UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Multidimensional IRT Models for Hierarchical Latent Structures

Juliane Venturelli Silva Lima

Belo Horizonte
2019

Juliane Venturelli Silva Lima

Multidimensional IRT Models for Hierarchical Latent Structures

Tese apresentada ao Programa de Pós-graduação em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutora em Estatística.

Orientador: Prof. Doutor Flávio Bambirra Gonçalves
Co-orientador: Prof. Doutor Dalton Francisco de Andrade

Belo Horizonte
2019

À Vó Irene e ao Tio Ju, por me balizarem sobre sentido de tanto trabalho.

Ao futuro Professor Doutor Rodolfo Victor.

Minhas avós estudaram até a quarta série, mas fizerem de tudo para que seus filhos estudassem. Graças ao esforço delas, acordando às quatro da manhã, andando quilômetros a pé para economizar o dinheiro da passagem, se privando de todo e qualquer luxo, meu pai e minha mãe conseguiram estudar e fazer faculdade. Fizeram-no trabalhando às vezes 12 horas por dia ou mais, em meio a plantões e plantações, para que eu e minhas irmãs chagássemos mais longe do que eles. Minha gratidão às minhas raízes por deixarem meu caminho um tanto quanto fácil, sem percalços ou maiores dificuldades.

Minha sincera gratidão ao professor Flávio Bambirra Gonçalves por todo suporte, dedicação e paciência: um Professor nato. Agradeço ao professor Dalton de Andrade, por aceitar fazer parte deste trabalho e por suas contribuições. Agradeço aos professores do Departamento de Estatística, que durante a graduação, mestrado e doutorado, contribuíram para minha formação acadêmica. Agradeço à professora Glaura Franco pelos anos de iniciação científica fundamentais para meu caminho acadêmico. Agradeço à professora Rosangela Loschi, apaixonada por dar aula, pelo exemplo e contribuições imensuráveis em minha formação. Meu agradecimento, em especial, ao amigo R.Q. pelo incentivo, cumplicidade e constante apoio. Agradeço a todos os funcionários do Departamento de Estatística, sempre dispostos a me atenderem.

Agradeço aos amigos e colegas de Departamento por tornarem essa jornada menos solitária. Minha querida Larissa Sayuri, obrigada!! Lívia, por estar ao meu lado no desespero e na alegria durante esses nove anos! Minha gratidão à Marina Muniz pelo compartilhamento de plantas e cafés, e à Vivian, minha Paixão, pela amizade cúmplice. Ao meu irmão e amigo Guilherme Borges e família. À Bruna Thaisa, pelo incentivo e companheirismo em grande parte dessa jornada. Agradeço às minhas alunas e alunos, talvez a maior razão para eu chegar aqui. Minha paixão, sem dúvida, é estar dentro de uma sala de aula.

Obrigada à minha família, tias, tios e primas e primos, pelo suporte inquestionável, pelo apoio e incentivo diário. Minha gratidão à Carol e Bruna, pelo prazer de ser irmã, Junior e James, pelos meus meninos Benjamin, Lucca, Charlie e próximos. À Aninha e ao Arthur, que muitas vezes busquei para sorrir. À Biga e Frida, pelo amor incondicional, e por saberem me escutar como ninguém.

À deusa da Aleatoriedade, que me levou a caminhos jamais programados.

Abstract

Many latent variables can be seen as having a hierarchical structure and some models have been proposed towards this. Motivated by the limitations of the existent literature approaches and on the importance of the theme, this work aims to propose an improved methodology in terms of both modeling and inference. From a modeling perspective, the proposed methodology allows for genuinely multidimensional items, which no past work have done, with a higher order structure for latent traits, in which all of them are in the same scale. We approach computational aspects, such as sampling jointly all latent traits and we carefully devise a MCMC algorithm to be efficient. Also, we prove some model properties that the past authors had done.

Contents

Chapter 1

Introduction

Latent variables, also known as constructs, latent traits or factors, are features that cannot be directly measured, but that can be indirectly measured by some manifests, such as responses to items. Intelligence, personality traits and depression are some examples of constructs. Moreover, some latent variables suggest the existence of an inherited hierarchical structure. Take, for instance, language constructs, which itself can be perceived as a higher dimension with subdimensions, such as reading and writing abilities. Going even further, the reading process could also be seen as a higher dimension formed by subdomains such as interpretation, evaluation and retrieval of information. It may be the case that the researcher is interested not only in measuring the general dimension, say language ability, but also in measuring those other subdomains. In modern theory, hierarchical structures are also used to model personality (DeYoung, 2006) and intelligence traits (Carroll, 1993). Recent intelligence theories consider a general ability in the hierarchical apex and specific cognitive abilities forming the lower level of the hierarchy (Brunner et al., 2012).

An example of hierarchical structure for constructs may be perceived for Enem, a large scale educational assessment exam in Brazil. Subjects are submitted to four multiple choice subtests, each containing 45 items, and one essay. The multiple choice subtests are Human Sciences (History, Geography, Philosophy and Sociology), Natural Sciences (Chemistry, Physics and Biology), Language (Portuguese, English or Spanish and Arts), and Mathematics. In the current approach, each subtest is individually modeled and the essay is rated in the traditional way, in which trained raters assign a score based on some previously defined criteria. The subject's final score is the average of the four subtests and the essay scores. It is plausible to consider that the specific latent traits, for each subset, form a complex structure, in which all of them are related. One way of dealing with this complex structure is to consider a hierarchical model in which the specific latent traits are related through a higher construct, say, a general academic ability. It is also reasonable to assume that some items are measured by more than one specific ability. For instance, Mathematics questions could demand high levels of interpretation.

By obtaining five abilities separately, Enem's current model does not take into account that a same subject is answering Mathematics and also Language items. It is expected, however, a significant correlation among latent traits responses of a given subject. By not considering this relationship, information is lost, thus reducing estimation accuracy in the individual's latent traits.

Some IRT models have been proposed to take into account a hierarchical structure in the construct level. De La Torre & Douglas (2004) proposed a higher-order latent trait model for cognitive diagnosis, however, the authors considered only dichotomous latent traits, which is very limited. Sheng & Wikle (2008) proposed a Bayesian framework approach to estimate both the general and specific abilities using a multiunidimensional IRT model on the item level. Sheng & Wikle (2009) proposed a model in which both general and specific abilities form an additive structure, so that each item measures both general an specific abilities directly, similar to a bifactor model approach. Following Sheng & Wikle (2009), Huang et al. (2013) added new hierarchical levels for the ability and extended its use to both dichotomous and polytomous items.

All the existing models consider a multiunidimensional structure at the item level, that is, items measure only one specific latent trait on the first level. The purpose of this work is to develop a class of multidimensional IRT models for complex hierarchical latent traits structures that accommodate both dichotomous and polytomous items. Therefore, the present work extends multidimensional models allowing a flexible hierarchical structure for the latent variables. This approach allows one to model more complex situations, commonly expected for cognitive latent variables. Some restrictions are assumed in order to set the scale of all the latent traits in an interpretable way without loosing model flexibility. Also, we devise an efficient MCMC algorithm to perform Bayesian inference.

This work is organized as follows. Chapter 2, presents a literature review covering the main current IRT models for dealing with higher order latent structures. In Chapter 3, the main contributions of this work are presented. In Chapter 4 we present some simulation results and a study case.

Chapter 2

IRT Literature Review: Models for Hierarchical Latent Structures

Some IRT hierarchical models have been developed in order to estimate ability structures that go beyond the unidimensional scope. Examples include the bifactor model and the higher order model. We present a brief description and comparison of those models and discuss their weaknesses that motivate our work.

2.0.1 Bifactor model

In the bifactor IRT model (Gibbons & Hedeker, 1992), each item is assumed to measure a general (common) and specific latent traits.

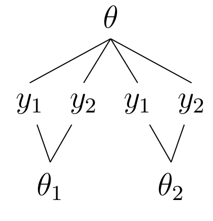


Figure 2.1: Bifactor model with independent dimensions.

Figure 2.1 illustrates the bifactor model, where θ is the general latent trait of main interest in the study and θ_q, $q = 1, 2$, are specific latent variables that are incorporated into the model to accommodate structures of dependency among items that belong to the same cluster. These specific latent traits are regarded as nuisance only.

To specify the bifactor model, let Y_{ijq} be the response to item $i = 1, ..., I_q$, belonging to cluster $q, q = 1, ..., Q$, answered by subject $j, j = 1, \ldots, J$. The probability of success is given by

$$P(Y_{ijq} = 1 | a_i, a_{iq}, \beta_{iq}, \theta_j, \theta_{qj}) = g(a_i \theta_j + a_{iq} \theta_{qj} - b_{iq}), \tag{2.1}$$

in which g(.) is either the probit or logit link, α_i is the discrimination parameter of item i associated to the general ability, α_{iq} is the discrimination parameter associated to the specific trait θ_{qj}, and β_{iq} is the location parameter of item i.

2.0.2 Second order model

As in the bifactor model, in the second order IRT model both general and specific latent traits are taken into account. However, in the second order model items only directly depend on their specific latent traits, which in turn depend on the general latent trait. This model assumes that the specific latent traits are conditionally independent from each other, so that all intercorrelation among lower order constructs are explained by the general trait. This implies that the general construct influences all first-order constructs, so that higher scores on the general construct are associated with higher scores on the first order factors (Brunner et al., 2012).

Sheng & Wikle (2008) proposed a second order model structure for latent trait using an IRT model. The authors proposed a Bayesian framework approach to estimate both the general and specific latent traits. In order to formally specify the model, consider a test consisting of I items and K clusters (subtests), each cluster containing multiple choice items. Let Y_{ijq} denote the binary response from the jth subject to the ith item of the qth cluster, in which $j = 1, 2, \ldots, J$ and $i = 1, 2, \ldots, I_q$. The probability of success is given by

$$P(Y_{ijq} = 1 | \theta_{qj}^{(1)}, a_{iq}, b_{iq}) = \Phi(a_{iq} \theta_{qj}^{(1)} - b_{iq}), \tag{2.2}$$

in which a_{iq} and b_{iq} are discrimination and difficulty parameters, respectively, and $\theta_{qj}^{(1)}$ is the specific latent trait related to the q^{th} subtest. The authors proposed two hierarchical IRT models for the latent trait structure, as follows.

Each specific latent trait is written as a linear function of the general latent trait such that

$$\theta_{qj}^{(1)} = \lambda_q \theta_j^{(2)} + \epsilon_{qj}, \tag{2.3}$$

in which $\epsilon_{qj} \overset{iid}{\sim} N(0,1)$ is the error term, $\theta_j^{(2)} \overset{iid}{\sim} N(0,1)$ is the jth subject's general latent trait parameter, and λ_q is a measure of association between the general latent trait and the qth specific latent trait parameters. These specifications imply that, marginally, $\theta_{qj}^{(1)} \sim N(0, 1+\lambda_q^2)$. Also, from the given specifications, $Cov(\theta_{qj}^{(1)}, \theta_j^{(2)}) = \lambda_q$ and $Corr(\theta_{qj}^{(1)}, \theta_j^{(2)}) = \frac{\lambda_k}{\sqrt{1+\lambda_q^2}}$.

Combining Equations (2.2) and (2.3), the probability of success becomes

$$P(Y_{ijq} = 1 | a_{iq}, b_{iq}, \lambda_q, \epsilon_{qj}, \theta_{qj}^{(2)}) = \Phi(a_{iq}\lambda_q\theta_j^{(2)} + a_{iq}\epsilon_{qj} - b_{iq}). \tag{2.4}$$

As noticed by the authors, the systematic component is similar to the one observed in the bifactor model, which has the form $(a_i\theta_j + a_{iq}\theta_{qj} - b_{iq})$. However, the bifactor model is a special case of model (2.4) when $\lambda_q = 1$ (Sheng & Wikle, 2008). In terms of specific and general latent trait, the bifactor model states that the specific latent trait can differ from the general latent trait only by an error since when $\lambda_q = 1$ Equation (2.3) becomes $\theta_{qj}^{(1)} = \theta_j^{(2)} + \epsilon_{qj}$.

Another model proposed by Sheng & Wikle (2008) assumes that the general latent trait is a linear combination of all the specific latent traits such that

$$\theta_j^{(2)} = \lambda_1\theta_{1j}^{(1)} + \cdots + \lambda_k\theta_{kj}^{(1)} + \epsilon_j, \tag{2.5}$$

where $\epsilon_j \sim N(0,1)$ and λ_q, $q = 1, \ldots, Q$, are the regression coefficient that relates the specific and general latent traits. According to the authors, both models differ in the prior belief about the latent structure. Model (2.3) assumes the general latent trait is a factor for each specific latent trait, so that $\theta_{qj}^{(1)} | \theta_j^{(2)}, \lambda_q \sim N(\lambda_q\theta_j^{(2)}, 1)$. On the other hand, Model (2.5) assumes the general latent trait is a linear function of all specific latent traits and $\theta_j^{(2)} | \boldsymbol{\theta}_j, \boldsymbol{\lambda} \sim N(\sum_q \lambda_q\theta_{qj}^{(1)}, 1)$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_Q)$ and $\boldsymbol{\theta}_j = (\theta_{1j}, \ldots, \theta_{Qj})^T$. However, it is not clear how this model is identified and how the coefficients $\boldsymbol{\lambda}$ are interpreted.

de la Torre & Song (2009) also proposed a second order multi-unidimensional IRT for continuous latent traits, in which an examinee's performance in each specific domain is represented by a specific latent trait, $\theta_{qj}^{(1)}$, for $q = 1, \ldots, Q$, and subject $j = 1, \ldots, J$. Correlations among specific latent traits are explained by a single higher order latent trait that can be viewed as the subject's general latent trait, $\theta_j^{(2)}$.

On the item level, the authors consider a 3PL model, given by

$$P(Y_{ijq} = 1|\theta_{qj}^{(1)}, a_{iq}, b_{iq}, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-(a_{iq}(\theta_{qj}^{(1)} - b_{iq}))}}. \tag{2.6}$$

The latent structure considered is given by

$$\theta_{qj}^{(1)} = \lambda_q \theta_j^{(2)} + \epsilon_{qj}, \tag{2.7}$$

where $\theta_j^{(2)} \overset{iid}{\sim} N(0,1)$, ϵ_{qj} is the error term that is independent of other error terms and of $\theta_j^{(2)}$, and such that $\epsilon_{qj} \sim N(0, 1 - \lambda_q^2)$, which implies that $\theta_{qj}^{(1)} \sim N(0,1)$ marginally. Under this formulation, it can be shown that λ_q is the correlation between $\theta_{qj}^{(1)}$ and $\theta_j^{(2)}$, and $Corr(\theta_{sj}^{(1)}, \theta_{tj}^{(1)}) = \lambda_s \lambda_t$, for $s \neq t$. Also, given the general latent trait $\theta_j^{(2)}$, the specific latent traits are independent.

Notice that this model structure is similar to that proposed by Sheng & Wikle (2008) as shown by Equation (2.7). The main difference would be the constrain, since de la Torre & Song (2009) constrain both $\theta_j^{(2)}$ and $\theta_{qj}^{(2)}$ to be normally distributed with mean zero and variance one, by setting $\epsilon_{qj} \sim N(0, 1 - \lambda_q^2)$. This also implies that both $\theta_j^{(2)}$ and $\theta_{qj}^{(1)}$ are, marginally, in the same scale. On the other hand, Sheng & Wikle (2008) constrain $\epsilon_{qj} \overset{iid}{\sim} N(0,1)$ and $\theta_j^{(2)} \overset{iid}{\sim} N(0,1)$, imposing that $\theta_{qj}^{(1)} \sim N(0, 1 + \lambda_q^2)$. While in the constrain imposed by de la Torre & Song (2009) λ_q is the correlation between $\theta_{qj}^{(1)}$ and $\theta_j^{(2)}$, in Sheng & Wikle (2008) λ_q is the covariance between $\theta_{qj}^{(2)}$ and $\theta_j^{(2)}$. We show in Subsection 2.1 the advantages of de la Torre & Song (2009) specification over Sheng & Wikle (2008). In de la Torre & Song (2009), Bayesian inference is performed via MCMC in a Metropolis Hastings within Gibbs algorithm.

To solve model indeterminacy, however, when only two specific latent traits were involved, the authors constrained the regression parameters to be equal, i.e., $\lambda_1 = \lambda_2$. This implies that the general latent trait is centered around the average of the specific latent traits. Finally, item parameters were assumed to be known, which is a highly restrictive assumption.

