

Outline

1. Introduction
 2. GLLAMM
dichotomous
 3. Bayesian
estimation
 4. Simulation
study
 5. Application
 6. Conclusions
- Appendix
- References

GLLAMM:
method, bayesian estimation, advantages, and
applications to educational data.

José Manuel Rivera Espejo

Master of Science in Statistics and Data Science

KU Leuven

Leuven, August 31, 2021

Document Outline

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

The document is organized as follows:

- ① Preliminary considerations
- ② The GLLAMM for dichotomous outcomes
- ③ Bayesian estimation
- ④ Simulation study
- ⑤ Application
- ⑥ Conclusions

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

1. Preliminary consideration

IRT local independence

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Comprised of two parts [1, 10]:

- local item independence
- local individual independence.

IRT models are **not robust** to the violation of local independence [28, 3, 11].

Educational data

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

often display **several** types of dependencies, violating the local item and/or individual independence, e.g.

- testlets [27];
- the measurement of multiple latent traits within individuals [24];
- cluster effects [23].

Proposed model

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

The GLLAMM follow a multilevel/hierarchical multidimensional approach to account for different dependencies.

- (**good**) control for dependencies in educational data
- (**important**) reach appropriate conclusion from the parameters

Computational implementation

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Under the Bayesian framework, but the integration will be:

- Complex and highly dimensional (in parameters)
- On sparse binary data

but complex parametrizations introduce pathologies that prevent MCMC methods to achieve ergodicity

[5, 6, 17, 18, 2], i.e. reach stationarity, convergence, and good mixing [15].

Computational implementation (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Four solutions are offered to solve the previous pathologies:

- Changing the settings of the MCMC method
 - ① increasing the number of iterations per chain, with large burn-in and thinning processes
 - ② designing model-specific MCMC algorithms.
- Readjusting the Bayesian model
 - ③ re-write the model in an alternative parametrization (**simple changes**)
 - ④ encode prior information through the prior distributions

Computational implementation (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

However, still no simple rotation/rescaling of the parameter, or the amount of data, allow to visit the posterior distribution properly [2].

Multiple authors showed that changing the **posterior sampling geometries**, i.e. removing the dependence of the parameters on other sampled parameters, **improves** the performance of the MCMC methods [5, 6, 17, 18, 2]

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

2. The GLAMM for dichotomous outcomes

Model definition

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Following Rabe-Hesketh et al. [20, 21], we define the GLLAMM in two parts:

- ① the response model
- ② the latent structure

Moreover, **the response model (1)** can be represented by a Generalized Linear Model (GLM) [16, 14] with:

- ① a distributional part
- ② a systematic part

The response model

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Conditional to all parameters $\Omega = \{\beta, \Lambda, \Theta, \Psi, \Gamma\}$; and the “stacked” vector of covariates \mathbf{X} and \mathbf{W} ; **the distributional part** is defined by:

$$f(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \Omega) = \pi_{jkd}^n (1 - \pi_{jkd})^{1-n} \quad (1)$$

Furthermore, **the systematic part** is defined in the following form:

$$P(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \Omega) = \pi_{jkd} = h(\tau_k + v_{jkd}) \quad (2)$$

where τ_k is k 'th item threshold, assumed to be zero for the binary case [20].

The response model (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Moreover, the inverse-link function $h(\cdot)$ can be defined in three ways:

$$h(x) = \begin{cases} \exp(x)[1 + \exp(x)]^{-1} \\ \Phi(x) \\ \exp(-\exp(x)) \end{cases} \quad (3)$$

corresponding to the logistic, standard normal $\Phi(x)$, and Gumbel (extreme value type I) cumulative distributions, respectively.

