this paper, following a Bayesian approach to MIRT models, the sparsity is induced by introducing prior probability distributions that favour shrinkage for the coefficients of the discrimination parameter matrix. Specifically, we address the sparse MIRT problem by introducing the sparsity-inducing prior suggested in the Stochastic Search Variable Selection (SSVS) approach for regression models (George and McCulloch, 1993).

2 Bayesian framework for the sparse multidimensional normal ogive model

We assume that there are M correlated constructs measured by K observed categorical items. The multidimensional normal ogive model (Béguin and Glas,

*This work is supported by a grant RPG-2013-211 from The Leverhulme Trust, UK.

2001) is given by

$$P(X_{ik} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k) = \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c-1}\right) - \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c}\right) \tag{1}$$

where X_{ik} is the observed response of person i ($i = 1,\ldots,N$) to item k ($k = 1,\ldots,K$); c denotes the category of the ordered response scale ($c = 1,...,C$) and Φ is the cumulative distribution function for the standard normal distribution. The probability of responding a certain category c, to a given item k, depends on the M-dimensional vector $\boldsymbol{\theta}_i$ of the unobserved latent scores for subject i, on the M-dimensional vector $\boldsymbol{\alpha}_k$ of item discrimination parameters and on the $(C-1)$-dimensional vector $\boldsymbol{\gamma}_k$ of category thresholds.

In an exploratory framework, the multidimensional normal ogive model requires identification restrictions given the over-parameterisation. For location and scale indeterminacy, we consider the constraint $\sum_{k=1}^{K}\gamma_{k,c} = 0$ for a given category c and assume that each latent component has unit variance. To address the rotational indeterminacy, we require the discrimination parameter matrix to be sparse lower triangular and of full rank.
With regards to the prior specification, on the person side of the model, we assume that all person parameters $\boldsymbol{\theta}_i$ are independent and identically distributed samples from a multivariate normal distribution, that is $\boldsymbol{\theta}_i \sim \mathcal{N}_M(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ (Béguin and Glas, 2001). The prior for the mean $\boldsymbol{\mu}_\theta$ is normal with mean $\boldsymbol{\mu}_0 = \mathbf{0}$ and covariance matrix $\sigma_\mu^2\mathbf{I}$ where $\sigma_\mu^2 = 100$. The prior for the inverse covariance matrix $\boldsymbol{\Sigma}_\theta$ is a Wishart with scale matrix $0.1\mathbf{I}$ and degree of freedom $M+1$.
On the item side, a uniform prior is assigned to the ordered thresholds $\gamma_{k,c} \sim uniform$, $c = 1,\ldots,C-1$, $\gamma_{k,1} \leq \cdots \leq \gamma_{k,C-1}$, $\forall k$. For the discrimination parameters, we assume that each free $\alpha_{k,m}$ arises from one of two normal mixture components, depending on a latent variable $\zeta_{k,m}$

$$\alpha_{k,m}|\zeta_{k,m} \sim (1-\zeta_{k,m})\mathcal{N}(0,\tau^2) + \zeta_{k,m}\mathcal{N}(0,c^2\tau^2) \tag{2}$$

where τ is positive but small, such that $\alpha_{k,m}$ is close to zero when $\zeta_{k,m} = 0$; c is large enough to allow reasonable deviations from zero when $\zeta_{k,m} = 1$. In addition, the prior probability that factor m has a nonzero effect is $P(\zeta_{k,m} = 1) = 1 - P(\zeta_{k,m} = 0) = p_k$. To obtain the normal mixture prior for $\boldsymbol{\alpha}_k$, George and McCulloch (1993) define a multivariate normal prior

$$\boldsymbol{\alpha}_k|\boldsymbol{\zeta}_k \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{D}_{\zeta_k}\boldsymbol{R}\boldsymbol{D}_{\zeta_k})$$

where $\boldsymbol{\alpha}_k = (\alpha_{k,1},\ldots,\alpha_{k,M})'$, $\boldsymbol{\zeta}_k = (\zeta_{k,1},\ldots,\zeta_{k,M})'$ and $\boldsymbol{D}_{\zeta_k} = diag(d_{k,1}\tau,\ldots,d_{k,M}\tau)$ with $d_{k,m} = 1$ if $\zeta_{k,m} = 0$ and $d_{k,m} = c$ if $\zeta_{k,m} = 1$. As prior correlation we assume the identity matrix $\mathbf{R} = \mathbf{I}_M$.