Another work has been proposed in the direction of modeling a hierarchical structure for the latent variables under and IRT framework. Sheng & Wikle (2009) proposed a model in which both general and specific latent traits form an additive structure, so that each item measures both general an specific latent traits directly. The proposed model is given by equation

$$p(Y_{ijq} = 1 | a_i, a_{iq}, b_i, \theta_j, \theta_{qj}) = \Phi(a_i \theta_j + a_{iq} \theta_{qj} - b_i), \tag{2.8}$$

in which θ_j and θ_{kj} are the general and specific latent traits, respectively, for subject j, a_i is the i-th item's discrimination parameter associated with the general latent trait, a_{iq} is the discrimination parameter associated to the specific latent trait θ_{qj} and b_i is the item difficulty parameter.

The proposed model is very similar to the bifactor model proposed by Gibbons & Hedeker (1992) (see Equation 2.1). Accordingly to Sheng & Wikle (2009), one main difference would be that the bifactor model assumes zero correlation among specific latent traits, while the additive model allows the modeling of their interdependence by using a covariance structure. As noticed by the authors, the model presented in Equation (2.3) could be rewritten as $\alpha_{iq}\beta_q\theta_j + \alpha_{iq}\epsilon_{qj} - b_{iq}$, where $\epsilon_{qj} \sim N(0, 1)$. Under this formulation, one could set $\beta_q = 1$, which would make the additive model a constrained version of the hierarchical model defined in Equation (2.3). However, although the models have similar structure, they differ substantially in their interpretation. In the additive model, θ_{qj} denotes the specific latent trait for the q^{th} subtest, and it can be correlated with other specific latent traits and with the general latent trait θ_j (Sheng & Wikle, 2009). In the hierarchical model (2.3), ϵ_{qj} are independent random error related to the q^{th} subtest with zero correlation with the general latent trait.

Based on Sheng & Wikle (2009) and de la Torre & Song (2009) works, Huang et al. (2013) added new hierarchical levels for the ability and extended its use to both dichotomous and polytomous items. Also, their model allows a lower-order trait to be governed by more than one higher order trait (Huang et al., 2013). Their model supposes that each of the first order latent traits is measured by a set of unidimensional items, i.e., a multi-unidimensional approach.

To specify their model, let $\theta_{qj}^{(k)}$ be the k^{th}-order latent trait q for subject j, such that the relationship between $\theta_{qj}^{(k)}$ and its higher order $\boldsymbol{\theta}_j^{(k+1)}$ is given by

$$\boldsymbol{\theta}_j^{(k)} = \boldsymbol{\lambda}^{(k)}\boldsymbol{\theta}_j^{(k+1)} + \epsilon_{qj}^{(k)}, \tag{2.9}$$

in which $\boldsymbol{\lambda}'^{(k)}$ the k^{th} order regression vector and $\epsilon_{qj}^{(k)}$ is the q^{th} order residual for latent trait q and subject j. For only two orders in the hierarchical level, the first order latent traits are governed by the same second order trait, so the structure becomes

$$\theta_{qj}^{(1)} = \lambda_q \theta_j^{(2)} + \epsilon_{qj}^{(1)}. \tag{2.10}$$

As it is noticed by the authors, in the second order model, latent traits could be also estimated indirectly with the multi staged confirmatory factor analysis (CFA) or the average approach, that is, to use the average of all first level latent traits as an estimate for the second level trait. However, these two approaches are less efficient than the direct estimation in the single-staged approach. In hierarchical IRT models, each person can obtain a parameter estimate together with a corresponding standard error for each latent trait at any order.

Model (2.10) has the same structure of the model presented by de la Torre & Song (2009). However Huang et al. (2013) extended the model to more than two levels and included item parameters estimation. Some important issues, however, are not mentioned by Huang et al. (2013). First, for model identification, the relation between the number of observed variables (p) and number of factors (m) must satisfy $p \geq 2m + 1$ (see Anderson & Rubin (1956)). Huang et al. (2013) explored examples in which this condition is not met and, in order to identify the model, one of the factors loadings was constrained to be one. However, this approach is quite problematic, as factors loadings are correlations between a lower and higher order latent traits. This means that a constrain such as that is not realistic. Moreover, their model allows more than one higher trait to influence a lower trait as illustrated in Figure (2.2). In doing so, latent traits are not in the same scale and model identifiability may be compromised. To see this, notice from Figure (2.2) that

$$\begin{aligned} \theta_2^{(1)} &= \lambda_1\theta_1^{(2)} + \lambda_2\theta_2^{(2)} + \epsilon_2^{(1)}, \\ \theta_2^{(1)} &\sim N(0, 1 + 2\lambda_1^2\lambda_2^2), \end{aligned} \tag{2.11}$$

since $Corr[\theta_1^{(2)}, \theta_2^{(2)}] = \lambda_1\lambda_2$. However, $\theta_1^{(1)} \sim N(0,1)$, which shows that latent traits marginals are in different scales, in particular, first level latent traits are themselves in different scales. Moreover, Huang et al. (2013) did not explain which restrictions were imposed on the variance of $\boldsymbol{\epsilon}$ when more than two latent traits of a higher level influences on a lower trait, since, in this case, $\epsilon \sim N(0, 1 - \lambda_1^2 - \lambda_2^2)$. Notice that if $\lambda_1^2 = \lambda_2^2 = 0.9$, for instance, the variance would be negative.

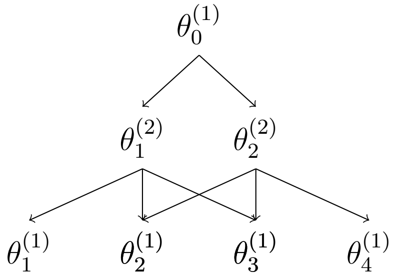


Figure 2.2: Huang et al. (2013) model.

2.1 Model's Equivalence

We now discuss the equivalence of the models propose by Sheng & Wikle (2008), de la Torre & Song (2009) and Huang et al. (2013), as well as the advantages and problems of each one.

Consider the factor equation with only two levels

$$\theta_k^{(1)} = \lambda_k\theta^{(2)} + \epsilon_k^{(1)}, \tag{2.12}$$

where $\theta_k^{(1)}$ denotes the k^{th} ability on the first level. The subscript of the j^{th} student is omitted.

A first model specification, defined as Model 1, is given by

$$\theta^{(2)} \sim N(0,1) \quad \text{and} \quad \epsilon_k^{(1)} \sim N(0, 1-\lambda_k^2), \tag{2.13}$$

which implies that $\theta_k^{(1)}|\lambda_k, \theta^{(2)} \sim N(\lambda_k \theta^{(2)}, 1-\lambda_k^2)$ and $\theta_k^{(1)} \sim N(0,1)$.

Another possible specification for Equation (2.12), defined as Model 2, is

$$\theta^{(2)} \sim N(0,1), \quad \text{and} \quad \epsilon_k^{(1)} \sim N(0,1), \tag{2.14}$$

which implies that $\theta_k^{(1)}|\lambda_k, \theta^{(2)} \sim N(\lambda_k \theta^{(2)}, 1)$ and $\theta_k^{(1)} \sim N(0, 1+\lambda_k^2)$.

Models 1 and 2 are actually the same but under different parametrization. If we take Model 2 and divide all terms by $\sqrt{1+\lambda_k^2}$, we get

$$\theta_k^{*(1)} = \gamma_k \theta^{(2)} + \epsilon_k^{*(1)}, \tag{2.15}$$

where $\theta_k^{*(1)} \sim N(0,1)$, $\gamma_k = \frac{\lambda_k}{\sqrt{1+\lambda_k^2}}$, and $\epsilon_k^{*(1)} \sim N(0, (1+\lambda_k^2)^{-1})$. However,

$$\lambda_k = \frac{\gamma_k}{\sqrt{1-\gamma_k^2}}, \tag{2.16}$$

which implies that

$$\epsilon_k^{*(1)} \sim N(0, 1-\gamma_k^2) \tag{2.17}$$

and Model 1 is recovered.

Another possible model specification, defined as Model 3, which is a generalization of Model 2, is given by

$$\theta^{(2)} \sim N(0,1), \quad \text{and} \quad \epsilon_k^{(1)} \sim N(0, \sigma_k^2), \tag{2.18}$$

which implies that $\theta_k^{(1)}|\lambda_k, \theta^{(2)} \sim N(\lambda_k \theta^{(2)}, \sigma_k^2)$ and $\theta_k^{(1)} \sim N(0, \sigma_k^2 + \lambda_k^2)$.

In a standard factor analysis set up, Models 1 and 2 consist of model restrictions. However, under the IRT framework, the $\boldsymbol{\theta}^{(1)}$ are not directly observed like in the factor analysis, and, therefore, the model is not identifiable. Some possible solutions include fixing $\prod a_i = 1$ and

$\sum b_i = 0$. The more intuitive restriction, however, is to fix the scale of the latent trait, say $\theta_k^{(1)} \sim N(0,1)$. Therefore, the parametrization used in Model 1 seems to be the most advantageous one.

2.2 Discussion

The parametrization adopted in Sheng & Wikle (2008) does not fix the scale of the latent traits. Nevertheless, the authors do not mention how model identifiability is achieved. Furthermore, the marginal distribution of the latent traits are not on the same scale, which makes model interpretation rather harder. de la Torre & Song (2009) also proposed a model in which both overall and specific latent traits (item level latent traits) are simultaneously estimated. Nevertheless, item parameters are assumed to be known, which is an unrealistic assumption. Huang et al. (2013) also proposed a higher order model, extending the work of de la Torre & Song (2009) to more than two orders and also simultaneously estimating item parameters, but under some inference drawbacks. Moreover, no author fully explored the inference process. The high dimensionality and correlation structure of the model makes inference a considerable challenging task. This means that any proposed MCMC algorithm should be carefully devised. This issue is not properly addressed in any of the references discussed here. For instance, Huang et al. (2013) only mentioned the Bayesian is performed using WinBUGS.

Motivated by the limitations of the approaches in the existent literature and on the importance of the theme, this work aims to propose an improved methodology in terms of both modeling and inference. From a modeling perspective, the proposed methodology allows for genuinely multidimensional items, with a higher order structure for the latent traits, in which all of them are in the same scale and model identifiability is guaranteed. We approach computational aspects, such as sampling jointly all latent traits of a same subject and we carefully devise a MCMC algorithm to be efficient. Also, we prove some model properties that the past authors had not done. We highlight that all the above references consider only a multi-unidimensional structure and we introduce the modeling and estimation of multidimensional items.

Chapter 3

Multidimensional IRT Model with Hierarchical Structure for the latent traits

Let $\boldsymbol{\theta}_j^{(k)}$ denote the latent trait vector $(\theta_{1j}^{(k)} \dots \theta_{Q_k j}^{(k)})^T$ for subject $j, j = 1, \dots, J$, on hierarchical level k, $k = 1, \dots, K$, $q = 1, \dots, Q_k$, where Q_k denotes the number of latent traits on level k. The following hierarchical structure is considered for the latent traits:

$$\boldsymbol{\theta}_j^{(k)} = \boldsymbol{\lambda}^{(k+1)} \boldsymbol{\theta}_j^{(k+1)} + \boldsymbol{\epsilon}_j^{(k)}, \qquad k = 1, \dots, K-1, \tag{3.1}$$

where $\boldsymbol{\lambda}^{(k+1)}$ is the coefficient matrix $Q_k \times Q_{k+1}$ relating $\boldsymbol{\theta}_j^{(k)}$ and $\boldsymbol{\theta}_j^{(k+1)}$, $\forall\, j$, and $\boldsymbol{\epsilon}_j^{(k)}$ is a vector $Q_k \times 1$ containing the error terms. We constrain $\boldsymbol{\lambda}^{(k+1)}$ such that each $\theta_{qj}^{(k)}$ is directly linked to only one trait of higher order $(k+1)$.

Let $\boldsymbol{\theta}_j = (\boldsymbol{\theta}_j^{(1)}, \dots, \boldsymbol{\theta}_j^{(K)})$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1 \dots \boldsymbol{\theta}_J)^T$, $\boldsymbol{\epsilon}_j = (\boldsymbol{\epsilon}_j^{(1)}, \dots, \boldsymbol{\epsilon}_j^{(K-1)})$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1 \dots \boldsymbol{\epsilon}_j)^T$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{(2)} \dots \boldsymbol{\lambda}^{(K)})$. To completely specify the proposed Bayesian model, it is assumed that

C1

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \left[\prod_{j}^{J} \prod_{q} \pi(\theta_{qj}^{(K)})\right] \left[\prod_{j}^{J} \prod_{k=1}^{K-1} \prod_{q}^{Q_k} \pi(\theta_{qj}^{(k)} | \lambda_{qq'}^{(k+1)}, \theta_{q'j}^{(k+1)})\right] \left[\prod_{k=2}^{K} \prod_{q=1}^{Q_k} \prod_{q'=1}^{Q_{k+1}} \pi(\lambda_{qq'}^{(k)})\right]. \tag{3.2}$$

C2

$$\theta_{qj}^{(K)} \overset{iid}{\sim} N(0,1), \quad q = 1, \ldots, Q_K. \tag{3.3}$$

C3

$$(\epsilon_{qj}^{(k)} | \lambda_{qq'}^{(k+1)}) \overset{ind}{\sim} N(0, 1 - (\lambda_{qq'}^{(k+1)})^2). \tag{3.4}$$

Both specifications in (3.3) and (3.4) have the advantage of establishing model identification and setting a common scale for all the latent traits in the model, as stated in the following proposition.

Proposition 1. *Given Equation (3.1) and conditions **C1**-**C3**, the marginal prior distribution of any $\theta_{qj}^{(k)}$ is normal with mean zero and variance one, i.e.,*

$$\theta_{qj}^{(k)} \sim N(0,1), \quad \forall \quad k, q, j. \tag{3.5}$$

Proof. Condition **C1** implies that $\boldsymbol{\theta}^{(K)}$ is independent of $\boldsymbol{\lambda}^{(K)}$, therefore

$$\begin{aligned}
\pi(\theta_{qj}^{(K-1)}) &= \int\int \pi(\theta_{qj}^{(K-1)}, \theta_{qj}^{(K)}, \lambda_q^{(K)}) \quad d\theta_{qj}^{(K)} d\lambda_q^{(K)} \\
&= \int\int \pi(\theta_{qj}^{(K-1)} | \theta_{qj}^{(K)}, \lambda_q^{(K)}) \pi(\theta_{qj}^{(K)}, \lambda_q^{(K)}) \quad d\theta_{qj}^{(K)} d\lambda_q^{(K)} \\
&= \int\int \pi(\theta_{qj}^{(K-1)} | \theta_{qj}^{(K)}, \lambda_q^{(K)}) \pi(\theta_{qj}^{(K)}) \pi(\lambda_q^{(K)}) \quad d\theta_{qj}^{(K)} d\lambda_q^{(K)} \\
&= \int_\lambda \left[\int_\theta \pi(\theta_{qj}^{(K-1)} | \theta_{qj}^{(K)}, \lambda_q^{(K)}) \pi(\theta_{qj}^{(K)}) d\theta_{qj}^{(K)} \right] \pi(\lambda_q^{(K)}) \quad d\lambda_q^{(K)} \\
&= \int_\lambda \phi(\theta_{qj}^{(K-1)}) \pi(\lambda_q^{(K)}) \quad d\lambda_q^{(K)} \\
&= \phi(\theta_{qj}^{(K-1)}) \int_\lambda \pi(\lambda_q^{(K)}) \quad d\lambda_q^{(K)} \\
&= \phi(\theta_{qj}^{(K-1)}).
\end{aligned} \tag{3.6}$$

For $k = K-2$, it follows that

$$\pi(\theta_{qj}^{(K-2)}) = \int\int \pi(\theta_{qj}^{(K-2)} | \theta_{q'j}^{(K-2)}, \lambda_q^{(K-2)}) \pi(\theta_{q'j}^{(K-1)}, \lambda_q^{(K-1)}) \quad d\theta_{q'j}^{(K-1)} d\lambda_q^{(K-1)} \tag{3.7}$$

where

$$\begin{aligned}\pi(\theta_{q'j}^{(K-1)}, \lambda_q^{(K-1)}) &= \int\int \pi(\theta_{q'j}^{(K-1)}|\theta_{q''j}^{(K)}, \lambda_{q'}^{(K)})\pi(\lambda_{qj}^{(K-1)}|\theta_{q''j}^{(K)}, \lambda_{q'}^{(K)})\pi(\lambda_{q'}^{(K)})\pi(\theta_{q'j}^{(K)})d\lambda_{q'}^{(K)}d\theta_{q'j}^{(K)}\\ &= \pi(\lambda_q^{(K-1)})\phi(\theta_{q'j}^{(K-1)}).\end{aligned} \tag{3.8}$$

For $k \leq K-3$, the proof is analogous and devised recursively such that the result in (3.8) for step k is used on the step for $k-1$. $\square$

Equation (3.1) and **C1-C3** also yields that for each subject j and for $k = 1, \ldots, K-1$

$$\boldsymbol{\theta}_j^{(k)}|\boldsymbol{\theta}_j^{(k+1)}, \boldsymbol{\lambda}^{(k+1)} \overset{ind}{\sim} N(\boldsymbol{\lambda}^{(k+1)}\boldsymbol{\theta}_j^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}), \tag{3.9}$$

where

$$\boldsymbol{\Sigma}^{(k+1)} = diag\left(1 - \sum_{q'=1}^{Q_{k+1}} (\lambda_{1q'}^{(k+1)})^2, \ldots, 1 - \sum_{q'=1}^{Q_{k+1}} (\lambda_{Q_kq'}^{(k+1)})^2\right). \tag{3.10}$$

Following the set notation, $\theta_{qj}^{(1)}$ denotes the j^{th} subject's q^{th} trait at the lowest level, that is, the q^{th} ability that is directly measured by items. For identifiability purposes, we impose that $Q_k \geq 2Q_{k+1} + 1$.

