KU LEUVEN

FACULTY OF SCIENCE

Generalized Linear Latent and Mixed Models:

method, estimation procedures, advantages, and applications to educational policy.

Jose Manuel Rivera Espejo

Supervisor: Prof. Geert Molenbegrhs
Affiliation *(optional)*

Co-supervisor: Prof. Wim Van den Noortgate *(optional)*
Affiliation *(optional)*

Thesis presented in fulfillment of the requirements for the degree of Master of Science in Statistics and Data Science for Social, Behavioral and Educational Sciences

Academic year 2020-2021

Dedication

To Manuel, for being my friend and father.
To Margarita, Susan, and Marysu, for their relentless encouragement.
To Ana, for showing me the value of family, here in this moorland.
To both of you, as you are always in my mind.
And to all that knowingly or not, help me to get here.
I am lucky due to all of you.
I hope I make you all proud.

A Manuel, por ser mi amigo y mi padre.
A Margarita, Susan y Marysu, por su incansable aliento.
A Ana, por mostrarme el valor de la familia, aquí en este páramo.
A ustedes dos, que siempre las tengo en mente.
Y a todos los que sabiendolo o no, me ayudaron a llegar aquí.
Soy un suertudo gracias todos ustedes.
Espero llenarlos de orgullo.

Acknowledgment

(work in progress)

Abstract

(work in progress)
Keywords:

Contents

List of Figures

List of Tables

Abbreviations

CFA	Confirmatory Factor Analysis.
DDM	Dual Dependence Models.
EFA	Exploratory Factor Analysis.
GLLAMM	Generalized Linear Latent and Mixed Model.
GLM	Generalized Linear Model.
GLMM	Generalized Linear Mixed Model.
HMC	Hamiltonian Monte Carlo.
IRT	Item Response Theory models.
MCMC	Markov Chain Monte Carlo.
ML	Maximum Likelihood.
SEM	Structural Equation Model.

Chapter 1

Introduction

1.1 Preliminar considerations

Local independence is one of the key assumptions of Item Response Theory (IRT) models, and it is comprised of two parts: (i) local item independence, and (ii) local individual independence [3, 21]. In the former case, the assumption entails that the individual's response to an item does not affect the probabilty of endorsing another item, after conditioning on the individual's ability. While in the case of the latter, the assumption considers that an individual's response to an item is independent of another person's response to that same item [44].

The literature has shown that IRT models are not robust to the violation of local independence. The transgression of the assumption affects model parameter estimates, inflates measurement reliabilities and test information, and underestimates standard errors, e.g. Yen [56], Chen and Thissen [9], and Jiao et al. [24].

However, item response data arising from educational assessments often display several type of dependencies, e.g. testlets, where items are constructed around a common stimulus [53]; the measurement of multiple latent traits within individuals [44]; cluster effects, where correlation among individuals results from the sampling and measurement mechanism used to gather the data [43]; among others. A good motivating example is the reading comprehension sub-test, from the Peruvian public teaching career national assessment. The test is designed to measure three hierarchically nested sub-dimensions of reading comprehension: literal, inferential and reflective abilities. Furthermore, the items are bundled together in testlets related to a common text or passage. Finally, multiple cluster effects are present, e.g. district or regional clustering, to mention a few.

Recent studies have proposed Dual Dependence Models (DDM) to deal with the teslet and individual clustering dependencies observed in the data [15, 14, 13, 24, 11, 12, 44, 7]. The majority of these representations have been developed under the bayesian framework, and they are similar in parametrization to multilevel models. On the other hand, an almost independent line of research, the Generalized Linear Latent and Mixed Models (GLLAMM) [38, 40, 48, 41], have extended the capabilities for the estimation of multiple latent traits at different hierarchical levels. These developments have been motivated mostly under the frequentist framework, and they are similar in parametrization to a hierarchical Structural Equation Model (SEM).

While the initial sense is that both developments are independent of each other, following their literature, one can easily notice that they share more than a resemblance. Both

followed a multilevel/hierarchical multidimensional approach to account for the clustering of persons within the samples and the item bundles (DDM), or the latent structures within the individuals (GLLAMM). However, it is important to point out that in some cases the model parametrization between the two developments differs in a way, that some of them appear to be useful only under their specific contexts. Fortunately, their integration under the bayesian framework is not only trivial, but it can be motivated under either type of models.

The benefits of the integration revolves around two facts: (i) the educational data often presents all of the aforementioned dependencies and more, as in the motivating example; and (ii) as it was point out in the second paragraph, in order to reach appropriate conclusions from the parameter estimates, IRT models need to account for all of these dependencies. The latter is particularly important as, more often than not, a researcher might be interested in produce inferences at the structural level of the model, i.e. how a different set of manifest variables explain the variability in the latent variables, or how the latent variables explain other manifest or latent variables, at different levels. As an example, one might be interested in finding evidence if the latent "abilities" of the teachers are explained by their initial educational conditions, i.e. if they were educated in an institute, university or both. The main purpose of this would be to identify the type of teacher that might benefit more from the "in-service training"[1], offered by the national educational authorities, making the intervention cost-effective.

From the previous description, one can infer that the proposed IRT representation would be complex and highly dimensional. Moreover, as educational assessments are usually described in a binary way, i.e. the individual either endorse or not the item, and because not all individuals are assessed by all the items, the model will be train on sparse data. From the modeling perspective, neither of previous points presents a challenge for the bayesian framework. However, it has long been recognized that complex parametrizations, that allow this powerful modeling schemes, introduce pathologies that make Markov Chain Monte Carlo methods (MCMC) face performance challenges [16, 17, 33, 34, 4]. This is highly relevant, because in order to make inferences about the posterior distribution of the parameters, the chains need to achieve three requirements, highly related to the performance of the method: stationarity, convergence and good mixing [30].

Throughout the bayesian IRT literature one often finds that four solutions are offered to ensure the fulfillment of the previous requirements, and they can be classified in two main groups: (i) solutions that involves changing the settings of the MCMC methods, and (ii) solutions that involves readjusting the bayesian model.