In order to draw samples for the posterior distribution of the parameters, it is convenient to use data augmentation technique. For each observed polytomous item, we assume that a continuous variable Z_k underlies the observed ordinal measure X_k and that there is a linear relationships between item and person parameters and the underlying variable such that $Z_{i,k} = \boldsymbol{\alpha}_k' \boldsymbol{\theta}_i + e_{i,k}$, with $e_{i,k} \sim \mathcal{N}(0,1)$. The relation between the observed item k and the underlying variable is given by the threshold model

$$X_{i,k} = c \text{ iff } \gamma_{k,c} < Z_{i,k} < \gamma_{k,c+1}. \quad (3)$$

The full conditional of most parameters can be specified in closed form which allows for a Gibbs sampler although a Metropolis-Hastings step is required to sample the threshold parameters.

3 Simulation study and Application

In this section, we evaluate the performance of the proposed method both on simulated and on a real world dataset.
For the simulation study, we consider different level of sparsity and both uncorrelated and correlated latent variables. We set the hyperparameters in Equation (2) as $\tau = 1$ and $c = 0.01$. In order to assess the performances of the proposed method, we compare the structure of the simulated discrimination parameter matrices with the ones retrieved by the algorithm. The results, displayed in Table 1, show that the normal mixture prior for $\boldsymbol{\alpha}_k$ is able to infer the sparse underlying structure. In particular, we notice that the algorithm correctly identifies the number of zero elements and their locations when it is assumed that each item measures only one latent construct (level of sparsity 75%). Overall, the procedure is promising, although there is the hint that the proportion of zeroes tends to be overestimated, especially for low levels of sparsity.

Original sparsity level	Uncorrelated $\boldsymbol{\theta}$		Correlated $\boldsymbol{\theta}$	
	Retrieved sparsity level	Correctly identified zero elements	Retrieved sparsity level	Correctly identified zero elements
0.75	0.75	1.00	0.75	1.00
0.69	0.67	0.98	0.73	0.96
0.58	0.62	0.96	0.71	0.87

Table 1. *Simulation results: proportions of zero entries and of correctly identified zero elements*

In the application, we consider data* analysed in Martin et al.(2003) and collected through the Humor Styles Questionnaire (HSQ) which assesses four dimensions relating to individual differences in uses of humor to enhance: the self (*Self-enhancing*); one's relationships with others (*Affiliative*); the self at the expense of others (*Aggressive*); relationships at the expense of self (*Self-defeating*). Consequently, in our analysis we consider a 4 dimensional solution with correlated factors. In figure 1 we provide a comparison of: the structure hypothesised in Martin et al.(2003) for the 32 items of the HSQ scale; the sparse structure retrieved considering the significant discrimination parameters (credible interval with probability 90%) estimated with our proposed method; the oblimin solution, with cut-off point set to .10, for the polythomic multidimensional IRT model estimated with the MIRT package (Chalmers, 2012).

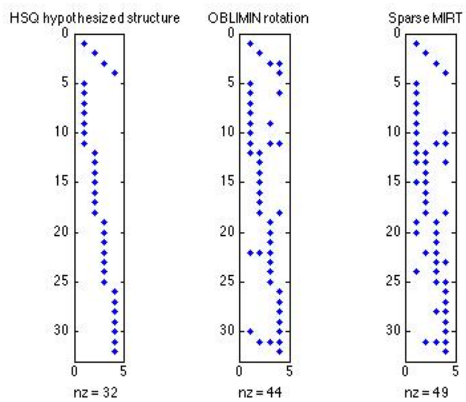

Figure 1. *Sparse structures of the discrimination parameter matrix for the HSQ scale.*

We demonstrated clearly that the sparse loadings are obtained in an optimal way and are easily interpretable. Moreover, the new sparse solution agrees with the classic ones.

References

BÉGUIN, A.A., & GLAS, C.A.W. 2001. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, **66**(4), 541–561.

*Data are available at <http://personality-testing.info/_rawdata/>

CHALMERS, R. P. 2012. mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, **48**(6), 1–29.

GEORGE, E. I., & MCCULLOCH, R. E. 1993. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, **88**(423), 881–889.

MARTIN, R. A., PUHLIK-DORIS, P., LARSEN, G., GRAY, J., & WEIR, K. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, **37**(1), 48–75.

RECKASE, M. D. 1997. The Past and Future of Multidimensional Item Response Theory. *Appl. Psych. Meas.*, **21**(1), 25–36.

WEST, M. 2003. Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. *Bayesian Statistics*, 723–732.