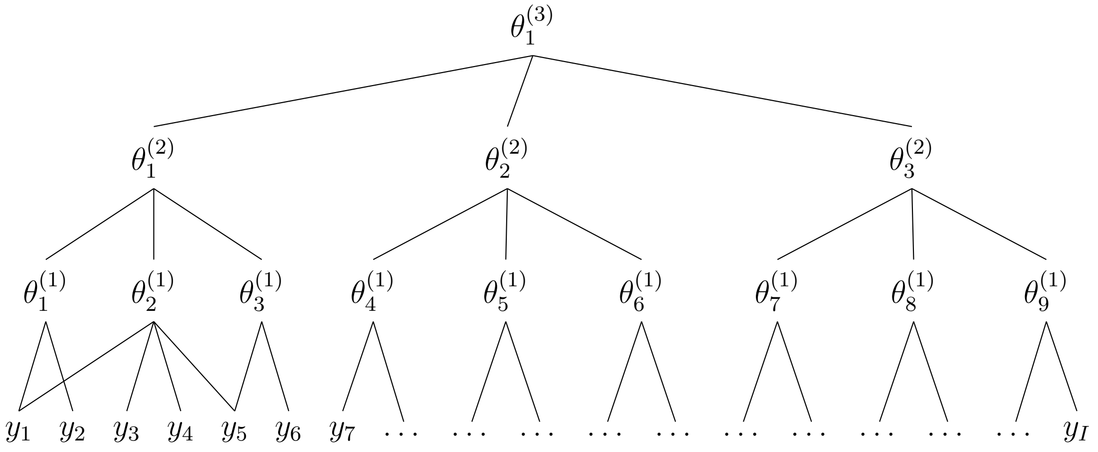


Figure 3.1: Hierarchical structure of order 3 for the latent traits.

For instance, take the structure illustrated by Figure (3.1), in which there are three levels, where the first level has nine traits ($Q_1 = 9$), the second level has three traits ($Q_2 = 3$) and the highest level has one trait ($Q_3 = 1$). Also, y_i are the item responses. In a matrix notation, for each subject j, the relation between the first and second orders is expressed as

$$\begin{pmatrix} \theta^{(1)}_{1j} \\ \theta^{(1)}_{2j} \\ \theta^{(1)}_{3j} \\ \theta^{(1)}_{4j} \\ \theta^{(1)}_{5j} \\ \theta^{(1)}_{6j} \\ \theta^{(1)}_{7j} \\ \theta^{(1)}_{8j} \\ \theta^{(1)}_{9j} \end{pmatrix} = \begin{pmatrix} \lambda^{(2)}_{11} & 0 & 0 \\ \lambda^{(2)}_{21} & 0 & 0 \\ \lambda^{(2)}_{31} & 0 & 0 \\ 0 & \lambda^{(2)}_{42} & 0 \\ 0 & \lambda^{(2)}_{52} & 0 \\ 0 & \lambda^{(2)}_{62} & 0 \\ 0 & 0 & \lambda^{(2)}_{73} \\ 0 & 0 & \lambda^{(2)}_{83} \\ 0 & 0 & \lambda^{(2)}_{93} \end{pmatrix} \begin{pmatrix} \theta^{(2)}_{1j} \\ \theta^{(2)}_{2j} \\ \theta^{(2)}_{3j} \end{pmatrix} + \begin{pmatrix} \epsilon^{(1)}_{1j} \\ \epsilon^{(1)}_{2j} \\ \epsilon^{(1)}_{3j} \\ \epsilon^{(1)}_{4j} \\ \epsilon^{(1)}_{5j} \\ \epsilon^{(1)}_{6j} \\ \epsilon^{(1)}_{7j} \\ \epsilon^{(1)}_{8j} \\ \epsilon^{(1)}_{9j} \end{pmatrix},$$

and the relation between the second and third levels is

$$\begin{pmatrix} \theta^{(2)}_{1j} \\ \theta^{(2)}_{2j} \\ \theta^{(2)}_{3j} \end{pmatrix} = \begin{pmatrix} \lambda^{(3)}_{11} \\ \lambda^{(3)}_{21} \\ \lambda^{(3)}_{31} \end{pmatrix} \begin{pmatrix} \theta^{(3)}_{1j} \end{pmatrix} + \begin{pmatrix} \epsilon^{(2)}_{1j} \\ \epsilon^{(2)}_{2j} \\ \epsilon^{(2)}_{3j} \end{pmatrix}.$$

3.1 The Model

Assume that a random sampling of J subjects is submitted to a test composed by I items. Denote by $\boldsymbol{Y} = (Y_{ij})_{I\times J}$, the $I \times J$ matrix of all responses to the test, where Y_{ij} is the indicator of a correct response given to item i by individual j, for i, $i = 1, \ldots, I$ and $j = 1, \ldots, J$. The proposed model is defined as follows

$$\begin{aligned} Y_{ij}|p_{ij} &\overset{ind}{\sim} Bernoulli(p_{ij}), \\ p_{ij} &= c_{ij} + (1 - c_{ij})\Phi(\mathbf{a}_i\boldsymbol{\theta}^{(1)}_j - b_i), \end{aligned} \tag{3.11}$$

where $\mathbf{a}_i = (a_{i1} \dots a_{iQ_1})$ is the vector of the discrimination parameters for item i, $\boldsymbol{\theta}_j^{(1)} = (\theta_{1j}^{(1)}, \dots, \theta_{Q_1j}^{(1)})^T$ is the vector of the specific latent traits for subject j, b_i and c_i, the intercept term and pseudo guessing parameter, respectively, for item i. The link function $\Phi(.)$ is the c.d.f. of the standard Normal distribution. Differently from an unidimensional IRT model, the proposed model allows item i to measure more than one latent trait. In order to identify model (3.11), some restrictions must be imposed. One possibility is $a_{iq} = 0$, for $i = 1, \dots, Q_1 - 1$, and $q = i+1, \dots, Q_1$ (Béguin & Glas, 2001).

To model the graded response items, we use the Graded Response Model (GRM) (Samejima, 1969), in which item responses are classified into ordered categories. Assume that there are M_i categories, denoted by $m = 1, \dots, M_i$. Therefore, for each graded item i, there are $M_i - 1$ location parameters b_{im}, such that

$$-\infty = b_{i0} < b_{i1} < b_{i2} < \dots < b_{i,M_i-1} < b_{i,M_i} = \infty. \tag{3.12}$$

Define $\boldsymbol{b^G} = (\boldsymbol{b}_1^G, \dots, \boldsymbol{b}_I^G)$, $\boldsymbol{b}_i^G = (b_{i1}, \dots, b_{i,M_i-1})$, $\mathbf{a}^G = (\boldsymbol{a}_1^G, \dots, \boldsymbol{a}_I^G)$ and $\boldsymbol{a}_i^G = (a_{i1}^G, \dots, a_{i,Q_1}^G)$. The GRM assumes that the probability that a subject j receives a score m is

$$P(Y_{ij} = m) = \Phi(b_{im} - \boldsymbol{a}_i^G \boldsymbol{\theta}_j) - \Phi(b_{im-1} - \boldsymbol{a}_i^G \boldsymbol{\theta}_j). \tag{3.13}$$

Models (3.11) and (3.13) can be jointly applied in tests composed by both dichotomous and graded response items.

The likelihood function of model (3.11) is given by

$$L(\boldsymbol{\theta}^{(1)}, \mathbf{a}, \mathbf{b} | \mathbf{Y}) = \prod_{j=1}^{J} \prod_{i=1}^{I} p_{ij}^{y_{ij}} (1 - p_{ij})^{1 - y_{ij}}, \tag{3.14}$$

in which $p_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_j^{(1)}, \mathbf{a}_i, b_i)$, $\mathbf{a} = (\mathbf{a}_1 \dots \mathbf{a}_I)$, $\boldsymbol{\theta}^{(1)} = (\boldsymbol{\theta}_{1,j}^{(1)}, \dots, \boldsymbol{\theta}_{Q_1,j}^{(1)})$, and $\mathbf{b} = (b_1, \dots, b_I)^T$. For graded response item, the likelihood function is given by

$$L_2(\boldsymbol{\theta}^{(1)}, \mathbf{a}^G, \mathbf{b}^G | \mathbf{Y}) = \prod_{j=1}^{J} \prod_{i=1}^{I} \prod_{m=1}^{M_i} [\Phi(b_{im} - \boldsymbol{a}_i^G \boldsymbol{\theta}_j^{(1)}) - \Phi(b_{im-1} - \boldsymbol{a}_i^G \boldsymbol{\theta}_j^{(1)})]^{\mathbb{I}(Y_{ij} = m)}, \tag{3.15}$$

in which $\mathbb{I}(Y_{ij} = m) = 1$ if subject j receives a score m in item i and it is 0 otherwise.

For tests composed by both dichotomous and graded response items, the likelihood function is is the corresponding product of terms in (3.14) and (3.15).

3.2 Prior Specifications

Define $\boldsymbol{\eta}_i = (\mathbf{a}_i, b_i)^T$, a vector $(Q_1 + 1) \times 1$. *A priori*, $\mathbf{a}_i$ and b_i are independent such that $\boldsymbol{\eta}_i \overset{ind}{\sim} N(\boldsymbol{\mu}_\eta, \boldsymbol{\Lambda}_\eta)$, where $\boldsymbol{\mu}_\eta$ is a $(Q_1 + 1) \times 1$ vector, and $\boldsymbol{\Lambda}_\eta$ is a $(Q_1 + 1) \times (Q_1 + 1)$ covariance matrix. Both hyperparameters, $\boldsymbol{\mu}_\eta$ and $\boldsymbol{\Lambda}_\eta$, are assumed to be known *a priori*, with $\boldsymbol{\Lambda}_\eta = diag(\sigma^2_{a_1}, \ldots, \sigma^2_{a_{Q_1}}, \sigma^2_{b_i})$.

For the graded response items, we consider the same prior distribution of the first Q_1 coordinates of $\boldsymbol{\eta}_i$ for $\boldsymbol{a}^G$ and a prior $N(\mathbf{0}, \mathbf{1})\mathbb{I}[b_{i1} < \cdots < b_{i,M_i}]$ for location parameters.

Parameters $\boldsymbol{\lambda}$ are constrained to be in [-1,1]. More specifically we set a Uniform(-1,1) prior for these parameters.

3.3 Bayesian Inference

Under the Bayesian approach, inference is based on the joint posterior distribution of the unknown parameters and unobservable variables of the model. The posterior distribution is the conditional distribution of all of those quantities given the item responses. Bayes Theorem implies that the density function of the posterior distribution is proportional to the product of the likelihood function and the prior densities of all model's unknown quantities. The hierarchical structure presented in Section 3.2 defines the prior distributions for all latent traits.

3.3.1 Auxiliary variables

Exploring analytically the target posterior distribution is impossible due to its high dimension and intractability. Instead, we obtain an (approximate) sample from it to perform Monte Carlo estimation of suitable posterior expectations. Such a sample is obtained via Markov Chain Monte Carlo (MCMC) and, in order to ease the derivation of an efficient algorithm, we introduce some auxiliary variables, proposed by Bambirra Gonçalves et al. (2018), that preserves the original model upon marginalization.

Define

$$Z_{ij} \sim Bernoulli(c_i), \qquad (X_{ij}|Z_{ij}) \sim N(\mathbf{a}_i\boldsymbol{\theta}_j - b_i, 1)\mathbb{I}(Z_{ij} = 0) + \delta_0\mathbb{I}(Z_{ij} = 1), \tag{3.16}$$

where $\mathbb{I}$ is the indicator function and δ_0 is a point-mass at 0, that is, $P(X_{ij} = 0|Z_{ij} = 1) = 1$.

The auxiliary variables $(\mathbf{Z}, \mathbf{X})$ are introduced in the model such that

$$Y_{ij} = \begin{cases} 1, & \text{if} \quad (Z_{ij} = 1) \quad \text{or} \quad (Z_{ij} = 0, X_{ij} \geq 0), \\ 0, & \text{if} \quad (Z_{ij} = 0, X_{ij} < 0), \end{cases} \tag{3.17}$$

where $\mathbf{Z} = \{Z_{ij}\}$, $\mathbf{X} = \{X_{ij}\}$, $\forall\ i, j$.

The model with the $\mathbf{X}$ and $\mathbf{Z}$ variables is a hierarchical version of the model in (3.11) which can be recovered by integrating out $(\mathbf{X}, \mathbf{Z})$. Also, notice that if $c_i = 0$, model (3.17) becomes

$$Y_{ij} = \begin{cases} 1, & \text{if} \quad X_{ij} \geq 0, \\ 0, & \text{if} \quad X_{ij} < 0, \end{cases} \tag{3.18}$$

in which $X_{ij} \sim N(\mathbf{a}_i\boldsymbol{\theta}_j^{(1)} - b_i, 1)$

For the model in (3.13), the following hierarchical representation is considered (Johnson & Albert, 2006)

$$Y_{ij} = \begin{cases} 0, & \text{if} \quad -\infty < X_{ij}^G < b_{i1}, \\ 1, & \text{if} \quad b_{i1} < X_{ij}^G < b_{i2}, \\ \vdots & \\ M_i, & \text{if} \quad b_{iM_i-1} < X_{ij}^G < \infty, \end{cases} \tag{3.19}$$

with $X_{ij}^G \overset{ind}{\sim} N(\mathbf{a}_i^G\boldsymbol{\theta}_j^{(1)}, 1)$.

The density of the posterior distribution under the augmented model is proportional to the joint density Y and $\Psi = (\mathbf{Z}, \mathbf{X}, \mathbf{X}^G, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{a}^G, \mathbf{b}^G)$ which is given by:

$$\begin{aligned}
\pi(\mathbf{Y}, \boldsymbol{\Psi}) =& \pi(\mathbf{Y}|\mathbf{X}, \mathbf{Z})\pi(\mathbf{X}|\mathbf{Z}, \mathbf{a}, \mathbf{b}, \boldsymbol{\theta}^{(1)})\pi(\mathbf{X}^G|\mathbf{a}^G, \mathbf{b}^G, \boldsymbol{\theta}^{(1)})\pi(\mathbf{Z}|\mathbf{c})\pi(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(2)}, \boldsymbol{\lambda}^{(2)})\ldots\pi(\boldsymbol{\theta}^{(K-1)}|\boldsymbol{\theta}^{(K)}, \boldsymbol{\lambda}^{(K)}) \\
& \times \pi(\mathbf{a}, \mathbf{b})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\theta}^{(K)})\pi(\mathbf{c})\pi(\mathbf{b}^G)\pi(\mathbf{a}^G). \\
=& \left[\prod_{i=1}^{I}\prod_{j=1}^{J}\pi(Y_{ij}|X_{ij}, Z_{ij})\pi(X_{ij}|Z_{ij}, \boldsymbol{\theta}_j^{(1)}, \mathbf{a}_i, b_i)\pi(Z_{ij}|c_i)\right]\left[\prod_{i=1}^{I}\prod_{j=1}^{J}\pi(X_{ij}^G|\boldsymbol{\theta}_j^{(1)}, \mathbf{a}_i^G, \mathbf{b}_i^G)\right] \\
& \times \left[\prod_{j=1}^{J}\prod_{k=1}^{K-1}\prod_{q=1}^{Q_1}\pi(\theta_{qj}^{(k)}|\boldsymbol{\theta}_j^{(k+1)}, \boldsymbol{\lambda}^{(k+1)})\right]\left[\prod_{q=1}^{Q_K}\prod_{j=1}^{J}\pi(\theta_{qj}^{(K)})\right]\left[\prod_{i=1}^{I}\prod_{q=1}^{Q_1}\pi(a_{iq}, b_i)\right]\left[\prod_{k=2}^{K}\prod_{q=1}^{Q_k}\prod_{q'=1}^{Q_{k+1}}\pi(\lambda_{qq'}^{(k)})\right] \\
& \times \left[\prod_{i=1}^{I}\pi(c_i)\right]\left[\prod_{i=1}^{I}\prod_{q=1}^{Q_1}\pi(a_{iq}^G)\right]\left[\prod_{i=1}^{I}\prod_{c=1}^{M_i-1}\pi(b_{ic}^G)\right].
\end{aligned} \tag{3.20}$$

3.3.2 MCMC algorithm

We devise a Gibbs sampling with blocks

$$(\mathbf{X}, \mathbf{Z}), \quad (\mathbf{a}, \mathbf{b}), \quad (\boldsymbol{\lambda}), \quad (\boldsymbol{\theta}), \quad (\boldsymbol{c}).$$

For tests with graded response items, the following blocks should be included

$$(\boldsymbol{X}^G), \quad (\boldsymbol{a}^G), \quad (\boldsymbol{b}^G).$$

We actually present and compare two possible blocking schemes. In the first one, all $\boldsymbol{\theta}$ is sampled together and in the second scheme, all $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are sampled together. Comparisons among the blocking schemes are presented in Section (A). All proposed schemes take advantage of the conditional independence among several components, which makes it straightforward to sample from the high dimensional blocks.