In the first category we find two proposals: (a) increasing the number of iterations per chains, with large burn-in and thinning processes, and (b) designing model-specific MCMC algorithms. The easiest and more prevalent is the former, e.g. Fujimoto [14] used chains with $60,000$ iterations, where $15,000$ were discarded and the remaining were thinned in jumps of 3; while Fujimoto [13] used $225,000$ iterations, with burn-in of $30,000$ and thinning with jumps of 15. Among the drawbacks of this solution are the large computational times, and the user involvement on deciding the specific setting for the process, which could be different for different parameters in the same model. On the other hand, several authors have developed high-tech MCMC algorithms that aim to

[1]Intervention designed with the purpose of potentiate specific abilities in teachers that are current part of the public teaching career.

optimize their performance within a particular class of models [34]. In these cases, the developers re-evaluate not only the use of the programming language, with the purpose of speeding and improving performance (e.g. Fujimoto [14]); but also the inclusion of ad-hoc model assumptions, like the ones used in large software packages like Mplus [31] or Stata (e.g. Rabe-Hesketh et al. [39]). It is clear from the previous that this solution is not accessible to all researchers, either because of the lack of programming skills or the restrictive cost of access involved in acquiring the software, but more importantly, this solutions is not always applicable to a wider framework of similar models [34].

In the second category we also find two proposed solutions: (a) re-write the model in an alternative parametrization, and (b) encode prior information through the prior distributions. On both solutions the purpose is to ensure the identification of the parameters within the model, which helps to stabilize the MCMC procedure [18]. An example of the former is Fujimoto [14], who decomposed the items' discriminatory parameters into overall and specific item discriminations. For the latter, Fujimoto [15] used informative priors also for the items' discrimination parameters.

More often than not, researchers use two or more of the aforementioned solutions to achieve a good performance in the chains. However, as point out by Betancourt and Girolami [4], even the most simple hierarchical models present formidable pathologies, that no simple correction can be performed to visit the posterior distribution properly. This is true no matter the rotation/rescaling of the parameter, or the amount of data. In this context, several authors [16, 17, 33, 34, 4] showed that prior information can be included in the model, not only through the prior distributions, but also by the encoding of the model itself, changing the posterior geometries and favoring the performance of the MCMC chains.

Given all of the above, the present research will focus its attention on showing how easy is to account for all of the dependencies present in educational data, under the GLLAMM framework. Furthermore, given that only the literature related to hierarchical models have shown the benefits of changing the posterior geometries, through the use of the non-centered parametrization [16, 17, 33, 34, 4], a similar assessment in the context of IRT models seems sensible. Finally, the research will apply the newfound knowledge to a data coming from a large Teacher's standardized educational assessments from Peru.

1.2 Objectives

As mentioned in the previous paragraph, the present research has a three-fold purpose:

1. **Motivate the GLLAMM for binary outcomes** [38, 40, 48, 41]. The representation will have a special emphasis in modeling multiple hierarchical latent structures and testlets, effectively blurring the division between the framework and IRT models.

2. **Empirically evaluate the benefits of changing the posterior geometry**. The emphasis here will be in evaluating the recovery of the parameters, and the benefits and shortcomings of changing the geometry of the bayesian models, through the centered and non-centered parametrization described in Gelfand et al. [16, 17], Papaspiliopoulos et al. [33, 34], and Betancourt and Girolami [4].

3. **Apply the model and its parametrization to a real data setting**. Here the emphasis will be to assess the conclusions arrived from the application of the model and what they imply for the educational authorities.

Given the aforementioned goals, the researcher believes the master's thesis contributes to the literature in two aspects:

1. In a the theoretical and methodological sense, as the research is focused describing a model that effectively controls for multiple dependencies observed in educational datasets; and

2. In a more practical sense, as it will provide empirical evidence that changing the posterior geometries benefit the performance of MCMC methods, allowing the modeler to arrive to proper inferences.

Finally, it is important to mention, that the computational implementation of the method will be developed in `Stan` [51] and `R` [35, 50].

1.3 Organization

Chapter 2, The Bayesian GLLAMM for binary outcomes, will describe the model, its components, characteristics, assumptions and properties for binary outcomes, under the bayesian framework.

Chapter 3, Simulation study,

Chapter 4, Application, will describe the instruments, its data collection process, the "dimensions" under analysis, and conclusions achieved by the application of the model.

Finally, **Chapter 5, Conclusions**, will discuss the conclusion for the research, under the aforementioned framework. Finally, it will outline the path of future research that can be derived from the present effort.

Chapter 2

The Bayesian GLLAMM for binary outcomes

The Generalized Linear Latent and Mixed Model (GLLAMM) is a framework that unifies a wide range of latent variable models. Developed by Rabe-Hesketh et al. [38, 40, 39], Skrondal and Rabe-Hesketh [48], Rabe-Hesketh et al. [41], the method was motivated by the need of a Multilevel Structural Equation Model (SEM) that accommodates for unbalanced data, noncontinuous responses and the use of cross-level effects among latent variables.

This chapter presents the definition, characteristics, assumptions and properties of such framework.

2.1 Definition

Following Rabe-Hesketh et al. [38, 41], we depart from the traditional multivariate framework for formulating factor and structural models, i.e. a "wide" data format, and adopt a univariate approach, i.e. "long" or vectorized format. In that sense, for each unit, the response variables are "stacked" in a single response vector, with different variables distinguished from each other, by a design matrix. With this structure, we proceed to outline the three parts of the framework:

1. The response model,
2. The structural latent variable model, and
3. The distribution of the latent variables.

For a detailed description of some of the special cases of multilevel SEM, that can be derived with this framework, refer to Appendix A.

2.1.1 Response model

As outlined by the authors, conditional on the latent variables, the response model is a Generalized Linear Model (GLM) defined by a systematic and a distributional part. For the systematic part, a linear predictor and a link function are selected, in accordance to

the characteristics of the manifest variables. On the other hand, for the distributional part, a distribution from the exponential family is selected.

In the following sections, we proceed to describe the linear predictor, the link function and the distributions accommodated by the framework.