The response model (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Finally, the linear predictor is defined by:

$$v_{jkd} = \sum_{p=1}^P x_{jp} \beta_p + \sum_{m=2}^{M+1} \sum_{k=1}^{K_{(m)}} \eta_k^{(m)} \alpha_k^{(m)} + \sum_{l=2}^{L+1} \sum_{d=1}^{D_{(l)}} \theta_{jd}^{(l)} \lambda_d^{(l)}$$
 (4)

Which after the appropriate “stacking”:

$$\begin{aligned} v_{jkd} &= \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\eta} \boldsymbol{\alpha} \mathbf{A}_j + \boldsymbol{\theta} \boldsymbol{\lambda} \mathbf{B}_j \\ &= \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\Theta} \boldsymbol{\Lambda} \mathbf{H}_j \end{aligned}$$
 (5)

The latent structure

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

The structural model for the latent variables is represented in the following form:

$$\boldsymbol{\Theta} = \boldsymbol{\Psi}_{(S \times S)(S \times 1)} \boldsymbol{\Theta} + \boldsymbol{\Gamma}_{(S \times Q)(Q \times 1)} \mathbf{W} + \boldsymbol{\zeta}_{(S \times 1)} \quad (6)$$

where $S = K + D$, $K = \sum_m K_m$, and $D = \sum_l D_l$.

Notice equation (6) is the generalization of a single-level Structural Equation Models (SEM) to a multilevel setting.