We now present the full conditional distribution for all the blocking schemes.

Full Conditional Distribution for X and Z

Given parameters $\boldsymbol{\theta}$, $\mathbf{a}$ and $\mathbf{b}$, it follows that the pairs (Z_{ij}, X_{ij}) are conditionally independent for all i and j, which means that we can sample the vector ($\mathbf{Z}$,$\mathbf{X}$) by sampling each pair (Z_{ij}, X_{ij}) individually. The full conditional posterior distribution is given by

$$Z_{ij}, X_{ij}|. = \begin{cases} \phi(x_{ij} - m_{ij})\mathbb{I}(Z_{ij} = 0)\mathbb{I}(X_{ij} < 0), \quad \text{if} \quad Y_{ij} = 0, \\ w_{ij}\mathbb{I}(Z_{ij} = 1)\mathbb{I}(X_{ij} = 0) + (1 - w_{ij})\frac{\phi(x_{ij} - m_{ij})}{\Phi(m_{ij})}\mathbb{I}(Z_{ij} = 0)\mathbb{I}(X_{ij} > 0) \quad \text{if} \quad Y_{ij} = 1, \end{cases} \tag{3.21}$$

in which $m_{ij} = \mathbf{a}_i\boldsymbol{\theta}_j^{(1)} - b_i$ and $w_{ij} = \frac{c_i}{c_i + (1 - c_i)\Phi(m_{ij})}$.

When $c_i = 0$,

$$(X_{ij}|\boldsymbol{\theta}^{(1)}, \mathbf{a}, \mathbf{b}, \mathbf{Y}) \sim \begin{cases} N(\mathbf{a}_i\boldsymbol{\theta}_j^{(1)} - b_i, 1)\mathbb{I}(X_{ij} > 0), \quad \text{if} \quad Y_{ij} = 1, \\ N(\mathbf{a}_i\boldsymbol{\theta}_j^{(1)} - b_i, 1)\mathbb{I}(X_{ij} < 0), \quad \text{if} \quad Y_{ij} = 0, \end{cases} \tag{3.22}$$

such that the full conditional posterior distribution of X_{ij} is a truncated Normal distribution in $\mathbb{R}^+$, if $Y_{ij} = 1$, and in $\mathbb{R}^-$, if $Y_{ij} = 0$.

Full Conditional Distribution for $\boldsymbol{\theta}$

We present the two sampling algorithms for the latent traits and these are valid for both dichotomous and graded response items. For the latter, X_{ij} is as defined by (3.19) and $b_i = 0$.

Full Joint Conditional Distribution for $\boldsymbol{\theta}$

Sampling the whole vector $\boldsymbol{\theta}$ jointly is possible due to the conditional independence of the respective full conditional distributions among different subjects. The joint full conditional distribution of all $\boldsymbol{\theta}_j$ can be factorized as

$$\pi(\boldsymbol{\theta}_j^{(1)}, \ldots, \boldsymbol{\theta}_j^{(K)}|.) \propto \pi(\boldsymbol{\theta}_j^{(K)}|.)\pi(\boldsymbol{\theta}^{(K-1)}|\boldsymbol{\theta}_j^{(K)}, .) \ldots \pi(\boldsymbol{\theta}_j^{(1)}|\boldsymbol{\theta}_j^{(2)}, \ldots, \boldsymbol{\theta}_j^{(K)}, .), \tag{3.23}$$

in which, for $k = 1, \ldots K$,

$$\pi(\boldsymbol{\theta}^{(k)}|.) \sim N_{Q_k}(\mu_j^{(k)}, \Lambda^{(k)}), \tag{3.24}$$

$$\Lambda^{(k)} = ({\Sigma^{(k+1)}}^{-1} + {\lambda^{(k+1)}}^T {\Sigma^{(k+1)}}^{-1} \lambda^{(k+1)} - {\lambda^{(k)}}^T {\Sigma^{(k)}}^{-1} \Lambda^{(k-1)} {\Sigma^{(k)}}^{-1} \lambda^{(k)})^{-1}, \tag{3.25}$$

and

$$\mu_j^{(k)} = \Lambda^{(k)} ({\Sigma^{(k+1)}}^{-1} \lambda^{(k+1)} \theta^{(k+1)} + \left[\prod_{i=2}^{k} {\lambda^{(i)}}^T {\Sigma^{(i)}}^{-1} \Lambda^{(i-1)}\right] a^T (X_j + b)), \tag{3.26}$$

where for k = K, $\lambda^{(k+1)} = 0$, $\theta^{(k+1)} = 0$ and $\Sigma^{(k+1)} = 1$.

We sample from (3.23) by first sampling $\boldsymbol{\theta}_j^{(K)}$ from $\pi(\boldsymbol{\theta}^{(K)}|.)$, then using these values to sample $\boldsymbol{\theta}_j^{(K-1)}$ from $\pi(\boldsymbol{\theta}_j^{(K-1)}|\boldsymbol{\theta}_j^{(K)}, .)$, and so forth, until $\boldsymbol{\theta}^{(1)}$. Also, for each j, if $z_{ij} = 0$, $(\mathbf{a}_i, b_i) = 0$.

Full Conditional Posterior Distribution for $(\boldsymbol{\theta}, \boldsymbol{\lambda})$

As an attempt to improve the mixing of the chain, we consider the algorithm that samples $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ in the same block. Since directly sampling from the full conditional distribution of ($\boldsymbol{\theta}$, $\boldsymbol{\lambda}$) if not possible, we consider a Metropolis-Hastings step.

The full conditional posterior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is proportional to

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\theta}|.) \propto \pi(\mathbf{X}|\mathbf{Z}, \mathbf{a}, \mathbf{b}, \boldsymbol{\theta}^{(1)}) \pi(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(2)}, \boldsymbol{\lambda}^{(2)}) \pi(\boldsymbol{\theta}^{(2)}) \pi(\boldsymbol{\lambda}) \tag{3.27}$$

The proposal distribution is $q(\boldsymbol{\lambda}, \boldsymbol{\theta}) = q(\boldsymbol{\lambda}) q(\boldsymbol{\theta}|\boldsymbol{\lambda})$, where $q(\boldsymbol{\lambda})$ is a Gaussian random walk and $q(\boldsymbol{\theta}|\boldsymbol{\lambda})$ is the joint full conditional distribution for $\boldsymbol{\theta}$ given in Subsection 3.3.2. The acceptance probability is proportional to the ratio of the following expression evaluated at the proposed values for $(\boldsymbol{\lambda}, \boldsymbol{\theta})$ and the actual values, such that

$$\begin{aligned}
a.p. &= min\Big\{1, \frac{\pi(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*|.) q(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}|.) q(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)}\Big\} \\
&= min\Big\{1, \frac{\pi(\boldsymbol{X}|{\boldsymbol{\theta}^{(1)}}^*, a, b) \pi({\boldsymbol{\theta}^{(1)}}^*|{\boldsymbol{\theta}^{(2)}}^*, \boldsymbol{\lambda}^*) \pi(\boldsymbol{\lambda}^*)}{\pi(\boldsymbol{X}|\boldsymbol{\theta}^{(1)}, a, b) \pi(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(2)}, \boldsymbol{\lambda}) \pi(\boldsymbol{\lambda})} \frac{\pi(\boldsymbol{X}|\boldsymbol{\theta}^{(1)}, a, b) \pi(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(2)}, \boldsymbol{\lambda}) / \kappa(\boldsymbol{\lambda}) \pi(\boldsymbol{\lambda})}{\pi(\boldsymbol{X}|{\boldsymbol{\theta}^{(1)}}^*, a, b) \pi({\boldsymbol{\theta}^{(1)}}^*|{\boldsymbol{\theta}^{(2)}}^*, \boldsymbol{\lambda}^*) / \kappa(\boldsymbol{\lambda}^*) \pi(\boldsymbol{\lambda}^*)}\Big\} \\
&= min\Big\{1, \frac{\pi(\boldsymbol{\lambda}^*)}{\pi(\boldsymbol{\lambda})} \frac{\kappa(\boldsymbol{\lambda}^*)}{\kappa(\boldsymbol{\lambda})}\Big\}, \qquad (3.28)
\end{aligned}$$

where

$$\kappa(\lambda) = \prod_{j=1}^{J}[|\Sigma|^{-1/2}|\Lambda_1|^{1/2}|\Lambda_2|^{1/2}exp\{-\mu_{2j}^T\Lambda_2^{-1}\mu_{2j} - (X_j + b)^T a\Lambda_1 a^T(X_j + b)\}], \tag{3.29}$$

and if $z_{ij} = 0$, $(\mathbf{a}_i, b_i) = 0$.

Full Conditional Distribution for (a, b) in dichotomous items

Define $\boldsymbol{\eta}_i = (\mathbf{a}_i, b_i)$, a vector $1\times Q_1$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{iJ})$, a vector 1×J, and let $\mathbf{C} = [\boldsymbol{\theta}^{\mathbf{(1)}^{\boldsymbol{T}}}, -1]$ be a matrix J$\times(Q_1 + 1)$. Sampling the whole vector (**a**, **b**) jointly is possible due to the conditional independence of the respective full conditional distributions among different items. We have that

$$(\mathbf{a}_i, b_i)|. \sim N(\boldsymbol{\mu}_{\eta_i}^*, \boldsymbol{\Lambda}_\eta^*), \tag{3.30}$$

where

$$\boldsymbol{\Lambda}_\eta^* = (\boldsymbol{\Lambda}_\eta^{-1} + \mathbf{C}^T\mathbf{C})^{-1}, \quad and \tag{3.31}$$

$$\boldsymbol{\mu}_{\eta_i}^* = \boldsymbol{\Lambda}_\eta^*(\boldsymbol{\Lambda}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\mu}_{\eta_i} + \boldsymbol{C}^T\mathbf{X}_i^T). \tag{3.32}$$

Full Conditional Distribution for $\boldsymbol{\lambda}$

Let $\boldsymbol{\lambda}_q^{(k)} = (\lambda_{q1}^{(k)}, \ldots, \lambda_{qQ_k}^{(k)})$, for $q = 1, \ldots, Q_{k-1}$, be a vector containing all coefficients that relates the q^{th} trait at level (k-1) to the level k. The full conditional distribution for λ is proportional to

$$\pi(\boldsymbol{\lambda}_q^{(k)}|.) \propto \left[\prod_j \frac{1}{\tau_q^{(k)}\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{\theta_{qj}^{(k-1)} - \boldsymbol{\lambda}_q^{(k)}\boldsymbol{\theta}_j^{(k)}}{\tau_q^{(k)}}\right)^2\right\}\right]\mathbb{I}(0 < \boldsymbol{\lambda}_q^{(k)} < 1), \tag{3.33}$$

where $\tau_q^{(k)} = \sqrt{1 - {\lambda_q^{(k)}}^2}$.

We are unable to sample directly from this distribution and, therefore, resort to a Metropolis-Hasting step. A Gaussian random walk proposal is adopted and properly tuned to optimize convergence speed (see Roberts et al. (1997)).

Full Conditional Posterior Distribution for c

Due to the prior independence of the vector $\boldsymbol{c}$ with prior $c_i \sim Beta(\alpha, \beta)$, all the c_i's are conditionally independent with full conditional posterior distribution given by

$$\pi(c_i|.) \sim Beta\bigg(\sum_{j=1}^{J} Z_{ij} + \alpha, J - \sum_{j=1}^{J} Z_{ij} + \beta\bigg), \tag{3.34}$$

where α and β are prior hyperparameters.

Full Conditional Distribution for X^G for Graded Response Items

Given parameters $\boldsymbol{\theta}$, $\mathbf{a}^G$ and $\mathbf{b}^G$, it follows from the posterior that $(X_{ij})^G$ are conditionally independent for all i and j, which means that we can sample the vector $(\mathbf{X}^G)$ by sampling each (X_{ij}^G) individually. Therefore, the full conditional posterior distribution is given by

$$(X_{ij}^G|\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}_G, \mathbf{Y}) \sim \begin{cases} N(\mathbf{a}_i^G \boldsymbol{\theta}_j^{(1)}, 1)\mathbb{I}(-\infty < X_{ij}^G < b_{i1}), & \text{if} \quad Y_{ij} = 0, \\ N(\mathbf{a}_i^G \boldsymbol{\theta}_j^{(1)}, 1)\mathbb{I}(b_{i1} < X_{ij}^G < b_{i2}), & \text{if} \quad Y_{ij} = 1, \\ \vdots & \\ N(\mathbf{a}_i^G \boldsymbol{\theta}_j^{(1)}, 1)\mathbb{I}(b_{iM_i-1} < X_{ij}^G < \infty), & \text{if} \quad Y_{ij} = M_i. \end{cases} \tag{3.35}$$

Full Conditional Distribution for $\mathbf{b}^G$ in the Graded Response model

For an item with M_i categories, there are $M_i - 1$ location parameters b_{ic} to be estimated. Due to the typically high number of X_{ij}^G variables, the full conditional distribution of the location parameters will be truncated in very small intervals, leading to very poor mixing of the chain. In order to overcome this issue, we define a collapsed Gibbs sampler by integrating out the X_{ij}^G variables. Also, due to the ordering constraints imposed to parameters b_{ic}, as shown in Equation 3.12, it is not straight forward to generate candidates. Therefore, we consider a reparametrization of the location parameters as follows:

$$\begin{aligned} b_{i1}^{t*} &= b_{i1}^{t} \\ b_{i2}^{t*} &= b_{i2}^{t} - b_{i1}^{t} \\ &\vdots \\ b_{i,M_i-1}^{t*} &= b_{i,M_i-1}^{t} - b_{i,M_i-2}^{t} \end{aligned}$$

in which $b_{i0}^{*} = b_{i0} = -\infty$ and $b_{i,M_i}^{*} = b_{i,M_i} = \infty$

We consider two different update steps for the reparametrized location parameters. The first one (Algorithm 1) only updates $b_{i,1}^{*}$ using a (properly tuned) Gaussian random walk. This implies in a translation of all the original parameters, preserving their differences. The second one (Algorithm 2) updates the $M_i - 2$ difference parameters $b_{i,2:M-1}^{*}$ using a (properly tuned) Gaussian random walk. This means that $b_{i,1}$ is preserved and the respective differences between the $b_{i,}$'s are updated. Each algorithm is as follows.

Algorithm 1 Metropolis-Hastings for $\pi(\mathbf{b}_G|.)$

```
for iteration $t$: do
    for graded item $i$: do
        Propose: $\delta_i^* \sim N(0, \sigma)$
        $\mathbf{b}_{i,1:M_i-1}^{C} = \delta_i + \mathbf{b}_{i,1:M_i-1}^{t}$
        $\alpha(\mathbf{b}_i^C, \mathbf{b}_i) = min\left\{1, \frac{\kappa(\mathbf{b}_{i,1:M_i-1}^{C})}{\kappa(\mathbf{b}_{i,1:M_i-1}^{t})}\right\}$
        $u \sim Uniform(0, 1)$
        if u $< \alpha$ then $\mathbf{b}_{i,1:M_i-1}^{t} \leftarrow \mathbf{b}_{i,1:M_i-1}^{C}$
        else
            $\mathbf{b}_{i,1:M_i-1}^{t} \leftarrow \mathbf{b}_{i,1:M_i-1}^{t-1}$
        end if
    end for
end for
```

for $\kappa(.)$ given by 3.15.