Linear predictor

For a model with L levels and M_l latent variables at $l > 1$ levels, the linear predictor takes the following form:

$$v = \mathbf{X}\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_{(l)}} \eta_m^{(l)}\mathbf{Z}_m^{(l)}\boldsymbol{\lambda}_m^{(l)} \tag{2.1}$$

where $\mathbf{X}$ is a design matrix that maps the parameter vector $\boldsymbol{\beta}$ to the linear predictor, $\eta_m^{(l)}$ the mth latent variable at level l ($m = 1, \ldots, M_{(l)}$ and $l = 1, \ldots, L$), and $\mathbf{Z}_m^{(l)}$ a design matrix that maps the vector of loadings $\boldsymbol{\lambda}_m^{(l)}$ to the mth latent variable at level l.

Note that wo do not use subscripts for the units of observation at different levels. This decision was made with the purpose of avoiding the use of mathematical definitions with large number of subscripts. However, a careful reader should consider that equation (2.1) rest on the assumption that each unit is identified at their appropriate level. For special cases of multilevel SEM, and their use of subscripts, refer to Appendix A.

Links and Distributions

As in the GLM framework, the model "links" the expectation of the conditional response, to the linear predictor, through a inverse-link function $h(\cdot)$, in the following form:

$$\mu = E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] = h(v) \tag{2.2}$$

where equation (2.2) can be re-written in terms of the link function $g(\cdot) = h^{-1}(\cdot)$:

$$g(\mu) = g(E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}]) = v \tag{2.3}$$

with $\boldsymbol{\eta} = \left[\eta^{(2)T}, \ldots, \eta^{(L)T}\right]^T$ and $\mathbf{Z} = \left[\mathbf{Z}^{(2)T}, \ldots, \mathbf{Z}^{(L)T}\right]^T$, as the "stacked" vector of latent variables, and the "stacked" design matrices of explanatory variables, for all L levels, respectively. Additionally, $\boldsymbol{\eta}^{(l)} = \left[\eta_1^{(l)}, \ldots, \eta_{M_{(l)}}^{(l)}\right]^T$ and $\mathbf{Z}^{(l)} = \left[\mathbf{Z}_1^{(l)T}, \ldots, \mathbf{Z}_{M_{(l)}}^{(l)T}\right]^T$, denotes the vector of latent variables, and the "stacked" design matrix of explanatory variables, at level l, respectively.

Finally, the response model specification is complete when we select an appropriate distribution from the family of exponential distributions. The types of responses that can be accommodated by the framework are the following:

Dichotomous:
It results from selecting an appropriate inverse-link function for the expected value of the manifest variable, which describe the probability of endorsing one of the two available categories,

$$\begin{aligned} \mu &= E[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi \\ &= h(\kappa - v) \end{aligned} \tag{2.4}$$

where κ is the decision threshold, and $h(\cdot)$ can be defined in three ways:

$$h(x) = \begin{cases} exp(x)[1 + exp(x)]^{-1} \\ \Phi(x) \quad \text{No closed form.} \\ exp(-exp(x)) \end{cases} \tag{2.5}$$

which corresponds to the logistic, standard normal $\Phi(x)$, and Gumbel (extreme value type I) *cumulative distributions*, respectively. In terms of link functions, the distributions corresponds to the well known logit, probit and complementary log-log link functions, respectively.

Alternatively, the same parametrization can be achieved using the concept of an underlying latent variable in the form $y^* = v + \epsilon^*$, where $y = 1$ if $y^* \geq \kappa$, and ϵ^* can have a distribution as the ones defined in equation (2.5). It is important to mention that under this parametrization, the threshold parameters κ and the $\boldsymbol{\beta}$ are confounded as they serve similar purposes, so only one would be estimated.

Finally, the distributional part is defined by a Binomial distribution,

$$\begin{aligned} f[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \binom{n}{k} \mu^k (1-\mu)^{n-k} \\ &= \binom{n}{k} \pi^k (1-\pi)^{n-k} \end{aligned} \tag{2.6}$$

where k denotes the number of successes in n independent Bernoulli trials.

Heteroscedasticity and over-dispersion in the response

Much like the Generalized Linear Mixed Model framework (GLMM), the GLLAMM allows to model heteroscedasticity, and over- or under-dispersion by adding random effects to the linear predictor, at level 1. The types of responses, in which such characteristics can be modeled, are the following:

Dichotomous:
In a more straightforward way, we model over- or under-dispersion by modifying equation (2.4), to include random intercepts at level 1, in the following form:

$$\begin{aligned} \mu &= P[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi \\ &= h(\kappa - v + \boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) \end{aligned} \tag{2.7}$$

2.1.2 Structural model for the latent variables

The structural model for the latent variables has the form:

$$\underset{(M\times 1)}{\boldsymbol{\eta}} = \underset{(M\times M)}{\mathbf{B}} \underset{(M\times 1)}{\boldsymbol{\eta}} + \underset{(M\times Q)}{\boldsymbol{\Gamma}} \underset{(Q\times 1)}{\mathbf{W}} + \underset{(M\times 1)}{\boldsymbol{\zeta}} \tag{2.8}$$

where $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are parameter matrices that maps the relationship between the latent variables $\boldsymbol{\eta}$, and the vector of "stacked" covariates $\mathbf{W}$, respectively; $\boldsymbol{\zeta}$ is a vector of errors or disturbances, and $M = \sum_l M_l$. Notice that while equation (2.8) resembles to

single-level structural equation models, the main difference lies in the fact that the latent variables may vary at different levels. Additionally, considering that $\boldsymbol{\eta}$ has no feedback effects, and it is permuted and sorted according to the levels, $\mathbf{B}$ is defined as a strictly upper triangular matrix. In this regard, it is important to mention that,

1. The absence of feedback loops implies that the method deals with non-recursive models, i.e. none of the latent variables are specified as both causes and effects of each other [27]; this in turn allows the easy estimation of the model parameters.

2. The strictly upper triangular structure reveals that the framework does not allow latent variables to be regressed on lower level latent or observed variables, as such specification is more related to the use of formative, rather than reflective, latent variables. For a detail explanation on the topic refer to Edwards and Bagozzi [10].

Notice, however, the previous restrictions does not hinder the ability of the method to model contextual effects, after controlling the lower level compositional effects. For examples of such refer to Appendix A.