Motivating example

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

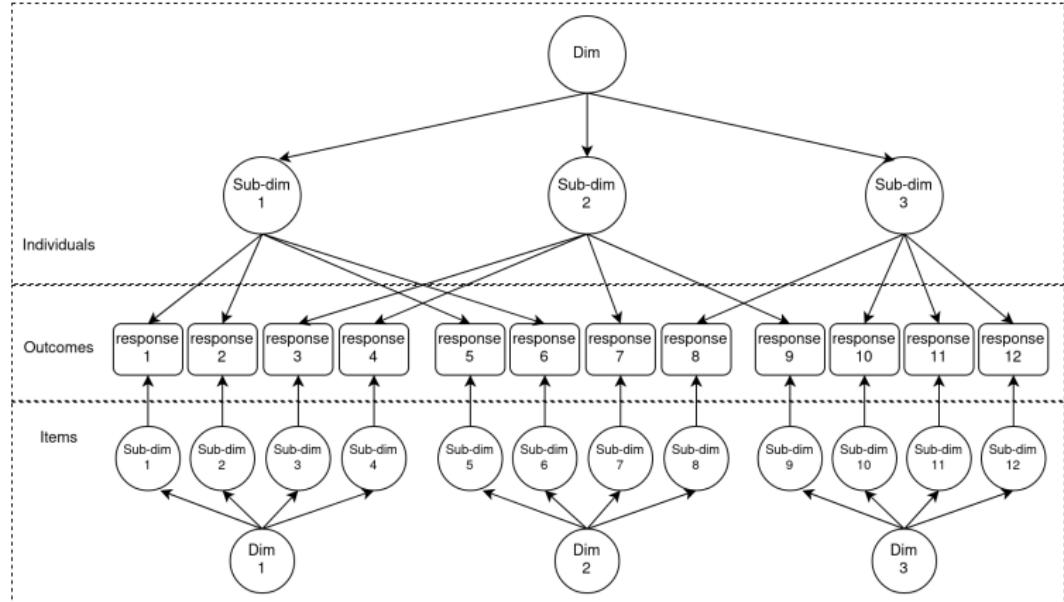
4. Simulation study

5. Application

6. Conclusions

Appendix

References



Motivating example (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- Empty level-1 covariates matrix \mathbf{X}_j ($P = 0$)
- $M = 2$ levels at the items block, with $K_2 = 12$ and $K_3 = 3$, i.e. $\boldsymbol{\eta}^{(2)} = [\eta_1^{(2)}, \dots, \eta_{12}^{(2)}]^T$ and $\boldsymbol{\eta}^{(3)} = [\eta_1^{(3)}, \eta_2^{(3)}, \eta_3^{(3)}]^T$
- $L = 2$ levels in the individuals block, with $D_2 = 3$ and $D_3 = 1$, i.e. $\boldsymbol{\theta}^{(2)} = [\theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)}]^T$ and $\boldsymbol{\theta}^{(3)} = \theta_1^{(3)}$.
- Specific regression relationship among latents $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\alpha}^{(3)} = [\alpha_{11}^{(3)}, \dots, \alpha_{15}^{(3)}, \alpha_{21}^{(3)}, \dots, \alpha_{25}^{(3)}, \alpha_{31}^{(3)}, \dots, \alpha_{35}^{(3)}]^T$ and $\boldsymbol{\lambda}^{(3)} = [\lambda_1^{(3)}, \lambda_2^{(3)}, \lambda_3^{(3)}]^T$
- Empty structural covariates \mathbf{W} .

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

3. Bayesian Estimation

Bayesian GLLAMM for dichotomous outcomes

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① **Posterior distribution.** Given that \mathbf{Y} is the observed data and $\Omega = \{\beta, \Lambda, \Theta, \Psi, \Gamma\}$ the parameters:

$$P(\Omega | \mathbf{Y}) = \frac{P(\mathbf{Y} | \Omega) P(\Omega)}{\int P(\mathbf{Y} | \Omega) P(\Omega) d\Omega} \quad (7)$$

- ② **Prior distributions.** Similar to Patz and Junker [19], we use an independent distributional structure for the joint priors:

$$\begin{aligned} P(\Omega) = & P(\beta) [P(\alpha) P(\lambda)] [P(\eta) P(\theta)] \\ & [P(\Psi_\eta) P(\Psi_\theta)] [P(\Gamma_\eta) P(\Gamma_\theta)] \end{aligned} \quad (8)$$

Bayesian GLLAMM for dichotomous outcomes (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

③ **Likelihood.** Following Rabe-Hesketh et al. [20], the likelihood function is build in a recursive way.

$$f(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \pi_{jkd}^n (1 - \pi_{jkd})^{1-n} \quad (9)$$