Algorithm 2 Metropolis-Hastings for $\pi(\mathbf{b}_G|.)$

```
for iteration $t$: do
    for graded item $i$: do
        Propose: $b^{C}_{i,2:M_i-1} \sim N_{(M_i-2)}(\mathbf{b}_i^{*t}, \mathbf{\Sigma})\mathbb{I}(b_{2:M_i-1} > 0)$
        $b^{C*}_{i,m} = \sum_{m=1}^{M_i-1} b^{C}_{im}$, for $m = 2, \ldots, M_i - 1$
        $\alpha(\mathbf{b}_i^{C*}, \mathbf{b}_i) = min\left\{1, \frac{\kappa(\mathbf{b}^{C}_{i,2:M_i-1})}{\kappa(\mathbf{b}^{t}_{i,2:M_i-1})}\right\}$
        $u \sim Uniform(0,1)$
        if  u $< \alpha$ then $\mathbf{b}^{t}_{i,1:M_i} \leftarrow \mathbf{b}^{C*}_{i,2:M_i}$
        else
            $\mathbf{b}^{t}_{i,1:M_i} \leftarrow \mathbf{b}^{t-1}_{2,1:M_i}$
        end if
    end for
end for
```

Full Conditional Distribution for $\mathbf{a}^G$ for Graded Response Items

For graded items, sampling the whole vector $\mathbf{a}^G = (\mathbf{a}^G_1, \ldots, \mathbf{a}^G_I)$ jointly is possible due to the conditional independence of the respective full conditional distributions among different items. We have that

$$\mathbf{a}_i^G|. \sim N(\boldsymbol{\mu}^*_{a_i}, \mathbf{\Lambda}^*_a), \tag{3.36}$$

which has dimension equal to the latent traits being measured by item i. Also,

$$\mathbf{\Lambda}^*_a = (\mathbf{\Lambda}_a{}^{-1} + \boldsymbol{\theta}^{(1)T}\boldsymbol{\theta}^{(1))})^{-1}, \quad and \tag{3.37}$$

$$\boldsymbol{\mu}^*_{a_i} = \mathbf{\Lambda}^*_a(\mathbf{\Lambda}_a{}^{-1}\boldsymbol{\mu}_a + \boldsymbol{\theta}^{(1)}\mathbf{X}_i^T). \tag{3.38}$$

Chapter 4

Simulation and Data Analysis

4.1 Comparison of number of students

We compare the efficiency of the proposed methodology to estimate the parameters under different specifications of subject's sizes. The data was originated by a two level hierarchical model. In the first level there were four latent traits and one in the second level, such that the true coefficient values were $\boldsymbol{\lambda} = (\lambda_1 = 0.95,\ \lambda_2 = 0.90,\ \lambda_3 = 0.85, \lambda_4 = 0.80)$. Each trait was measured by 45 unidimensional dichotomous items and the pseudo-guessing parameter was set to zero in this study. The model used to estimate the parameters had the same structure of the model that generated the data set. In Subsection 4.2.1, we present a simulation study in which the pseudo-parameter is considered as well multidimensional and polytomous items. Parameters $\boldsymbol{\theta}_j$ were jointly sample, as presented in Subsection 3.3.2, separately from $\boldsymbol{\lambda}$.

Table 4.1 presents the posterior statistics for $\boldsymbol{\lambda}$ for the three sample sizes (500, 2k and 5k), in which StandDev denotes standard deviation and $CI_{0.025}$ and $CI_{0.975}$ denote percentile 2.5 and 97.5, respectively. Notice that the $\boldsymbol{\lambda}$ parameters were satisfactorily recovered for the three sample size. Also, as expected, the posterior standard deviation is impacted by sample size, decreasing as the sample size increase. Moreover, the higher the λ value, the smaller its posterior standard deviation. Figure A.3, in Subsection A.0.2, shows the trace plots for λ, indicating fast converge. Figure A.4 (see Appendix) shows the posterior mean versus true values for the first and second level θ, and Figure A.6 (see Appendix) shows the posterior mean versus true values for the item parameters. Also, notice that, under a highly complex structure, even for a small sample size as 500, all model parameters were fairly well recovered. Figure A.5 shows an example of the trace plot for a first and second level latent trait related to a random subject and its respectively ACF plots, related to Simulation 1, indicating fast convergence.

Simulation	Sample Size	λ	Mean	StandDev	$CI_{0.025}$	$CI_{0.975}$
1	500	0.95	0.957	0.010	0.937	0.974
		0.90	0.905	0.013	0.879	0.928
		0.85	0.852	0.017	0.817	0.883
		0.80	0.836	0.018	0.797	0.869
2	2.000	0.95	0.946	0.005	0.937	0.956
		0.90	0.904	0.006	0.892	0.917
		0.85	0.865	0.008	0.849	0.879
		0.80	0.791	0.010	0.770	0.810
3	5.000	0.95	0.949	0.003	0.942	0.955
		0.90	0.900	0.004	0.891	0.908
		0.85	0.853	0.005	0.842	0.862
		0.80	0.794	0.006	0.781	0.807

Table 4.1: Posterior values for $\boldsymbol{\lambda}$ for different subject sample sizes.

4.2 Other Studies

We now study the impact of not considering a hierarchical structure when there is one and vice versa. In Simulation 4, data was originated from hierarchical model, such that $\boldsymbol{\lambda} = (\lambda_1 = 0.95, \lambda_1 = 0.90, \lambda_1 = 0.85, \lambda_1 = 0.80)$, but the fitted model ignored the hierarchical structure imposing that $\lambda = (0, 0, 0, 0)$, that is, four unidimensional unrelated tests. In Simulation 5, the dataset was generated considering uncorrelated subtests, that is, $\lambda = (0, 0, 0, 0)$, but the estimated model allowed a hierarchical structure. In Simulation 6, dataset was originated from an unidimensional model, that is, all $\lambda = 1$, but a hierarchical structure was allowed to fit the data, assuming four first level and one second level latent traits. In Simulation 7, the dataset was generated with a hierarchical structure, such that $\boldsymbol{\lambda} = (\lambda_1 = 0.95, \lambda_1 = 0.90, \lambda_1 = 0.85, \lambda_1 = 0.80)$, and the same structure was considered to estimate parameters, under four first level and one second level traits. For Simulation 4-7, pseudo-guessing parameters were set to zero, i.e. a 2-PNO (parameter normal ogive), items were dichotomous (45 for each first level trait), and 5K subjects. Also, item values were fixed at their real value to isolate noise effects.

Notice that, when all $\lambda = 0$ the whole test is composed by uncorrelated subtests and when all $\lambda = 1$ the test is unidimensional. Therefore, both models are particular cases of the hierarchical model. Moreover, one can use the proposed model to test unidimensionality or uncorrelated traits, as we show in these studies. This means that, the proposed model can be used even when the researcher is not sure about the existence of a hierarchical structure.

Table 4.3 shows the Root Mean Squared Error (RMSE), $\sqrt{\sum_{j=1}^{J}(\hat{\theta}_{qj}^{(k)} - \theta_{qj}^{(k)})^2/J}$, for latent traits, under these studies. The lowest RMSE is for Simulation 6, where dataset was unidimensional, i.e., $\lambda = 1$ for all subtests. This is expected as all 180 are basically measuring only one trait. For these studies, the RMSE for Simulation 4 and 5 were slightly higher than the RMSE for Simulation 7. Also, RMSE for second level $\theta_1^{(2)}$ (Simulation 4) was obtained by averaging the four first level latent traits posterior mean. One of the most import result is that $\theta^{(2)}$ are better estimated under the hierarchical approach (lower RMSE). This is also verify in Figure 4.1, which presents the density of real and estimated values (posterior mean) of $\theta_1^{(2)}$ for both Simulation 4 and 7. As expected, the model that considers the hierarchical structure recoverer better $\theta^{(2)}$ than a simple average. This is expected mostly when the values of $\boldsymbol{\lambda}$ greatly differ, meaning that lower level traits are not equally related to the higher latent traits. When the subtests form an unidimensional test, indicating almost perfect correlation among first level traits, all λ's are estimated near one. When there are uncorrelated subtests, all λ are estimated near zero (Table 4.2).

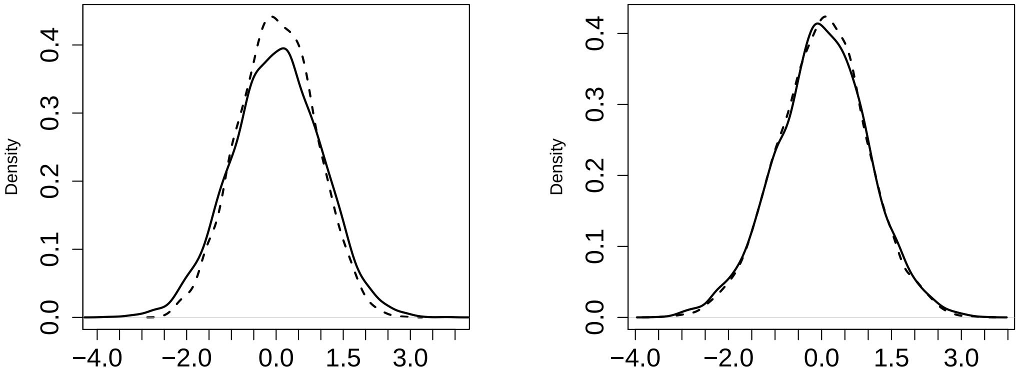
Although the model's structure is very complex, these studies show item and subject's parameters are satisfactorily recovered. As noticed before, other models, included the unidimensional one, are sub models of the hierarchical model. Not only we present the model and its estimation processes, as we construct in a way to preserve flexibility and interpretation. Moreover, notice that, in education tests, for instance, it is crucial to estimate latent traits with maximum precision, as these results are used to classify subjects.

Simulation	Real	Mean	StandDev	$CI_{0.025}$	$CI_{0.975}$	Range
5	0	-0.023	0.123	-0.280	0.212	0.492
	0	0.023	0.146	-0.248	0.289	0.538
	0	-0.022	0.120	-0.338	0.168	0.506
	0	-0.001	0.121	-0.255	0.212	0.466
6	1.00	0.9992	2e-04	0.9986	0.9995	0.0010
	1.00	0.9999	0e+00	0.9998	0.9999	0.0001
	1.00	0.9992	5e-04	0.9982	0.9997	0.0015
	1.00	0.9985	4e-04	0.9976	0.9991	0.0016

Table 4.2: Posterior statistics for $\boldsymbol{\lambda}$ (Simulations 5 and 6).

Simulation	$\theta_1^{(1)}$	$\theta_2^{(1)}$	$\theta_3^{(1)}$	$\theta_4^{(1)}$	$\theta_1^{(2)}$
4	0.255	0.238	0.251	0.248	0.313
5	0.250	0.248	0.240	0.249	-
6	0.127	0.126	0.126	0.127	0.1260
7	0.221	0.225	0.221	0.241	0.271

Table 4.3: Root Mean Squared Error.



(a)

(b)

Figure 4.1: Density plot of real (dotted line) and estimated (solid) values for $\theta^{(2)}$: Simulation 4 (left) and Simulation 7 (right)

4.2.1 Simulation 8

We also ran an example simulating an exam with both dichotomous and polytomous items, with some multidimensional and also considering the pseudo-guessing parameter for all items.

The hierarchical structure was composed by five first level traits and one second level ability, such that $\boldsymbol{\lambda} = (\lambda_1 = 0.95,\ \lambda_2 = 0.90,\ \lambda_3 = 0.85,\ \lambda_4 = 0.80,\ \lambda_5 = 0.80)$. The dataset consisted of 10k subjects, 180 dichotomous items, among some multidimensional, and five polytomous items. Five items of trait $\theta_2^{(1)}$ were related to $\theta_1^{(1)}$, five items of trait $\theta_3^{(1)}$ were related to $\theta_1^{(1)}$, six items of trait $\theta_4^{(1)}$ were related to $\theta_1^{(1)}$, and one item of $\theta_5^{(1)}$ (polytomous item) was related to $\theta_1^{(1)}$. This simulation study structure is done in order do replicate the real data analysis, which will be latter explored (Section 4.3). The pseudo-guessing of all items were randomly selected from a Uniform(0,0.2).

All $\boldsymbol{\lambda}$ were very well recovered (Table 4.4) and Figure A.8 shows the fast converge of these parameters. Also, as said before, the $\boldsymbol{\lambda}$ chains are uncorrelated (Table 4.5). Figure A.7 (Appendix) shows the posterior mean versus true values for all first and second traits, indicating that these parameters were well recovered, and Figure A.9 presents the posterior mean versus true values for all item parameters, also satisfactorily recovered, even under model high complexity. The pseudo parameters were middling recovered for some items, but this behavior is expected for these parameters.

λ	Mean	StandDev	$CI_{0.025}$	$CI_{0.975}$	CI Range
0.95	0.951	0.002	0.947	0.956	0.009
0.90	0.897	0.003	0.890	0.902	0.012
0.85	0.850	0.004	0.842	0.857	0.014
0.80	0.800	0.004	0.791	0.808	0.018
0.75	0.742	0.006	0.731	0.754	0.023

Table 4.4: Posterior statistics for $\boldsymbol{\lambda}$ (Simulation 8).

	λ_1	λ_2	λ_3	λ_4
λ_2	-0.05			
λ_3	0.00	0.14		
λ_4	-0.01	0.17	0.17	
λ_5	0.03	0.15	0.19	0.20

Table 4.5: Posterior correlation among $\boldsymbol{\lambda}$'s chains (Simulation 8).

4.3 Real data analysis - Enem

We apply the proposed methodology to a dataset from the High School National Exam (Enem) from Brazil, related to 2017. The exam is annually applied to high school students and is organized by the National Institute of Educational Studies and Researches Anísio Teixeira (INEP) from the Ministry of Education (MEC). It aims to assess students who are concluding or have already concluded high school in the previous years. The exam is used in many universities admission processes country wide. Enem is composed of five sub exams: Humanities (H), Natural science (NS), Languages (Lang) and Maths (MT), each with 45 dichotomous unidimensional items, and an Essay. The five sub-exams are corrected independently, using the 3-parameter logistic IRT model. The Essay is corrected in the classical way, in which two referees independently grade the Essay over five matters, each one with grade 0, 20, 40, 60, 80, 100, 120, 140, 160, 180 or 200. The five grades are summed for each referee and the final score is the average both scores. This means that the Essay is not in the same scale of the other four tests, i.e., it is not under a normal scale with mean zero and variance one. The student's final score is the average of the five scores. This approach could be improved since it averages scores with different scales. Figure 4.2 shows Enem's current structure, in which the first 45 items are related to Humanities, items 46 to 90 are related to Natural Sciences, items 91 to 135 are related to Language, items 136 to 180 are related to Mathematics, and, finally, items 181 to 185 are related to the Essay. We consider data from the Languages exam consisting of a random sample of 52,210 students

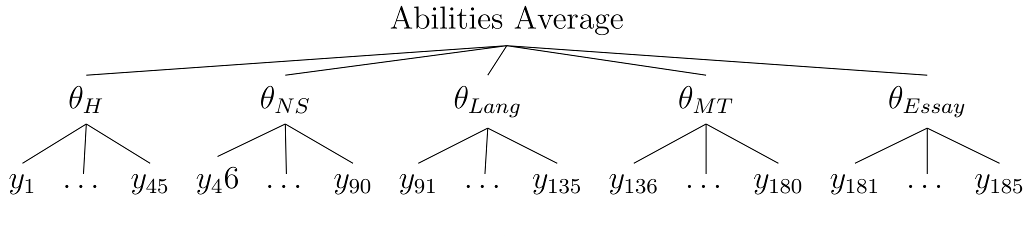


Figure 4.2: Enem's current structure.

It is reasonable to expect a multidimensional behavior in some items, since, for instance, some Mathematics items require more interpretation skills or some Natural Science items require some Mathematics skills. Moreover, it is expected a positive correlation among a same subjetc's abilities. In fact, this is evident when computing the correlation matrix for scores obtained from Enem's grading system, for a random sample of 44,000 students (Table 4.6).

	Humanities	Natural Sciences	Language	Mathematics
Natural Sciences	0.67			
Language	0.62	0.75		
Mathematics	0.63	0.64	0.60	
Essay	0.48	0.53	0.54	0.47

Table 4.6: Correlation of the latent traits in the Enem 2017 exam.

The results in Table 4.6 show a significant correlation among Enem's estimated abilities. Notice that the scores were obtained from an analysis where latent traits were estimated separately for each sub-test and, therefore, without considering any relation between specific skills. A more coherent statistical analysis requires that the relationship among the specific abilities and the definition of the overall ability to be conceived in the adjusted model. In the same way, the possible multidimensional structure of some items must also be considered. Moreover, no source of uncertainty should be ignored in the estimation of traits and their relationships. Figure 4.3 presents the proposed multidimensional hierarchical structure for Enem dataset analysis.

