Quaderni di Statistica
Vol. 14, 2012

Multidimensional extensions of IRT models and their application to customer satisfaction evaluation

Federico Andreis, Pier Alda Ferrari

Department of Economics, Business and Statistics - Università degli Studi di Milano

Summary: Multidimensional IRT models (MIRTM), developed in the fields of psychometrics and ability assessment, are here considered in connection with the problem of evaluating customer satisfaction. Different models, that allow us to take into account more complex and, possibly, more realistic latent constructs than those usually assumed, are presented and discussed. Eventually, these models are applied to a real dataset, MCMC techniques for the estimation are implemented and analogies and differences with results from previous analyses on the same survey in the literature are discussed.

Keywords: BINARY RESPONSES, COMPENSATORY MODELS, MIRT.

1. Introduction

It has become increasingly clear in the last decades that, in complex settings where unobservable (latent) quantities are of interest, the usual IRT assumption of a single underlying component influencing the observable outcomes might be not realistic [5, Chap.3]. Multidimensional IRT models (MIRTM) arise from the fields of psychometrics and ability assessment (as do their ancestors, the unidimensional IRT models, UIRTM), their aim being to overcome this limitation. The rationale behind these techniques is to provide an instrument capable of describing the usually not trivial apparatus of skills that a person brings to a test, in order to obtain a diagnostic tool about several subscales simultaneously and a way to model the interaction between examinees and test items' characteristics. It is important to point out, though, that this approach is not intended to provide a measurement of the latent trait in the sense Rasch introduced [4] (i.e. objective measurement), rather than to investigate from a modeling point of view the complexity of such unobservable phenomenon, thus following the statistical logic of model tuning in order to fit data, i.e. choosing the best tool available. Such distinction has been highlighted since the beginnings: the Rasch model is a theoretical ideal, a definition of measurement, as opposed to statistical modeling, which is a toolbox we reckon being useful to try and disentangle not trivial intertwinings within the underlying (often very articulated) reality. Along these lines, our interest is into evaluating the possibility of employing MIRTM in a field they haven't previously been applied to, specifically customer satisfaction, assessing interpretability of models' parameters and comparing this approach to current methodology.

2. Multidimensional IRT models

Different approaches to multi-dimensionality have been carried out in the literature, leading to the definition of two main classes of models, the 'compensatory' and 'non (or partially) compensatory' ones [5, Chap. 4] and to the introduction of the concepts of 'between-items' and 'within-items' dimensionality [1]. Compensatory models allow, through proper parameterization, latent traits' effects to compensate for each other, e.g. a high level of ability on one dimension can make up for a low level on another - think of an additive functional form for the parameters; non-compensatory (even though it would be more correct to say 'partially' compensatory) models do not admit such strong compensation - think of a multiplicative form. Between-items and within-items dimensionality embody assumptions regarding how latent traits are represented by items in a questionnaire, i.e. if each item is related to one, and only one, of the latent traits (between-items), or is to be linked to more than one at the same time (within-items).

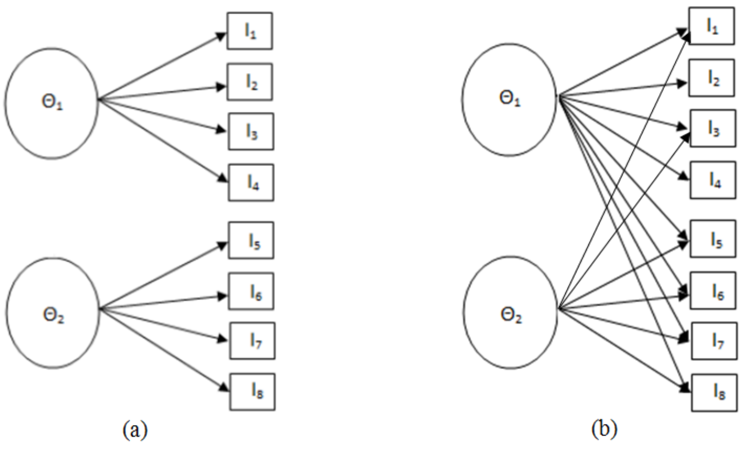

Figure 1. Between-items (a) and within-items (b) dimensionality with two latent traits.

Multidimensional extensions of classic IRT models have been introduced; they permit to work with dichotomous or polytomous test responses, but also to include covariates [5, Chap. 4].
Since MIRTM require a greater number of parameters to be estimated than unidimensional IRT models, estimation issues arise and are addressed. In the literature, MCMC methods are advocated as a helpful tool to obtain accurate results, but implementation of the algorithms and assessment of convergence to the desired posterior distribution require careful assessment [6].