$$f(\mathbf{y} = \mathbf{1} | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \prod_{j=1}^J \prod_{d=1}^D \prod_{k=1}^K f(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (10)$$

$$f_{(m)}^{(l)} (\mathbf{y} = \mathbf{1} | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \int \left[\prod f_{(m-1)}^{(l-1)} (\mathbf{y} = \mathbf{1} | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \right] P(\boldsymbol{\Theta}_{(m)}^{(l)}) \quad (11)$$

$$\mathcal{L}(\mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \prod_{m=2}^{M+1} \prod_{l=2}^{L+1} f_{(m)}^{(l)} (\mathbf{y} = \mathbf{1} | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (12)$$

$$\ell(\mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \log \mathcal{L}(\mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (13)$$

Computational implementation

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ➊ Hamiltonian Monte Carlo (HMC) and Stan [26].
- ➋ **No burn-in and thinning.** Warm-up phase to “tune-up” the number of steps (leapfrogs), and the step size [26].
- ➌ We use 3,000 **effective iterations:** 3 chains of 2,000 iterations each, where 1,000 are spent on warm-up.
- ➍ **Initial starts** sampled from the priors defined in the model.
- ➎ Prior distributions selected based on **prior predictive simulations.**
- ➏ We test the **centered (CP)** and **non-centered parametrizations (NCP).**

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

4. Simulation studies

Objectives

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① **Performance.** In terms of achieving ergodicity, under the CP and NCP,
- ② **Recovery capacity.** Capacity to recover the parameters of interest, especially the structural regression parameters.
- ③ **Retrodictive accuracy.** Capacity to retrodict the data of interest, according to a set of aggregating dimensions.

Evaluation criteria

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① **Performance.** Trace, trunk and ACF plots with support of Rhat and n_eff statistics developed by Gelman et al. [7] (pp. 284 – 287).
- ② **Recovery capacity.** we used the between replica root mean squared error (RMSE_B).
- ③ **Retrodictive accuracy.** we used the average within $\overline{\text{RMSE}}_W$ and between prediction root mean squared error RMSE_B , of the responses' predictive proportion \hat{p} , versus the observed proportion p .