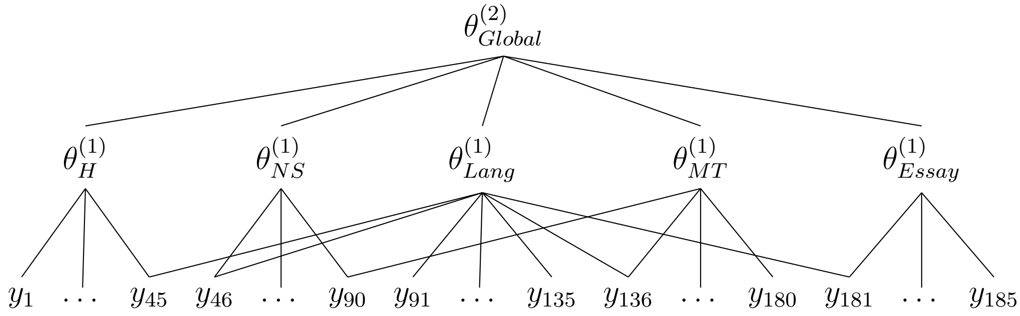


Figure 4.3: Proposed hierarchical structure applied to Enem.

We consider a random sample of 44,000 students from all over the country. There are four types of exams, differing only by item's ordering and we used items order according to the Blue version of the test. For the Languages exam, students can choose between English or Spanish but the sample was chosen only among students who chose the English exam. Also, we only selected students who completed the five exams and that were not disqualified in any of the exams. We subjectively chose some multidimensional items based on textual and mathematics complexity (see Table 4.9). After a pilot analysis, some items were estimated as having negative discrimination and were excluded from the analysis (10, 23, 25, 30, 36, 65, 123, 143 and 157). This happens when the characteristic curve is not monotonically increasing. Figure A.11 (Appendix) shows the accuracy rate for estimated trait levels related to Item 36 (Humanities) and 143 (Mathematics) indicating non monotonically increasing characteristic curves.

Table 4.7 presents the posterior statistics for factor weights $\boldsymbol{\lambda}$, showing the highest value was related to the Humanities ($\lambda_H = 0.98$), followed by Natural Sciences (0.964) and the lowest was Essay (0.686). These results show that, as expected, all the subtests are highly correlated. Notice that, if the exam as a whole was unidimensional, all weights $\boldsymbol{\lambda}$ would be close to one. Moreover, if the tests were each unidimensional, but uncorrelated, as it is currently assumed to compute scores, all weights $\boldsymbol{\lambda}$ would be close to zero. Therefore, it a simple average over the five scores is not the most robust way to estimate the general ability. Figure A.10 (Appendix) shows $\boldsymbol{\lambda}$'s trace plots (the initial chain values for all $\boldsymbol{\lambda}$ were 0.9).

Test	Mean	StandDev	$CI_{0.025}$	$CI_{0.975}$	CI Range
Humanities	0.980	0.001	0.979	0.982	0.004
Natural Sciences	0.964	0.001	0.962	0.967	0.005
Language	0.936	0.001	0.934	0.939	0.005
Mathematics	0.892	0.002	0.886	0.896	0.010
Essay	0.686	0.003	0.680	0.692	0.012

Table 4.7: Posterior values for $\boldsymbol{\lambda}$ - Enem.

Table 4.8 presents the posterior mean for the location parameters for the Essay. Essay's first item was assumed to be multidimensional with the Language trait. Again, it is worth noticing that Enem does not score the Essay via IRT. Table A.1 (Appendix) presents the posterior mean for discrimination, location and guessing parameters for all items and Table 4.9 shows the posterior mean for item parameters concerning only the multidimensional items. In fact, most of those items presented low multidimensionality effect. Item 84, measuring the Natural Sciences trait, however, showed a considerable multidimensionality effect with the Language trait. Moreover, Item 86, also related to Natural Sciences trait, was more correlated to the Math trait ($a_{MT} = 1.07$) than to Natural Sciences trait ($a_{NS} = 0.65$). Although we only considered a few multidimensional items, it should be noticed that these features have impact on subjects' scores and should not be ignored. Even more, the possibility of having multidimensional items should be indicated *a priori* by the specialists who develop the items. Although Enem' specialists formulate items with the intention of them being unidimensional, it is not always easy to control this feature. Figure A.12 (Appendix) shows trace plots for some item parameters, indicating fast convergence.

Item	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
C1	-8.66	-6.08	-4.77	-3.87	-2.55	-1.59	0.26	1.26	3.63	4.83
C2	-6.80	-6.52	-3.63	-3.18	-2.47	-1.60	1.27	1.72	2.41	3.14
C3	-9.07	-7.42	-5.18	-4.47	-2.38	-1.02	1.41	2.68	4.38	5.60
C4	-7.37	-6.26	-5.04	-3.93	-2.08	-0.88	1.07	2.03	3.36	4.32
C5	-2.33	-1.89	-1.25	-0.79	-0.13	0.39	1.15	1.78	2.55	3.28

Table 4.8: Posterior values for essay's location parameters - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
18	0.91		0.31			1.44	0.24
43	0.63		0.40			1.52	0.21
84		1.00		0.73		3.35	0.08
86		**0.65**		**1.07**		4.66	0.09
156			0.24	2.23		5.28	0.21
181			0.20		1.31		

Table 4.9: Posterior mean for item parameters for the multidimensional item.

Table 4.10 presents the correlation matrix for the posterior latent traits means. Notice that the correlations are higher than the correlations in Table 4.6, in which the hierarchical structure was not considered.

	Humanities	Natural Sciences	Language	Mathematics
Natural Sciences	0.99			
Language	0.97	0.96		
Mathematics	0.95	0.95	0.92	
Essay	0.73	0.73	0.70	0.69

Table 4.10: Enem's correlations among posterior latent trait means.

Chapter 5

Conclusion

Although IRT models with hierarchical structure for latent variables have already been proposed in the literature, all the existing approaches have serious concerns and limitations in terms of both modeling and inference. Motivated by the great practical importance of the topic, the aim of this thesis was to propose a robust, functional and practical methodology to deal with multidimensional IRT models with hierarchical structure for the latent traits. In fact, the multidimensional feature and hierarchical structure had never been consider simultaneously in the literature before.

The general IRT model proposed in the thesis is specified in a way that all the latent traits are in the same (known) scale, so model identification is guaranteed and interpretation is not lost. Furthermore, we showed how different restrictions imposed by past authors are actually the same model under different parametrizations. We also discussed what is the best parametrization in terms of model identification and interpretation, without losing model flexibility.

We studied computational aspects and carefully devised an MCMC algorithm in order to efficiently obtain an approximated sample from the posterior distribution. Different blocking schemes were considered and the best performance was achieved when the latent traits and factor loading are sampled separately.

Simulation studies were performed to investigate the efficiency of the proposed MCMC algorithm under different sample sizes. Moreover, we studied the impact of not considering the hierarchical structure when this exists and vice-versa. We showed that the model is capable of detecting whether the test is unidimensional, multidimensional with uncorrelated latent traits or multidimensional with correlated latent traits.

Finally, we applied the proposed methodology to a real data set from the High School National Exam (Enem) from Brazil, which is the most important education exam in the country. Differently from the currently methodology, we considered a hierarchical structure for the latent traits, allowing us to estimated a more robust and informative general latent trait. We also considered some multidimensional items for which such feature was confirmed. Finally, the Essay's score was estimated via IRT together with the other sub exams, which also contributed to the quality of the general latent trait.

We believe that this work offers substantial contribution to the field of Item Response Theory and educational assessment. We also think that it may inspire and be the basis for further developments in the area.

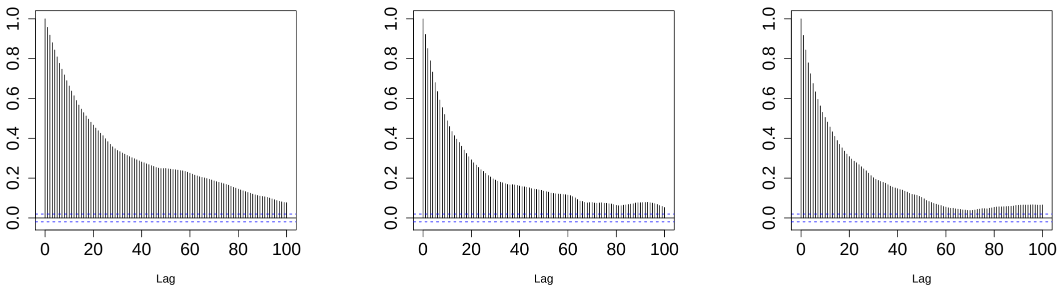
Appendix A

Appendix

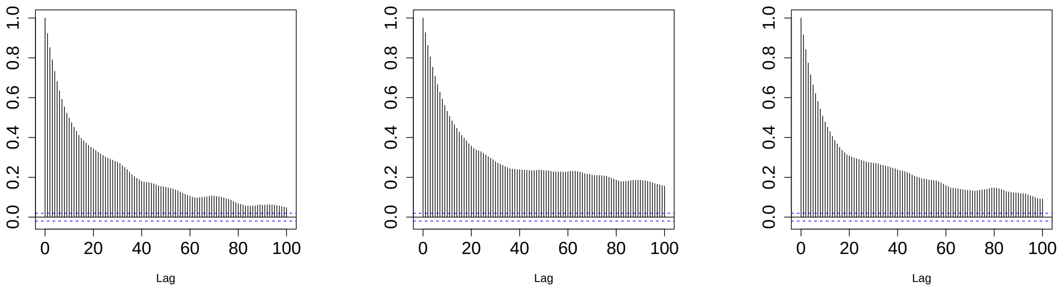
A.0.1 Comparison among blocking methods

We compare the blocking schemes, presented in Subsection 3.3.2, for different subject sizes to understand the impact of blocking scheme on the autocorrelation of $\boldsymbol{\lambda}$'s chains. The models that originated the datasets were hierarchical, such that $\boldsymbol{\lambda} = (\lambda_1 = 0.95, \lambda_2 = 0.90, \lambda_3 = 0.85, \lambda_4 = 0.80)$. The datasets consisted of 2K, 5K and 20K subjects and 180 unidimensional items (45 for each first level latent traits). The estimated model also supposed the same model structure.

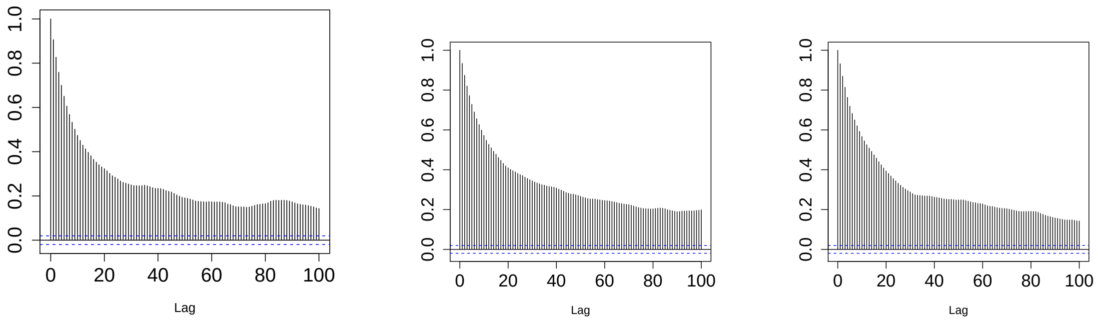
Figure A.1 shows the ACF plots related to the blocking scheme when $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are jointly sampled and Figure (A.2) shows ACF plots related to the blocking scheme when $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are sampled separately. The latter method seems to be generally better, and for this reason, in the simulation studies and in the real data analysis, we sampled all $\boldsymbol{\theta}$ jointly separately from $\boldsymbol{\lambda}$.



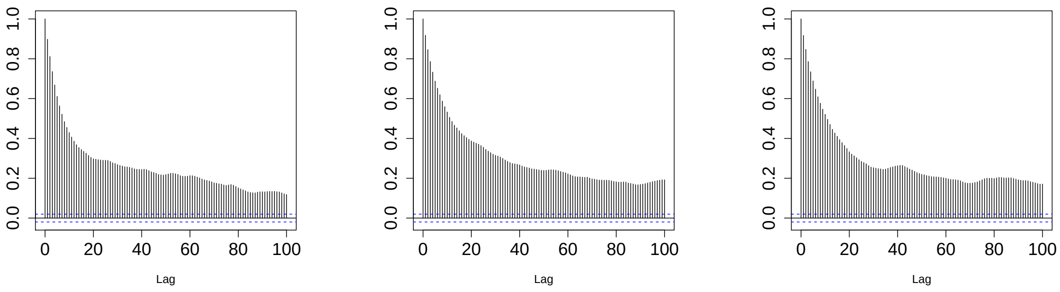
(a) $\lambda_1 = 0.95$ 2k subjects (b) $\lambda_1 = 0.95$ 5k subjects (c) $\lambda_1 = 0.95$ 20k subjects



(d) $\lambda_2 = 0.90$ 2k subjects (e) $\lambda_2 = 0.90$ 5k subjects (f) $\lambda_2 = 0.90$ 20k subjects

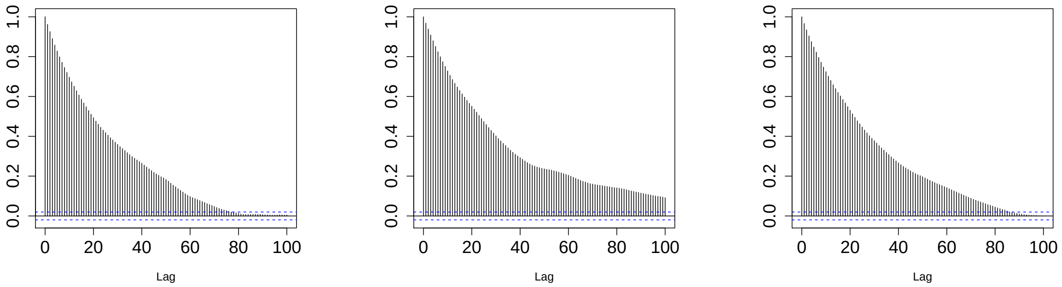


(g) $\lambda_3 = 0.85$ 2k subjects (h) $\lambda_3 = 0.85$ 5k subjects (i) $\lambda_3 = 0.85$ 20k subjects

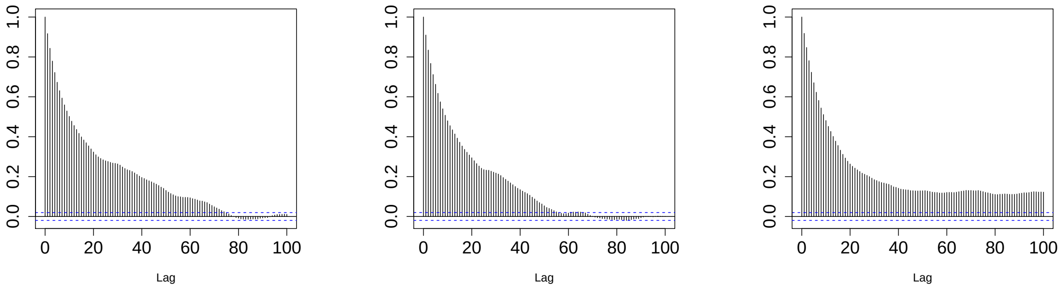


(j) $\lambda_4 = 0.80$ 2k subjects (k) $\lambda_4 = 0.80$ 5k subjects (l) $\lambda_1 = 0.80$ 20k subjects

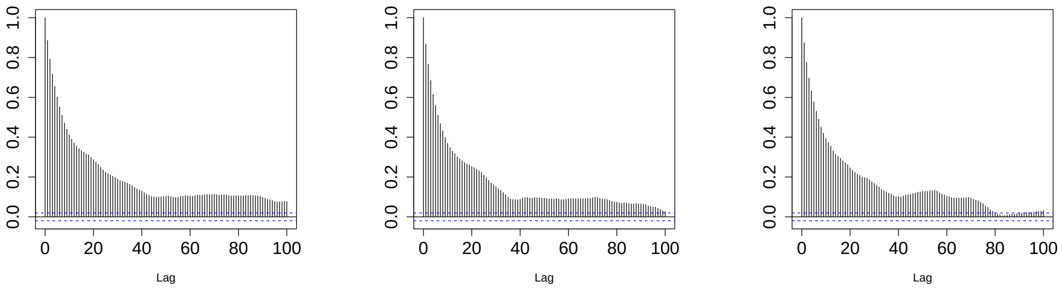
Figure A.1: $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ jointly.