2.1.3 Distribution of the latent variables

Finally, to fully specify the framework, and provide a scale for the latent variables, we have to make assumptions for either the distribution of the disturbances $\boldsymbol{\zeta}$ or the latent variables $\boldsymbol{\eta}$. If our research interest lies in the structural equation model, it is more convenient to make assumptions for the distribution of the disturbances; otherwise, we make assumptions for the distributions for the latent variables.

Furthermore, as in the hierarchical framework, it is assumed the latent variables at different levels are independent, whereas latent variables at the same level may present dependency. In that sense, we presume all latent variables at level l to have a multivariate normal distribution with zero mean and covariance matrix Σ_l, i.e. $\boldsymbol{\eta}^{(l)} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_l)$. It is important to emphasize that, while the multivariate normal distribution is widely used in these settings, it is not the only distribution that can be assumed. Rabe-Hesketh, Skrondal and Pickles [37] have provided evidence that it can be even left unspecified, by using non-parametric maximum likelihood estimation.

2.2 Model identification

(work in progress)

The structure of the latent variables is specified by the number of levels L and the number of latent variables Ml at each level. A particular level may coincide with a level of clustering in the hierarchical dataset. However, there will often not be a direct correspondence between the levels of the model and the levels of the data hierarchy.

2.3 The bayesian estimation

The practical use of GLLAMM requires the estimation of the parameters associated with the items and the individuals' latent abilities. These can be obtained within two frame-

works: the classical (frequentist), and the bayesian. The current chapter center its attention on describing the bayesian framework using the Markov Chain Monte Carlo method (MCMC). For a full development of GLLAMM under the frequentist estimation framework refer to Rabe-Hesketh et al. [38, 40], Skrondal and Rabe-Hesketh [48], Rabe-Hesketh et al. [41].

2.4 Benefits and shortcomings

The reasons on why bayesian statistics is attractive to perform the estimation of the parameters of any model, and especially under the GLLAMM framework, are:

1. The bayesian estimates are at least as good as the frequentist estimates [2, 54, 23].

2. It is built on a simulation-based estimation method, therefore, it can handle all kinds of priors and data-generating processes [12]. This is especially useful with highly complex and over-parameterized models, where other methods are unfeasible or work poorly [2, 26].

3. The model definitions, i.e. the likelihood for the data and priors for the parameters, are used to estimate the corresponding posterior distributions. However, the definitions can also be used in a generative way, i.e. simulate observations, allowing us to test the ability of the method/data to recover the parameters of interest [30].

4. It allow us to integrate prior beliefs or knowledge about the parameters beyond the observed responses [12, 48]. This is especially useful when we have issues of non-convergence or improper estimation of the parameters under the Maximum Likelihood methods (ML). Examples of these cases are:

   (a) Estimating abilities when individuals have null scores or aberrant response patterns, i.e. examinees that answered some relatively difficult and discriminating items correctly, while answering some of the easiest incorrectly. [21, 1].

   (b) Estimating parameters that need to be confined to a permitted parameter space, e.g. the estimation of positive unique factors variances, where the opposite is known as 'Heywood cases' [28]

   (c) Estimating parameters under a sparse data structure, where the asymptotic theory is unlikely to hold [12];

Finally, in terms of shortcomings, the bayesian framework has the following inconveniences:

1. It exposes the user to arbitrary" decisions about the running of the chains, e.g. how many iterates does the chain need to achieve precise estimates?, what is the right size for the burn-in and warm-up phases?, how should the thining procedure be performed? [48].

2. The user has many options to assess if the chain achieves stationarity, convergence or good mixing, and most of them are visual. This makes it hard to assess if the chain converges to a proper distribution [19].

3. The procedure makes it hard to discover parameters' lack of identification [48]. Inadequate mixing of the chain could lead us to think unidentified parameters have been estimated with precision, when in fact they have a 'flat' posterior [25].

4. Sometimes the geometry of the model makes it hard to find proper solutions for the parameter space. This is especially true in hierarchical models. Under this circumstances, the scientist needs to re-parameterize the model to a non-centered form, i.e. remove the dependence of the parameters on other sampled parameters [20]. In those cases, the complexity of the transformation limit the ability of the scientist to communicate/share the implementation [30].

5. The procedure requires more time to achieve a proper solution, compared to the classical methods. This is especially true in models with high complexity [52, 45].

Although some of the shortcomings has made the use of bayesian methods a "controversial" issue, most of these already have an acceptable solution.

For the first point, a popular approach to solve the issues is to use a large number of iterates, or multiple chains with different initial states. This is mostly applicable under the Metropolis-Hastings and Gibbs sampling methods. However, as we will see in section 2.6, the Hamiltonian Monte Carlo method (HMC) [4] implements a different sampling mechanism that is less reliant on these decisions.

About the second shortcoming, it is well accepted that the visual assessment of stationarity and convergence is easier, and this procedure usually has additional support from statistics like `Rhat` [18]. On the contrary, a visual evaluation of 'good' mixing remains as a hard task. A popular approach to increase the possibility of a well mixed chain is to change the geometry of the model [30]. However, the implementation of the approach does not necessarily ensure the required property.

On the third point, the most common solution is to use regularizing priors, i.e. priors that are more 'skeptical' of wider parameter spaces [30]. However, it is important to mention, there are scenarios where one can achieve poor parameter estimates, even in the presence of 'enough' data and regularizing priors, e.g. the estimation of the variance parameters in random effects models [48], but this is also applicable to the classical estimation procedures.

Finally, the fourth and fifth points can be considered as the 'price' a scientist has to pay to be able to fit complex models, that are in more accordance with the observed data generating processes.

2.5 Bayesian framework

2.5.1 Prior distribution

2.5.2 Initial start

2.5.3 Likelihood

2.5.4 Posterior distribution

2.6 Computational implementation

(work in progress)
see
- Gelman et al (2011) - Handbook of Markov Chain Monte Carlo
- McElreath (2020) - Statistical Rethinking

Rethinking: Warmup is not burn-in. Other MCMC algorithms and software often discuss burn- in. With a sampling strategy like ordinary Metropolis, it is conventional and useful to trim off the front of the chain, the "burn-in" phase. This is done because it is unlikely that the chain has reached stationarity within the first few samples. Trimming off the front of the chain hopefully removes any influence of which starting value you chose for a parameter. 156 But Stan's sampling algorithms use a different approach. What Stan does during warmup is quite different from what it does after warmup. The warmup samples are used to adapt sampling, to find good values for the step size and the number of steps. Warmup samples are not representative of the target posterior distribution, no matter how long warmup continues. They are not burning in, but rather more like cycling the motor to heat things up and get ready for sampling. When real sampling begins, the samples will be immediately from the target distribution, assuming adaptation was successful.