Results (Performance)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① the non-centered parametrization (NCP) largely improved the performance of the MCMC chains, towards achieving ergodicity.

This is true across models, simulated sample sizes, and simulated replicas.

- ② No large difference in performance was observed in either the sub-dimensions' correlation or loading parameters.

However, no evidence supported the idea the parameters suffered from a further lack of identification.

Results (Performance, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

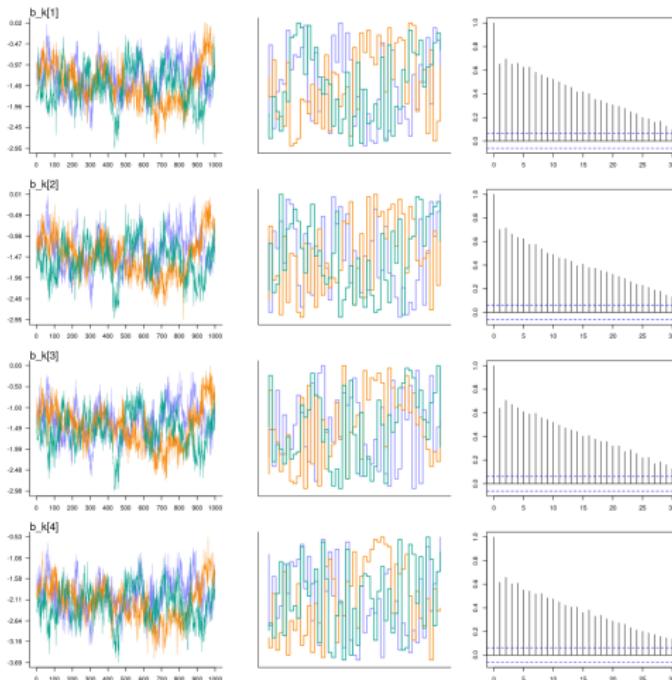
4. Simulation
study

5. Application

6. Conclusions

Appendix

References



Results (Performance, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

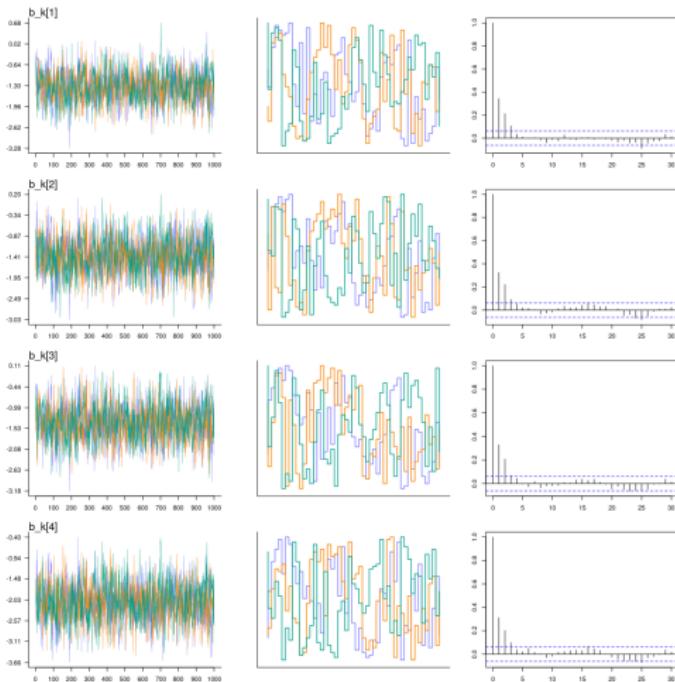
4. Simulation
study

5. Application

6. Conclusions

Appendix

References



Results (Recovery capacity)

Outline

- 1. Introduction
 - 2. GLLAMM dichotomous
 - 3. Bayesian estimation
 - 4. Simulation study
 - 5. Application
 - 6. Conclusions
- Appendix
- References

- ① The GLLAMM was able to recover most of the simulated parameters with good precision.
 - ② However, the model still had issues estimating the sub-dimensions correlations and loadings.
- Under the Confirmatory Factor Analysis theory (CFA), a SOLV model is only justified, if the lower-level correlations are high enough, usually above 0.8.
- ③ It is surprising that CP and NCP achieved similar levels of recovery capacity.