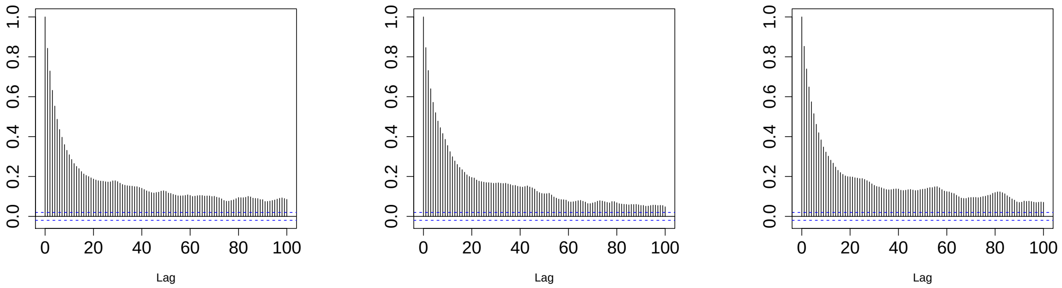
(a) $\lambda_1 = 0.95$ 2k subjects (b) $\lambda_1 = 0.95$ 5k subjects (c) $\lambda_1 = 0.95$ 20k subjects



(d) $\lambda_2 = 0.90$ 2k subjects (e) $\lambda_2 = 0.90$ 5k subjects (f) $\lambda_2 = 0.90$ 20k subjects



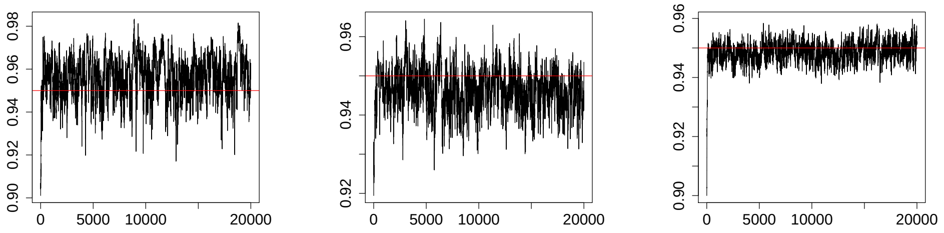
(g) $\lambda_3 = 0.85$ 2k subjects (h) $\lambda_3 = 0.85$ 5k subjects (i) $\lambda_3 = 0.85$ 20k subjects



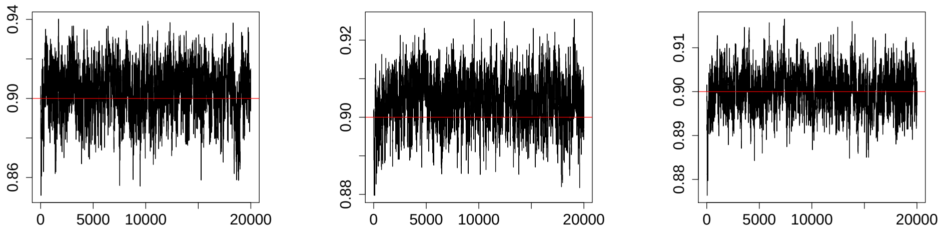
(j) $\lambda_4 = 0.80$ 2k subjects (k) $\lambda_4 = 0.80$ 5k subjects (l) $\lambda_4 = 0.80$ 20k subjects

Figure A.2: ($\boldsymbol{\theta}$) and ($\boldsymbol{\lambda}$) separately.

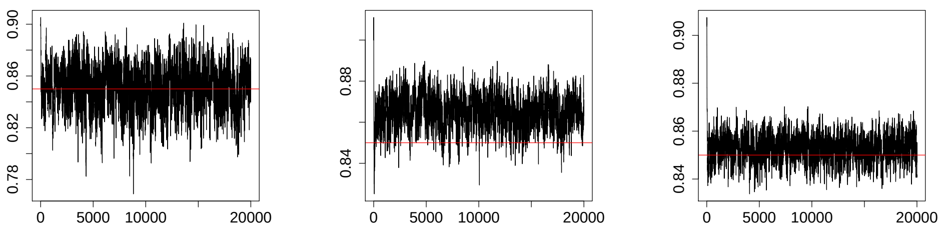
A.0.2 MCMC plots



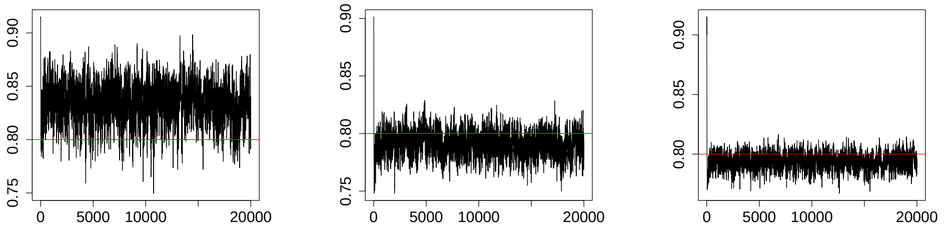
(a) $\lambda_1 = 0.95$ - 500 (b) $\lambda_1 = 0.95$ - 2k (c) $\lambda_1 = 0.95$ - 5k



(d) $\lambda_2 = 0.90$ - 500 (e) $\lambda_2 = 0.90$ - 2k (f) $\lambda_2 = 0.90$ - 5k

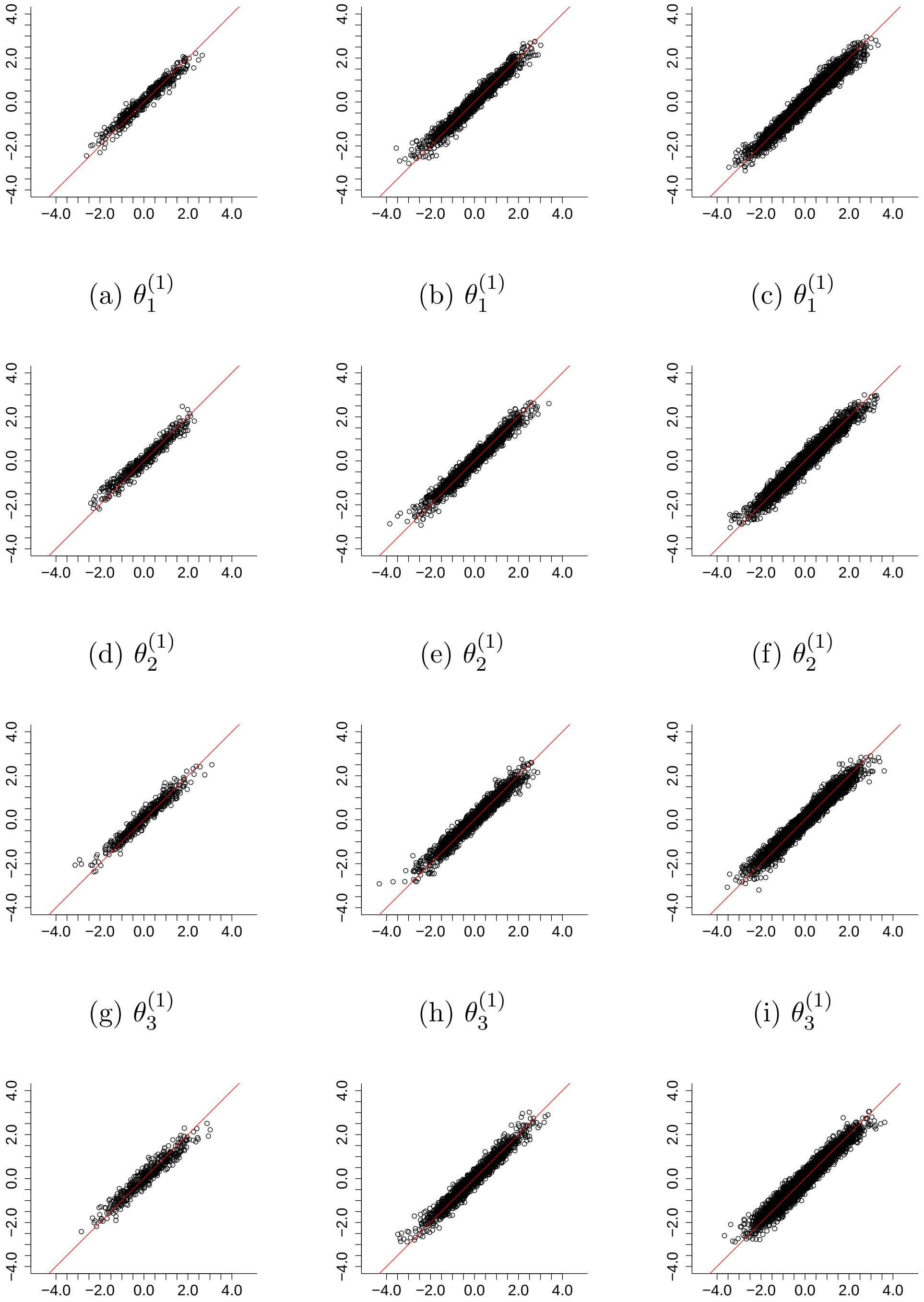


(g) $\lambda_3 = 0.85$ - 500 (h) $\lambda_3 = 0.85$ - 2k (i) $\lambda_3 = 0.85$ - 5k



(j) $\lambda_4 = 0.80$ - 500 (k) $\lambda_4 = 0.80$ - 2k (l) $\lambda_4 = 0.80$ - 5k

Figure A.3: Trace plots for $\boldsymbol{\lambda}$ for different subject sample sizes (Simulations 1-3): line is the parameter true value.



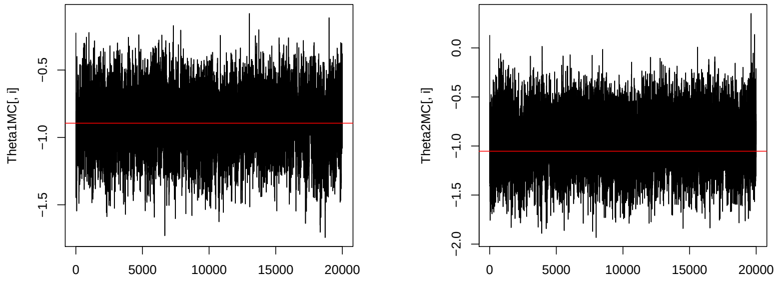
(a) $\theta_1^{(1)}$ (b) $\theta_1^{(1)}$ (c) $\theta_1^{(1)}$

(d) $\theta_2^{(1)}$ (e) $\theta_2^{(1)}$ (f) $\theta_2^{(1)}$

(g) $\theta_3^{(1)}$ (h) $\theta_3^{(1)}$ (i) $\theta_3^{(1)}$

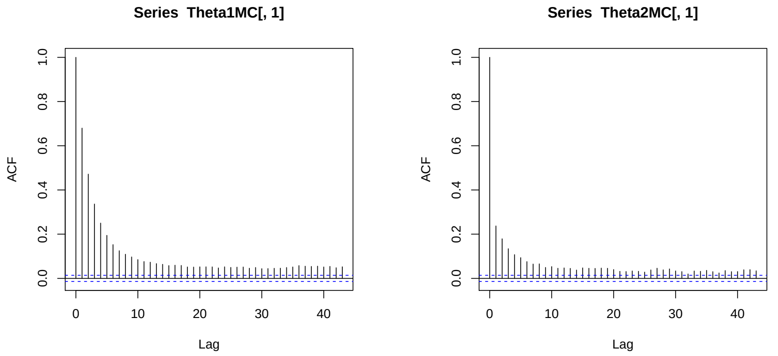
(j) $\theta_4^{(1)}$ (k) $\theta_4^{(1)}$ (l) $\theta_4^{(1)}$

Figure A.4: Posterior mean (y-axis) versus true (x-axis) values for 500 (left column), 2k (middle column) and 5k (right column) subjects (Simulations 1-3).



(a) Trace plot of subject 1 for $\theta_1^{(1)}$.

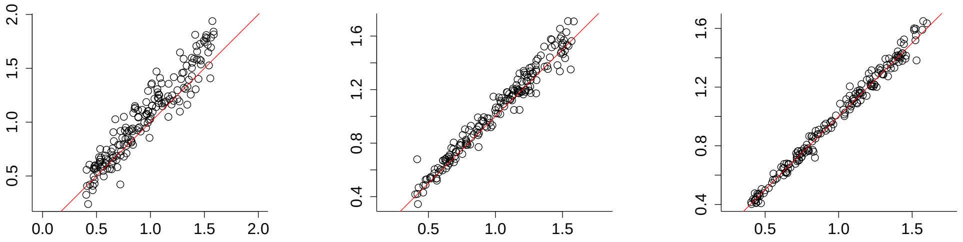
(b) Trace plot of subject 1 for $\theta_1^{(2)}$.



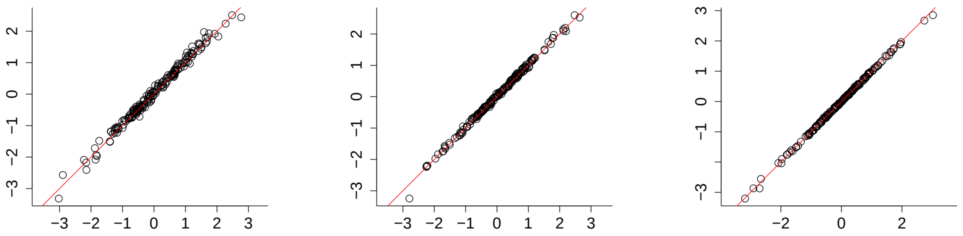
(c) ACF of subject 1 for $\theta_1^{(1)}$.

(d) ACF of subject 1 for $\theta_1^{(2)}$

Figure A.5: Trace plots (horizontal line is the parameter true value) and ACF for subject 1 (Simulation 1).

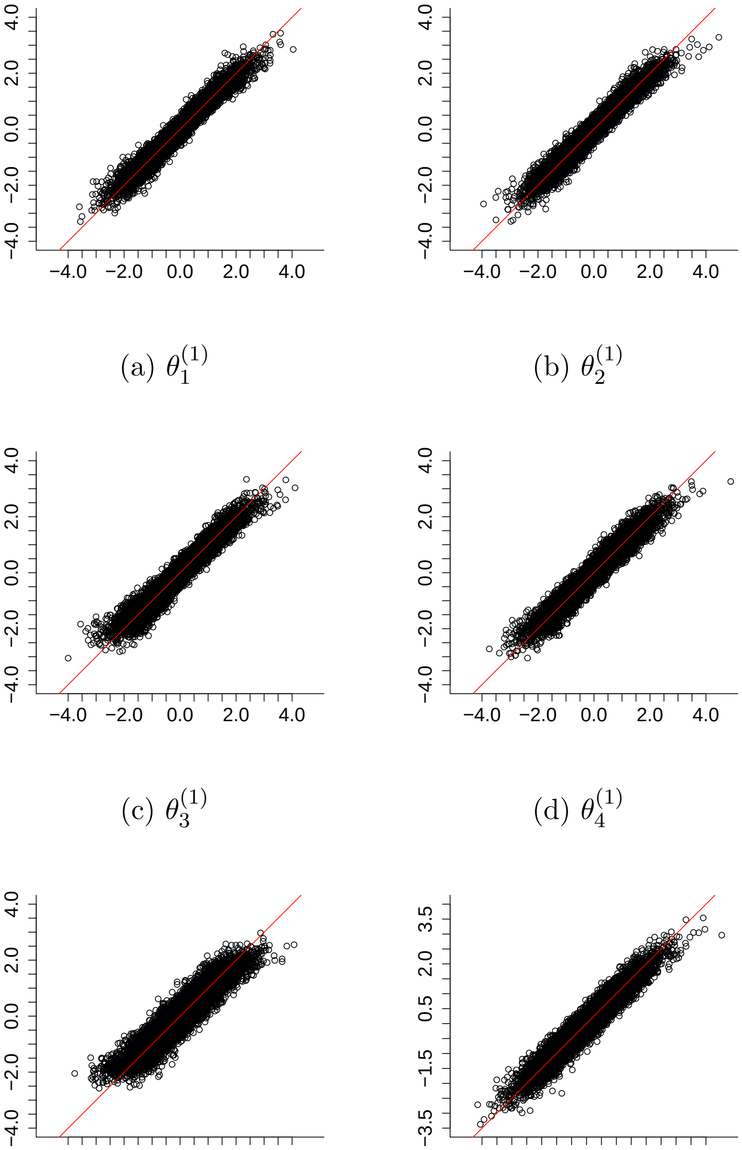


(a) Discrimination - 500. (b) Discrimination - 2k. (c) Discrimination - 5k.



(d) Location - 500. (e) Location - 2k. (f) Location - 5k.