3. Customer satisfaction assessment with MIRTM

Recently, unidimensional IRT models have been applied to the field of customer satisfaction (CS) evaluation, via a convenient re-interpretation of the role of their parameters [2]. Since it is legitimate to think about the existence of more than one latent factor in CS-related investigations too, the aim of this work is to evaluate the possibility of identifying and applying suitable MIRTM to this context.
In order to better appreciate the meaning and additional contribution of the extension to more than one latent trait, we first review the basic one-dimensional dichotomous model, whose expression is given by:

$$P(X_{ij} = x|\theta_i, \beta_j) = \frac{e^{x(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \tag{1}$$

where $x = 0, 1$ if the customer is, respectively, unsatisied or satisfied, θ_i is *satisfaction* for the $i-$th respondent and $(-\beta_j)$ the quality of the $j-$th item. This model is algebraically equivalent to that introduced by Rasch [4] in order to evaluate ability tests, θ_i being the person ability and β_j the item difficulty.

The multidimensional extension we consider is called *multidimensional 2 parameter logistic* (M2PL) model, and has the following expression:

$$P(X_{ij} = x|\boldsymbol{\theta}_i, \mathbf{a}_j, d_j) = \frac{e^{x(\mathbf{a}'_j\boldsymbol{\theta}_i + d_j)}}{1 + e^{\mathbf{a}'_j\boldsymbol{\theta}_i + d_j}} \tag{2}$$

where both $\mathbf{a}_j$ and $\boldsymbol{\theta}_i$ are $m-$dimensional vectors, so that the sum in the exponent of e can be rewritten as $\mathbf{a}'_j\boldsymbol{\theta}_i + d_j = \sum_{t=1}^{m} a_{jt}\theta_{it} + d_j$. This model assumes that m latent traits characterize the satisfaction, with single scores θ_{it} whose relevance is weighted by the a_{jt} parameters, known as items' discrimination related to the $t-$th dimension, and here intended to describe the relevance of the $j-$th item to the $t-$th trait. The M2PL belongs to the class of compensatory models, since it adopts a parameterization that is a linear combination of the satisfaction parameters (one for each presumed dimension of the latent construct) for each individual (customer/user), i.e. a high satisfaction value θ_{it} on one latent dimension can compensate for a low θ_{is} on another.

The a_{jt} parameters may be estimated from the data or be fixed according to particular assumptions. If all are assumed to be equal to a fixed value $a_{jt} = a^*$, then $\mathbf{a}_j'\theta_i + d_j = a^* \sum_{t=1}^m \theta_{it} + d_j$, and we assume that the discriminating power of all the items is the same, also across all dimensions; this means that different items have different quality, but the same relevance to each trait. If $a^* = 1$, we obtain the initial model, with the positions $\theta_i = \sum_{t=1}^m \theta_{it}, d_j = -\beta_j$, thus obtaining a mere reparameterization of the θ_i parameters from the unidimensional model, not taking into account dimensionality, i.e. collapsing the model back to one latent trait only. Another possibility is to fix the a_{jt} either to be zero or different from zero: this way, it becomes possible to embody assumptions about between- or within-items dimensionality. If the researcher believes a certain item (or group of items) to be dependent on one latent trait only, for example the $s-$th, this can be implemented in the model by letting $a_{jt} = 0, t \neq s$, and a_{is} either to be estimated or fixed to some non-zero value; this is the structure in Fig.1 - (a).

4. Application

Eventually, the models we discuss are applied to real data, presented in [3, Chap. 2]; the dataset consists of responses (on a 1-5 Likert scale, that we dichotomize for our purposes) to a questionnaire concerning satisfaction for a large firm's services. Analogies and differences with previous IRT-oriented analyses on the same survey [3, Chap. 14] are investigated. The results are then compared and discussed.

References

[1] Adams, R.J., Wilson, M. and WANG, W. (1997), The Multidimensional Random Coefficients Multinomial Logit Model, *Applied Psychological Measurement,* **21-1**, *1-23* .

[2] De Battisti, F., Nicolini, G. and Salini, S. (2005), The Rasch Model to Measure Service Quality, *The ICFAI Journal of Services Marketing,* **III-3**, *58-80.*

[3] Kennet, R.S. and Salini, S. (2012), *Modern Analysis of Customer Satisfaction Surveys: with applications using R*, New York: Wiley.

[4] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research).

[5] Reckase, M.D. (2009), *Multidimensional Item Response Theory*, Springer, Statistics for Social and Behavioral Sciences.

[6] Sinharay, S. (2004), Experiences With Markov Chain Monte Carlo Convergence Assessment in Two Psychometric Examples, *Journal of Educational and Behavioral Statistics,* **29**, *461-488* .