Results

(Recovery capacity, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

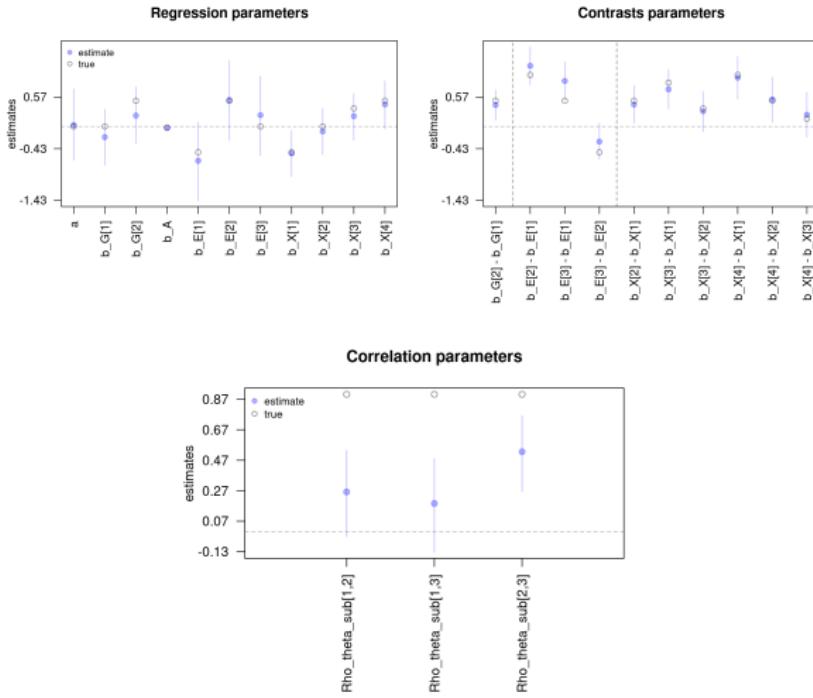
4. Simulation study

5. Application

6. Conclusions

Appendix

References



Results (Retrodictive accuracy)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① the models managed to capture the traits of the data, while avoiding its exact replication.

These results were consistent across models, simulated sample sizes and replicates.

Results

(Retrodictive accuracy, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

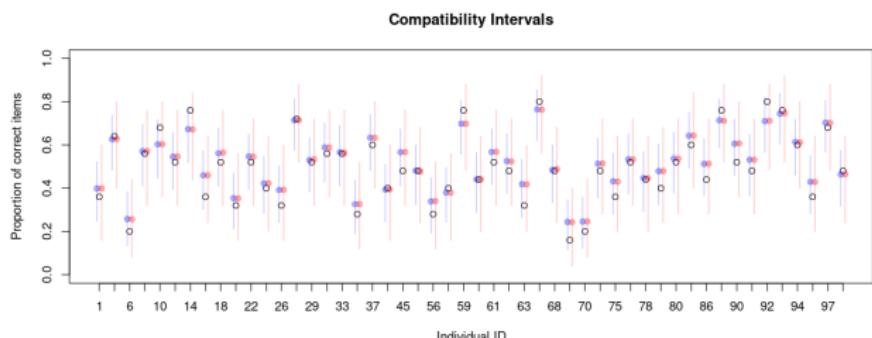
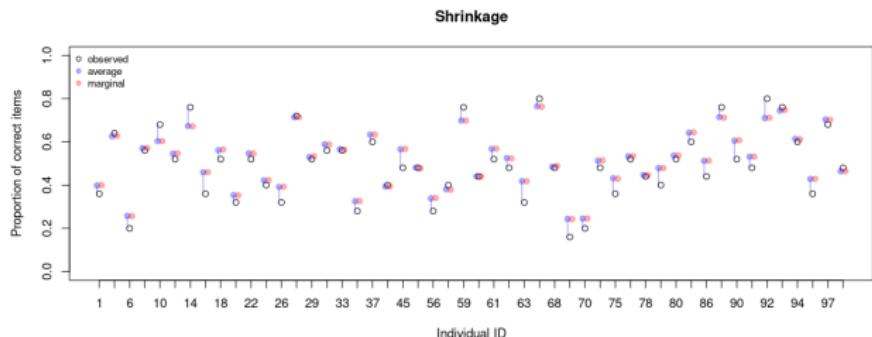
4. Simulation
study