The procedure will be with the aid of `Stan` [51] and `R` [35, 50] to retrieve .

(to be erased)

2.7 Relationship with other modeling schemes

From section 2.1, it is evident that the GLLAMM framework shares some common ground, and even extends, some of the most important modeling schemes, such as the GLM, GLMM, SEM, and the Generalized Latent Model framework, from which the Item Response Theory Model (IRT) stand out.

2.7.1 Generalized Linear and Mixed Models

The Generalized Linear Model (GLM) framework, presented by Nelder and Wedderburn [32], and further developed by McCullagh and Nelder [29], was formulated with the purpose of expanding the linear regression model to other types of responses, like dichotomous, and counts. The scheme generalizes the linear model by "linking" the mean response variable to a linear predictor, and further allowing the magnitude of the variance, of each measurement, to be a function of its predicted value. Finally, the scheme is fully defined after selecting a distribution, from the exponential family, to model the distribution of the response variable.

As expressed in the previous paragraph, the GLM framework fixes the relationship of the modeled dispersion to the mean value, e.g. in the counts case $\mu = \lambda$, and $v(\mu) = \lambda$. However, in practice, this assumption is often violated as the data can present over- or under-dispersion. Even in the continuous response case, where the mean and variance function are not related, the model assumes that the errors are homoscedastic, identical and independently distributed. However, this assumption is often violated when the units of analysis are correlated or belong to a cluster, e.g. when students are nested in schools, and these are further nested in districts or states.

It is important to mention that, while the GLM framework can model heteroscedasticity, over- or under-dispersion, it does it in a way that does not allow them to be dependent on covariates, something that might be of interest for a researcher.

Given the restrictions of GLM, the Generalized Linear Mixed Model (GLMM) framework was developed. The method handled the hierarchical or clustered structure in the data, and in doing so, indirectly modeled the heteroscedasticity, over- or under-dispersion by adding latent variables, called "random effects", to the linear predictor. Under the framework, the random variables are often interpreted as the effects of unobserved covariates, at different levels, that induce dependence among lower-level units [41], and can be further explained by additional observed covariates.

From the previous description, it is easy to notice that the GLLAMM framework uses the same generalization and distributional assumptions, for the response variables, as the GLM; while it borrows the idea of modeling the hierarchical or clustered structure in the data, by including random effects; from the GLMM. However, it is clear that the GLLAMM further generalize both, by allowing the framework to model measurement error at different levels of the hierarchy in the data.

2.7.2 Structural Equation Models

Considering that, in practice, researchers are often faced with variables that cannot be measured directly or reflect measurement error, e.g. intelligence, depression, student abilities, among other; the statistical literature was instigated to develop methods that can handle such data characteristics.

(work in progress)
The disciplinary seeds of Structural Equation Models (SEM) were set by [**?**], with a factor model on intelligence testing, passing through [55], with a path analysis in the context of genetic and biology, to finally land in the sociological field, with the work of [5].

to include several features of the previous modeling scheme, i.e. generalized linear mixed models, the framework is characterized by the fact that it is a method that can impute relationships between unobserved factors or latent variables, and observable or manifest variables. Under this framework, it is assumed that such "common factors" are responsible for the variation and dependence in the manifest variables.

mention Factor Models, Item Response Theory and Generalized Latent Models, and Multilevel Structural Equation Models 2.1.3

multilevel structural equation models represent a synthesis between multilevel regression models and structural equation models. Considering that

2.8 Advantages and Disadvantages

Chapter 3

Simulation Study

Chapter 4

Application

4.1 Instruments

4.2 Data

4.2.1 Collection

4.2.2 Sample scheme

4.3 Results

Chapter 5

Conclusion and Discussion

5.1 Discussion

5.2 Conclusions

5.3 Future development

Appendix A

Additional Theory

A.1 Other links and distributions

1. **Continuous:**
   It results form selecting an identity link function for the scaled mean response,

$$\begin{aligned} \mu^* &= E[y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= v \end{aligned} \tag{A.1}$$

   where $\mu^* = \mu\sigma^{-1}$, $y^* = y\sigma^{-1}$, and σ denotes the standard deviation of the errors.

   On the other hand, the distributional part is defined by a Standard Normal distribution $\phi(x) = (2\pi)^{-1/2}exp(-x^2/2)$,

$$\begin{aligned} f(y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}) &= \phi(\mu^*)\sigma^{-1} \\ &= \phi(v)\sigma^{-1} \end{aligned} \tag{A.2}$$

   Notice that the same parametrization can be achieved considering $y^* = v + \epsilon^*$, and $\epsilon^* \sim N(0, 1)$. Additionally, the decision to standardize the response variables has been made with the purpose of making the estimation process easier, as such distribution is free of unknown parameters.

2. **Polytomous:**
   It results from selecting a generalized logistic inverse-link function [6] for the expected value of the response, which in this case, describe the probability of endorsing one of the S unordered available categories,

$$\begin{aligned} \mu_s &= E[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi_s \\ &= h(v_s) \end{aligned} \tag{A.3}$$

   where v_s is the linear predictor for category s ($s = 1, \ldots, S$), and $h(\cdot)$ is defined as:

$$h(x) = exp(x) \cdot \left[\sum_{s=1}^{S} exp(x)\right]^{-1} \tag{A.4}$$

It is important to note that, as in the dichotomous case, the same parametrization can be achieved using the concept of underlying continuous responses in the form $y_s^* = v_s + \epsilon_s$, where $y = s$ if $y_s^* > y_k^* \; \forall s, s \neq k$, ϵ_s have a Gumbel (extreme value type I) distribution, as the one defined in equation (2.5), and y_s denotes the random utility for the s category.