Figure A.6: Posterior mean (y-axis) versus true value (x-axis) for three subject sample sizes (Simulations 1-3).

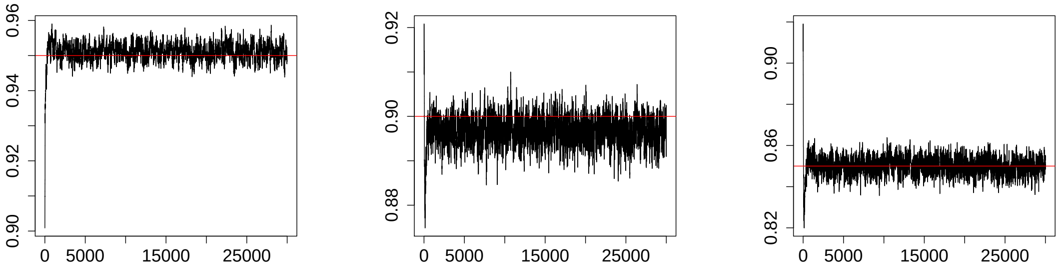


(a) $\theta_1^{(1)}$ (b) $\theta_2^{(1)}$

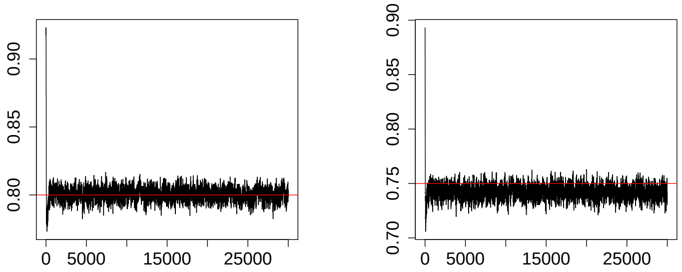
(c) $\theta_3^{(1)}$ (d) $\theta_4^{(1)}$

(e) $\theta_5^{(1)}$ (f) $\theta_1^{(2)}$

Figure A.7: True value (x-axis) versus posterior mean (y-axis) for latent traits (Simulation 8).

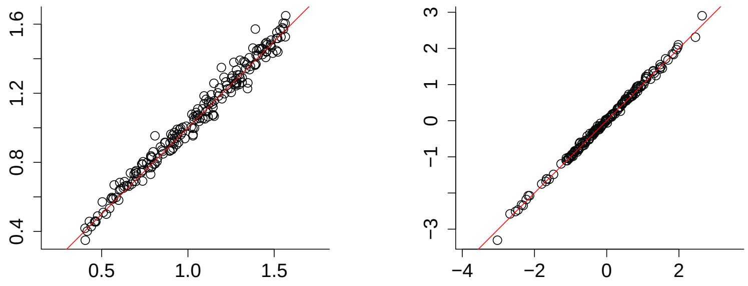


(a) λ_1 (b) λ_2 (c) λ_3



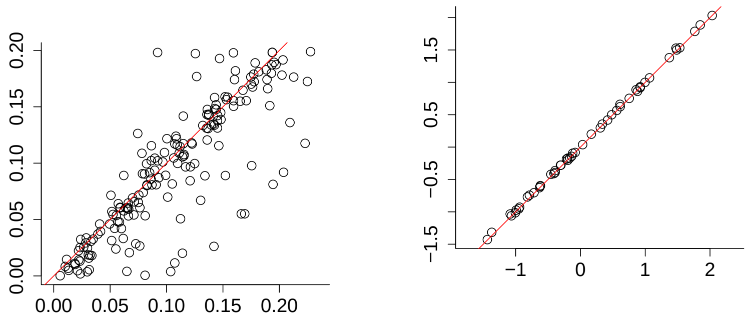
(d) λ_4 (e) λ_5

Figure A.8: Trace plots for $\boldsymbol{\lambda}$ from Simulation 8: horizontal line is the parameter true value.



(a) Discrimination of dichotomous and polytomous items.

(b) Location of dichotomous items.

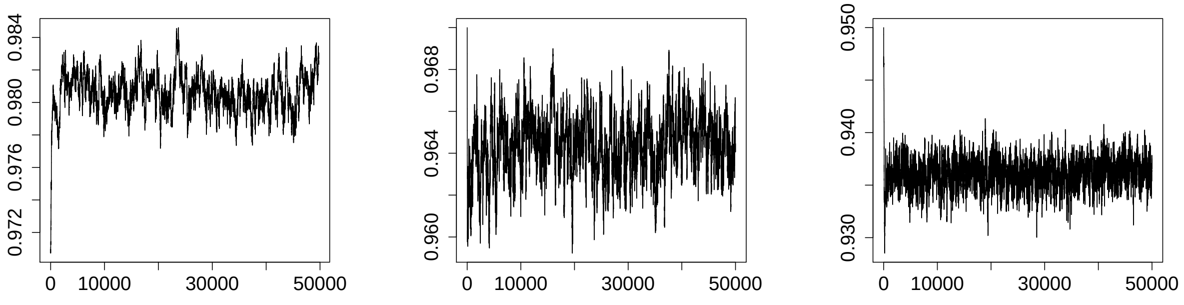


(c) Pseudo-guessing.

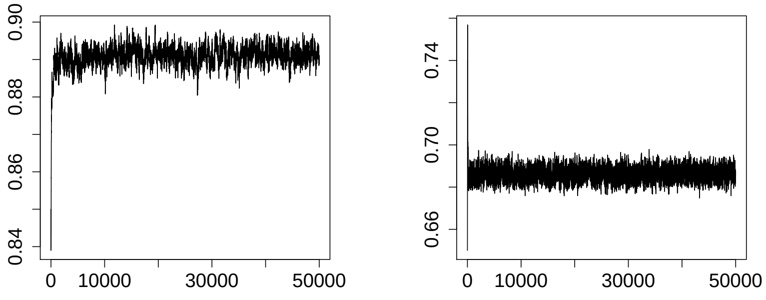
(d) Location for polytomous items.

Figure A.9: True value (x-axis) versus posterior mean (y-axis) for Simulation 8.

A.0.3 Further results from real data analyses

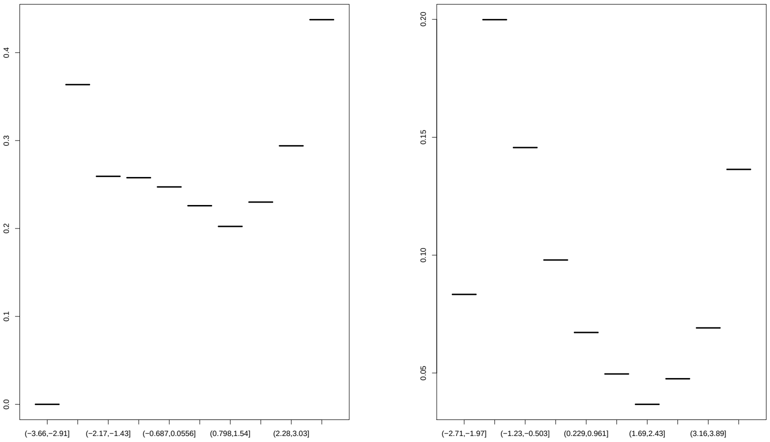


(a) λ_{CH} (b) λ_{CN} (c) λ_{LC}



(d) λ_{MT} (e) λ_{Essay}

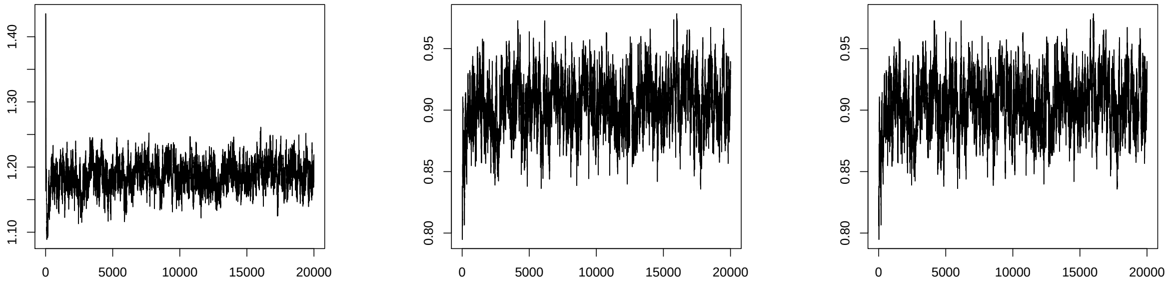
Figure A.10: Trace plots for $\boldsymbol{\lambda}$: initial chain values 0.90. (Figure (a): showing chain from iteration 200) - Enem.



(a) Item 36.

(b) Item 143.

Figure A.11: Accuracy rate for different latent trait levels.



(a) Discrimination.

(b) Location.

(c) Pseudo-guessing.

Figure A.12: Trace plots for discrimination, location and pseudo-guessing parameters of Item 1 - Enem.

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
1	1.19					0.91	0.10
2	0.91					-0.16	0.14
3	0.32					1.59	0.18
4	0.51					-0.07	0.04
5	0.76					0.31	0.19
6	0.58					-0.34	0.02
7	1.34					1.04	0.17
8	1.30					3.79	0.20
9	1.15					0.86	0.11
10							
11	0.48					0.24	0.04
12	1.01					1.94	0.17
13	1.33					-0.21	0.2
14	0.30					0.56	0.12
15	1.00					0.68	0.17
16	0.36					-0.06	0.06
17	1.48					1.30	0.19
18	0.91		0.31			1.44	0.24
19	0.19					0.49	0.09
20	0.56					0.35	0.05
21	0.64					1.55	0.18
22	0.56					0.39	0.16
23							
24	0.35					0.60	0.02
25							
26	0.48					0.03	0.05

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
27	1.18					0.58	0.14
28	1.36					2.09	0.22
29	0.66					0.95	0.21
30							
31	0.83					0.73	0.18
32	1.31					1.67	0.24
33	1.47					0.27	0.22
34	0.82					1.96	0.19
35	1.84					1.44	0.12
36							
37	1.30					3.19	0.23
38	0.56					1.27	0.13
39	0.62					0.18	0.12
40	1.20					2.38	0.19
41	1.12					0.27	0.18
42	0.88					1.18	0.28
43	0.63		0.40			1.52	0.21
44	1.65					3.28	0.11
45	0.51					2.07	0.09
46		1.61				3.41	0.15
47		1.32				1.67	0.11
48		1.24				2.70	0.32
49		1.30				0.74	0.16
50		2.53				5.05	0.10
51		0.68				1.55	0.33
52		0.94				1.71	0.19

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
53		0.27				0.69	0.04
54		0.54				0.53	0.41
55		0.79				0.38	0.18
56		0.35				-0.01	0.19
57		1.44				3.94	0.10
58		0.74				2.46	0.16
59		1.19				2.70	0.31
60		0.71				0.80	0.26
61		0.86				3.02	0.06
62		1.51				1.59	0.30
63		1.31				3.69	0.10
64		0.62				0.67	0.14
65							
66		1.65				2.03	0.14
67		0.68				2.23	0.20
68		1.28				3.54	0.14
69		1.04				1.94	0.13
70		0.72				0.15	0.11
71		2.01				4.65	0.17
72		1.55				3.31	0.14
73		1.17				0.87	0.29
74		1.94				4.97	0.13
75		2.21				3.97	0.18
76		1.01				2.78	0.18
77		1.84				3.78	0.24
78		0.28				1.76	0.21

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
79		2.10				5.66	0.23
80		1.80				3.06	0.28
81		1.46				1.29	0.16
82		1.31				3.78	0.23
83		2.21				3.23	0.19
84		1.00		0.73		3.35	0.08
85		0.85				1.56	0.25
86		0.65		1.07		4.66	0.09
87		1.01				0.68	0.28
88		0.78				2.01	0.13
89		1.23				2.40	0.17
90		1.18				2.16	0.19
91			1.2			0.44	0.21
92			0.99			0.63	0.21
93			1.26			1.46	0.09
94			0.91			0.41	0.25
95			1.00			0.73	0.23
96			0.55			-0.64	0.03
97			0.64			0.52	0.17
98			0.22			0.16	0.03
99			0.52			-0.08	0.06
100			0.59			0.75	0.1
101			0.73			-0.75	0.01
102			0.62			1.14	0.11
103			0.44			1.15	0.12
104			1.17			0.61	0.15

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
105			0.19			0.06	0.03
106			0.41			0.39	0.06
107			0.46			-0.14	0.02
108			0.63			1.08	0.13
109			0.76			-0.47	0.01
110			0.94			1.68	0.08
111			0.35			-0.52	0.02
112			0.72			0.66	0.21
113			0.60			-0.47	0.02
114			0.66			0.57	0.1
115			1.40			3.08	0.14
116			0.20			0.08	0.04
117			0.86			-0.96	0.02
118			0.36			-0.29	0.04
119			1.40			2.89	0.12
120			0.51			0.75	0.15
121			0.87			-0.08	0.09
122			1.00			0.39	0.14
123							
124			0.71			-1.13	0.06
125			0.7			1.49	0.22
126			0.07			1.62	0.22
127			0.43			-0.24	0.02
128			0.68			0.83	0.16
129			0.94			2.25	0.16
130			0.98			0.64	0.25

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
131			0.41			1.53	0.23
132			0.77			0.24	0.14
133			0.41			-0.22	0.02
134			1.37			0.76	0.21
135			1.11			1.32	0.12
136				0.72		2.43	0.16
137				0.71		2.40	0.13
138				1.94		2.43	0.21
139				1.32		1.40	0.33
140				2.50		6.01	0.1
141				1.96		3.40	0.07
142				0.55		1.41	0.14
143							
144				1.58		2.21	0.25
145				1.40		2.79	0.22
146				1.11		1.88	0.19
147				1.58		2.49	0.22
148				1.44		2.08	0.25
149				0.73		0.72	0.11
150				1.16		0.62	0.16
151				1.86		3.18	0.21
152				0.63		1.25	0.31
153				1.55		3.05	0.14
154				0.87		-0.61	0.06
155				0.56		-0.16	0.03
156			0.24	2.23		5.28	0.21

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
157				0.90		0.58	0.2
158							
159				1.24		1.78	0.18
160				1.22		3.08	0.05
161				1.44		2.7	0.23
162				1.48		4.53	0.13
163				0.75		1.71	0.23
164				0.8		0.46	0.2
165				2.21		3.12	0.07
166				1.36		3.72	0.19
167				1.62		3.59	0.14
168				0.52		1.13	0.17
169				1.92		3.54	0.11
170				0.74		0.82	0.18
171				2.11		3.39	0.28
172				0.82		2.06	0.07
173				2.31		4.87	0.29
174				1.09		1.74	0.18
175				0.09		1.01	0.11
176				1.30		3.20	0.23
177				1.10		1.90	0.21
178				0.54		-0.09	0.09
179				1.88		3.86	0.25
180				1.26		2.85	0.14
181			0.20		1.31		
182					1.43		

Table A.1: Item parameters posterior mean - Enem.

Item	a_H	a_{NS}	a_{Lang}	a_{MT}	a_{Essay}	b	c
183					2.45		
184					1.85		
185					0.97		

Bibliography

Anderson, T., & Rubin, H. (1956). Statistical inference in factor analysis in: Neyman j (ed) proceedings of the 3rd berkeley symposium on mathematical statistics and probability, vol v.

Bambirra Gonçalves, F., da Costa Campos Dias, B., & Machado Soares, T. (2018). Bayesian item response model: a generalized approach for the abilities' distribution using mixtures. *Journal of Statistical Computation and Simulation*, *88*(5), 967–981.

Béguin, A. A., & Glas, C. A. (2001). Mcmc estimation and some model-fit analysis of multidimensional irt models. *Psychometrika*, *66*(4), 541–561.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846.
URL `http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x`

Carroll, J. (1993). Human cognitive abilities: A survey of factor-analytic studies. 10.1017. *CBO9780511571312*.

De La Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353.

de la Torre, J., & Song, H. (2009). Simultaneous estimation of overall and domain abilities: A higher-order irt model approach. *Applied Psychological Measurement*, *33*(8), 620–639.

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of personality and social psychology*, *91*(6), 1138.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436.

Huang, H.-Y., Wang, W.-C., Chen, P.-H., & Su, C.-M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, *37*(8), 619–637.

Johnson, V. E., & Albert, J. H. (2006). *Ordinal data modeling*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, *7*(1), 110–120.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.

Sheng, Y., & Wikle, C. K. (2008). Bayesian multidimensional irt models with a hierarchical structure. *Educational and psychological measurement*, *68*(3), 413–430.

Sheng, Y., & Wikle, C. K. (2009). Bayesian irt models incorporating general and specific abilities. *Behaviormetrika*, *36*(1), 27–48.