Finally, the distributional part is defined by a Multinomial distribution,

$$\begin{aligned} f[y = \{y_1, \cdots, y_S\}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \mu_s^{y_s} \\ &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \pi_s^{y_s} \end{aligned} \tag{A.5}$$

where y_s denotes the number of "successes" in category s.

3. **Ordinal and discrete time duration:**
For the ordinal case, the linear predictor is "linked" to the probability of endorsing category s, against all previous categories, in the following form:

$$\begin{aligned} \mu_s &= E[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y \leq y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[y \leq y_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(\kappa_s - v_s) - h(\kappa_{s-1} - v_{s-1}) \end{aligned} \tag{A.6}$$

where κ_s denotes the thresholds for category s. For discrete time duration, the linear predictor is "linked" to the probability of survival, in the sth time interval, as follows:

$$\begin{aligned} \mu_s &= E[t_{s-1} \leq T \leq t_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[T \leq t_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[T \leq t_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(v_s + t_s) - h(v_{s-1} + t_{s-1}) \end{aligned} \tag{A.7}$$

where T is the unobserved continuous time, and t_s its observed discrete realization. Additionally, for both type of responses, $h(\cdot)$ can be defined as the logistic, standard normal, and Gumbel (extreme value type I) *cumulative distributions*, as in equation (2.5).

Similar to the dichotomous and polytomous case, the same parametrization can be achieved using the concept of underlying latent variables with $y_s^* = v_s + \epsilon_s$, where $y = s$ if $\kappa_{s-1} < y_s^* \leq \kappa_s$, $\kappa_0 = -\infty$, $\kappa_1 = 0$, $\kappa_S = +\infty$, ϵ_s has one of the distributions in equation (2.5), and y_s denotes the random utility for the s category.

It is important to note, for discrete time duration responses, the logit link corresponds to a *Proportional-Odds model*, while the complementary log-log link to a *Discrete Time Hazards model* [42]. Other models for ordinal responses, such as the *Baseline Category Logit* or the *Adjacent Category Logit* models can be specified as special cases of the generalized logistic response function, defined in equation (A.4).

Finally, the distributional part is defined by a Multinomial distribution, as the one defined in equation (A.5).

4. **Counts and continuous time duration:**
It results from selecting an exponential inverse-link function (log link) for the expected value of the response,

$$\begin{aligned} \mu &= E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \lambda \\ &= exp(v) \end{aligned} \tag{A.8}$$

and a Poisson conditional distribution for the counts,

$$\begin{aligned} f[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= exp(-\mu)\mu^y(y!)^{-1} \\ &= exp(-\lambda)\lambda^y(y!)^{-1} \end{aligned} \tag{A.9}$$

It is important to mention that unlike the models for dichotomous, polytomous and ordinal responses, model for counts cannot be written under the random utility framework.

5. **Rankings and pairwise comparisons:**
Following Skrondal and Rabe-Hesketh [46], the parametrization for polytomous responses can serve as the building block for the conditional distribution of rankings. Selecting a "exploded logit" inverse-link function [8] for the expected value of the response, which describes the probability of the full rankings of category s,

(work in progress)

$$\begin{aligned} \mu_s &= P[\mathbf{R}_s = \{r_s^1, \dots r_s^1\}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi_s \\ &= h(v_s) \end{aligned} \tag{A.10}$$

where v_s is the linear predictor for category s ($s = 1, \dots, S$), and $h(\cdot)$ is defined as:

$$h(x) = \prod_{s=1}^{S} exp(x^s) \left[\sum_{s=1}^{S} exp(x^s) \right]^{-1} \tag{A.11}$$

Again, as in specific previous cases, the same parametrization can be achieved using the concept of underlying latent variables.

Finally, the distributional part is defined by a Multinomial distribution,

$$\begin{aligned} f[y = \{y_1, \cdots, y_S\}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \mu_s^{y_s} \\ &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \pi_s^{y_s} \end{aligned} \tag{A.12}$$

where y_s denotes the number of "success cases" in category s.

6. **Mixtures:**
Given the previous definitions, the framework easily lends itself to model five additional settings:

(a) **Different links and distributions for different latent variables**. This can be easily achieved by setting different links and distributions for each of the M_2 latent variables located at level 2.

(b) **Left- or right-censored continuous responses**. Common in selection models (e.g. [22]), they can be achieved by specifying an identity link and Normal distribution for the uncensored scaled responses, as in equations (A.1) and (A.2); and a scaled probit link and Binomial distribution otherwise, as in equations (2.5) and (2.6).

(c) **zero-inflated count responses**. where a log link and a Poisson distribution is set for the counts, as in equations (A.8) and (A.9); and a logit link and Binomial distribution is specified to model the zero center of mass, as in equations (2.4) and (2.6).

(d) **Measurement error in covariates**. this setting occurs when standard models use variables, with measurement error, as covariates, e.g. a logistic regression with a continuous covariate that presents measurement error. For more details on this type of setting see Rabe-Hesketh, Skrondal and Pickles [37], Rabe-Hesketh, Pickles and Skrondal [36], and Skrondal and Rabe-Hesketh [47].

(e) **Composite links**. Useful for specifying proportional odds models for right-censored responses, for handling missing categorical covariates and many other model types. For more details on this type of settings see Skrondal and Rabe-Hesketh [49].

Heteroscedasticity and over-dispersion in the response

Much like the Generalized Linear Mixed Model framework (GLMM), the GLLAMM allows to model heteroscedasticity, and over- or under-dispersion by adding random effects to the linear predictor, at level 1. The types of responses, in which such characteristics can be modeled, are the following:

1. **Continuous:**
   We model **heteroscedasticity** in the following form:

$$\sigma = exp(\boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) \tag{A.13}$$

   Notice that the previous formula implies that equation (A.2) can be re-written in the following form:

$$f(y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}) = \phi(v + \boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) \tag{A.14}$$

   where $\mathbf{Z}^{(\mathbf{1})}$ is the design matrix that maps the random effects $\boldsymbol{\alpha}$. Notice that equation (A.14) effectively corresponds to a model that includes random intercepts at level 1.

2. **Ordinal, and discrete time duration:**
   Similar to the dichotomous case, by including random intercepts at level 1 in equation (A.6), we can model over- or under-dispersion:

$$\begin{aligned} \mu_s &= P[y \leq y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[y \leq y_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(\kappa_s - v_s + \boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) - h(\kappa_{s-1} - v_{s-1} + \boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) \end{aligned} \tag{A.15}$$

A similar parametrization can be used for discrete time duration.

3. **Counts, and continuous time duration:**
Finally, modifying equation (A.8) allow us to model over- or under-dispersion under a counts model:

$$\begin{aligned} \mu &= E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \lambda \\ &= exp(v + \boldsymbol{\alpha}^T \mathbf{Z}^{(\mathbf{1})}) \end{aligned} \tag{A.16}$$

A.2 Sampling scheme

Bibliography

[1] Azevedo, C. [2003]. *Métodos de estimação na teoria de resposta ao item*, Master's thesis, Universidade de São Paulo (USP).
url: https://teses.usp.br/teses/disponiveis/45/45133/tde-05102004-163906/pt-br.php.

[2] Baker, F. [1998]. An investigation of the item parameter recovery characteristics of a gibbs sampling procedure, *Applied Psychological Measurement* **22**(22): 153–169.
doi: https://doi.org/10.1177/01466216980222005.

[3] Baker, F. [2001]. The basic of item response theory, *Technical report*, ERIC Clearinghouse on Assessment and Evaluation.

[4] Betancourt, M. and Girolami, M. [2012]. Hamiltonian monte carlo for hierarchical models.
url: arxiv.org/abs/1312.0906v1.

[5] Blalock, H. [1961]. Causal inferences in nonexperimental research.

[6] Bock, R. [1972]. Estimating item parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika* **37**(1).
doi: https://doi.org/10.1007/BF02291411.

[7] Bradlow, E., Wainer, H. and Wang, X. [1999]. A bayesian random effects model for testlets, *Psychometrika* **64**(2): 153–168.
doi: https://doi.org/10.1007/BF02294533.

[8] Chapaaan, R. and Staelin, R. [1982]. Exploiting rank ordered choice set data within the stochastic utility model, *Journal of Marketing Research* **19**(3): 288–301.
doi: https://www.doi.org/10.1177/002224378201900302.

[9] Chen, W. and Thissen, D. [1997]. Local dependence indexes for item pairs using item response theory, *Journal of Educational and Behavioral Statistics* **22**(3): 265–289.
doi: https://doi.org/10.3102/10769986022003265.

[10] Edwards, J. and Bagozzi, R. [2000]. On the nature and direction of relationships between constructs and measures, *Psychological Methods* **5**(2): 155–174.
doi: https://www.doi.org/10.1037/1082-989X.5.2.155.

[11] Flores, S. [2012]. *Modelos testlet logísticos y logísticos de exponente positivo para pruebas de compresión de textos*, Master's thesis, Pontificia Universidad Católica del Perú.

[12] Fox, J. [2010]. *Bayesian Item Response Modeling, Theory and Applications*, Statistics for Social and Behavioral Sciences, fienberg, s. and van der linden, w. edn, Springer Science+Business Media, LLC.

[13] Fujimoto, K. [2018a]. The bayesian multilevel trifactor item response theory model, *Educational and Psychological Measurement* **79**(3): 462–494.
doi: https://doi.org/10.1177/0013164418806694.

[14] Fujimoto, K. [2018b]. A general bayesian multilevel multidimensional irt model for locally dependent data, *Br J Math Stat Psychol* **71**(3): 536–560.
doi: https://doi.org/10.1111/bmsp.12133.

[15] Fujimoto, K. [2020]. A more flexible bayesian multilevel bifactor item response theory model, *Journal of Educational Measurement* **57**(2): 255–285.
doi: https://doi.org/10.1111/jedm.12249.

[16] Gelfand, A., Sahu, S. and Carlin, B. [1995]. Efficient parametrisations for normal linear mixed models, *Biometrika* **82**(3): 479–488.
doi: https://doi.org/10.1093/biomet/82.3.479.

[17] Gelfand, A., Sahu, S. and Carlin, B. [1996]. Efficient parameterizations for generalised linear models (with discussion), *in* J. Bernardo, J. Berger, A. Dawid and a. Smith (eds), *Bayesian Statistics*, Vol. 5, pp. 165–180.

[18] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. [2014]. *Bayesian Data Analysis*, Texts in Statistical Science, third edn, Chapman and Hall/CRC.

[19] Gelman, A. and Rubin, D. [1996]. Markov chain monte carlo methods in biostatistics, *Statistical Methods in Medical Research* **5**(4): 339–355.
doi: https://doi.org/10.1177/096228029600500402.

[20] Gorinova, M., Moore, D. and Hoffman, M. [2019]. Automatic reparameterisation of probabilistic programs.
url: https://arxiv.org/abs/1906.03028.

[21] Hambleton, R., Swaminathan, H. and Rogers, H. [1991]. *Fundamentals of Item Response Theory*, SAGE Publications Inc.

[22] Heckman, J. [1979]. Sample selection bias as a specification error, **47**(1): 153–161.
doi: https://www.doi.org/10.2307/1912352.
url: https://www.jstor.org/stable/1912352.

[23] Hsieh, M., Proctor, T., Hou, J. and Teo, K. [2010]. A comparison of bayesian mcmc and marginal maximum likehood methods in estimating the item parameters for the 2pl irt model, *International Journal of Innovative Management, Information and Production* **1**(1): 81–89.
url: http://ismeip.org/IJIMIP/contents/imip1011/10IN15T.pdf.

[24] Jiao, H., Kamata, A., Wang, S. and Jin, Y. [2012]. A multilevel testlet model for dual local dependence, *Journal of Educational Measurement* **49**: 82–100.
doi: https://doi.org/10.1111/j.1745-3984.2011.00161.x.

[25] Keane, M. [1992]. A note on identification in the multinomial probit model, *Journal of Business and Economic Statistics* **10**(2): 193–200.
doi: https://doi.org/10.2307/1391677.
url: https://www.jstor.org/stable/1391677.

[26] Kim, S. and Cohen, A. [1999]. Accuracy of parameter estimation in gibbs sampling under the two-parameter logistic model, *Annual Meeting of the American Educational Research Association*, American Educational Research Association.
url: https://eric.ed.gov/?id=ED430012.

[27] Kline, R. [2012]. Assumptions in structural equation modeling, *in* R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, The Guilford Press, chapter 7, pp. 111–125.

[28] Martin, J. and McDonald, R. [1975]. Bayesian estimation in unrestricted factor analysis: A treatment for heywood cases, *Psychometrika* (40): 505–517.
doi: https://doi.org/10.1007/BF02291552.

[29] McCullagh, P. and Nelder, J. [1989]. *Generalized Linear Models*, Monographs on Statistics Applied Probability, Chapman Hall/CRC Press.

[30] McElreath, R. [2020]. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Texts in Statistical Science, 2 edn, Chapman and Hall/CRC.
doi: https://doi.org/10.1201/9780429029608.

[31] Muthén, L. and Muthén, B. [1998-2011]. *Mplus User's Guide*, CA: Muthén Muthén.

[32] Nelder, J. and Wedderburn, W. [1972]. Generalized linear models, *Royal Statistical Society* **135**(3): 370–384.
doi: https://doi.org/10.2307/2344614.
url: https://www.jstor.org/stable/2344614.

[33] Papaspiliopoulos, O., Roberts, G. and Skold, M. [2003]. Non-centered parameterisations for hierarchical models and data augmentation, *Bayesian Statistics* **7**: 307–326.
url: http://econ.upf.edu/ omiros/papers/val7.pdf.

[34] Papaspiliopoulos, O., Roberts, G. and Skold, M. [2007]. A general framework for the parametrization of hierarchical models, *Statistical Science* **22**(1): 59–73.
doi: https://www.doi.org/10.1214/088342307000000014.

[35] R Core Team [2015]. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
url: http://www.R-project.org/.

[36] Rabe-Hesketh, S., Pickles, A. and Skrondal, A. [2003]. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation, *Statistical Modelling* **3**(3): 215–232.
doi: https://www.doi.org/10.1191/1471082X03st056oa.

[37] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2003]. Maximum likelihood estimation of generalized linear models with covariate measurement error, *The Stata Journal* **3**(4): 386–411.
doi: https://www.doi.org/10.1177/1536867X0400300408.
url: https://journals.sagepub.com/doi/pdf/10.1177/1536867X0400300408.

[38] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004a]. Generalized multilevel structural equation modeling, *Psychometrika* **69**(2): 167–190.
doi: https://www.doi.org/10.1007/BF02295939.

[39] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004b]. *GLLAMM Manual*, UC Berkeley Division of Biostatistics.
url: http://www.biostat.jhsph.edu/ fdominic/teaching/bio656/software-gllamm.manual.pdf.

[40] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004c]. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, *Journal of Econometrics* **128**(2): 301–323.
doi: https://www.doi.org/10.1016/j.jeconom.2004.08.017.
url: http://www.sciencedirect.com/science/article/pii/S0304407604001599.

[41] Rabe-Hesketh, S., Skrondal, A. and Zheng, X. [2012]. Multilevel structural equation modeling, *in* R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, The Guilford Press, chapter 30, pp. 512–531.

[42] Rabe-Hesketh, S., Yang, S. and Pickles, A. [2001]. Multilevel models for censored and latent responses, *Statistical Methods in Medical Research* **10**(6): 409–427.
doi: https://www.doi.org/10.1177/096228020101000604.

[43] Raudenbush, S. and Bryk, A. [2002]. *Hierarchical linear models: Applications and data analysis methods (Vol. 1)*, Advanced Quantitative Techniques in the Social Sciences, SAGE Publications Inc.

[44] Reckase, M. [2009]. *Multidimensional Item Response Theory*, Statistics for Social and Behavioral Sciences, Springer Science+Business Media, LLC.

[45] Rivera, J. [2019]. *El modelo de respuesta nominal: Aplicación a datos educacionales*, Master's thesis, Pontificia Universidad Católica del Peru.
url: http://hdl.handle.net/20.500.12404/14600.

[46] Skrondal, A. and Rabe-Hesketh, S. [2003a]. Multilevel logistic regression for polytomous data and rankings, *Psychometrika* **68**: 267–287.
doi: https://www.doi.org/10.1007/BF02294801.

[47] Skrondal, A. and Rabe-Hesketh, S. [2003b]. Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling, *Norsk Epidemiologi* **13**(2): 265–278.

[48] Skrondal, A. and Rabe-Hesketh, S. [2004a]. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Interdisciplinary Statistics, Chapman Hall/CRC Press.

[49] Skrondal, A. and Rabe-Hesketh, S. [2004b]. Generalized linear latent and mixed models with composite links and exploded likelihoods, *in* BiggeriA., E. Dreassi, C. Lagazio and M. Marchi (eds), *Proceedings of the 19th International Workshop on Statistical Modeling*, Firenze University Press, Florence, Italy, pp. 27–39.
url: http://www.gllamm.org/composite$_c onf.pdf$.

[50] Stan Development Team [2020a]. RStan: the R interface to Stan. R package version 2.21.2.
url: http://mc-stan.org/.

[51] Stan Development Team. [2020b]. *Stan Modeling Language Users Guide and Reference Manual, version 2.26*, Vienna, Austria.
url: https://mc-stan.org.

[52] Tarazona, E. [2013]. *Modelos alternativos de respuesta graduada con aplicaciones en la calidad de servicios*, Master's thesis, Pontificia Universidad Católica del Perú (PUCP).
url: http://hdl.handle.net/20.500.12404/6175.

[53] Wainer, H., Bradlow, E. and Wang, X. [2007]. *Testlet response theory and its applications*, Cambridge University Press.

[54] Wollack, J. A., Bolt, D. M., Cohen, A. S. and Lee, Y.-S. [2002]. Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and markov chain monte carlo estimation, *Applied Psychological Measurement* **26**(3): 339–352.
doi: https://www.doi.org/10.1177/0146621602026003007.

[55] Wright, S. [1920]. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs, *Proceedings of the National Academy of Sciences* **6**(6): 320–332.
doi: https://doi.org/10.1073/pnas.6.6.320.

[56] Yen, W. [1984]. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model, *Applied Psychological Measurement* **8**(2): 125–145.
doi: https://doi.org/10.1177/014662168400800201.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be