5. Application

6. Conclusions

Appendix

References



Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

5. Application

Objectives

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① **Evaluate the performance of the parametrizations.** Ergodicity of CP vs NCP.
- ② **Evaluate the retrodictive accuracy.**
- ③ **Assess the psychometric properties.** Special interest in determine how difficult the items were.
- ④ **Test research hypothesis.** About the explanatory power a set of covariates had on the latent dimensions, and their implications for the educational authority.

Instrument and data

Outline

1. Introduction
 2. GLLAMM dichotomous
 3. Bayesian estimation
 4. Simulation study
 5. Application
 6. Conclusions
- Appendix
- References

- ➊ **Instrument.** Reading comprehension sub-test composed of 25 binary scored items, from the Peruvian public teaching career national assessment.
- ➋ **Access.** Legal requirement of open information to the Ministry of Education of Peru (MINEDU).
- ➌ **Sample scheme.** Simple random sample of 2,000 (from approx. 195,0000).

Results (Performance)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① the non-centered parametrization (NCP) largely improved the performance of the MCMC chains, towards achieving ergodicity.
- ② No large difference in performance was observed in either the sub-dimensions' correlation or loading parameters.

results are similar to simulation study, albeit in some cases more extreme.

Results (Performance, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

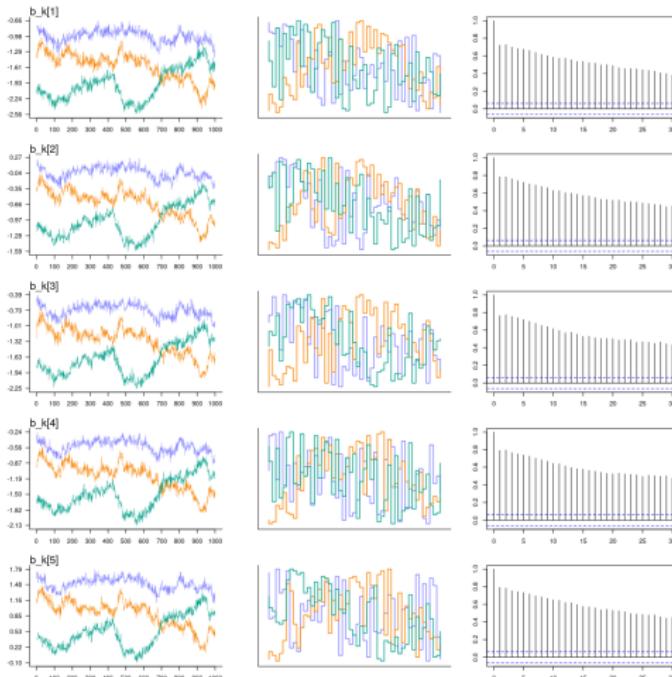
4. Simulation
study

5. Application

6. Conclusions

Appendix

References



Results (Performance, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

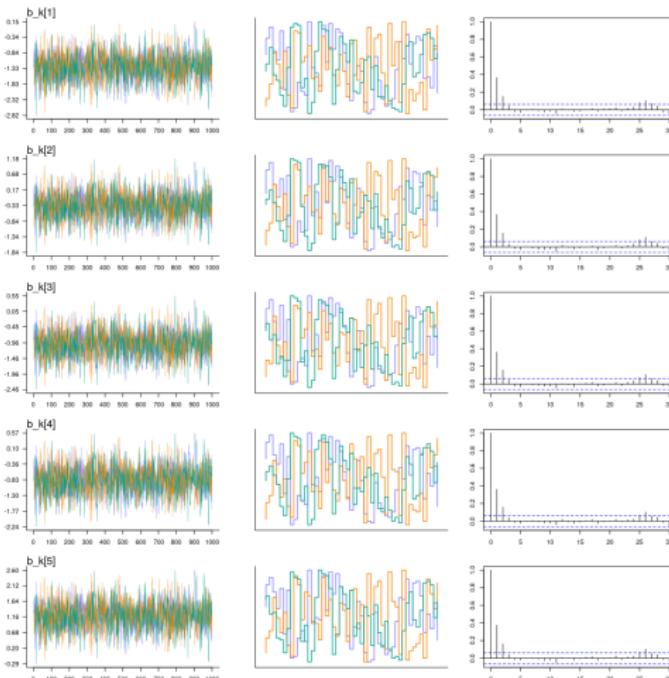
4. Simulation study

5. Application

6. Conclusions

Appendix

References



Results (Retrodictive accuracy)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① the models managed to capture the traits of the data, while avoiding its exact replication.

Similar to simulation study.

Results

(Retrodictive accuracy, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

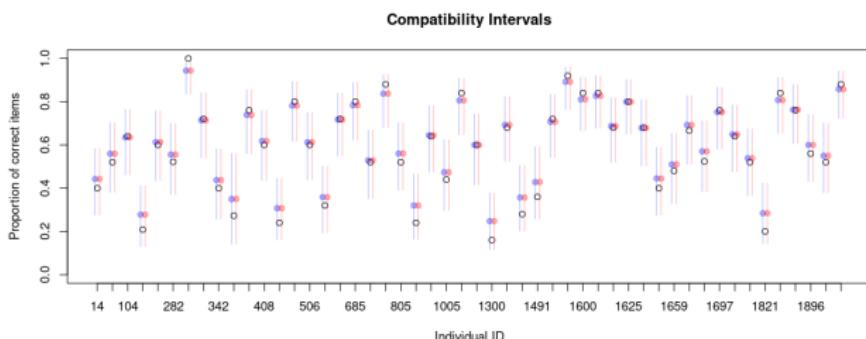
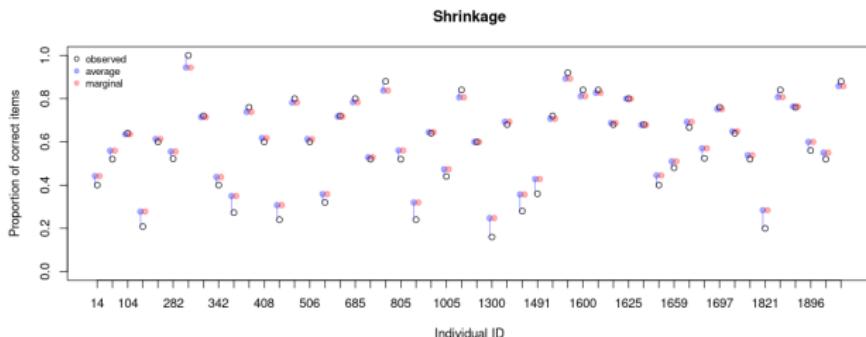
4. Simulation study

5. Application

6. Conclusions

Appendix

References



Results (Psychometric properties)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① the items were scattered throughout a significant portion of the abilities range.
- ② Specific benefit of the implementation allowed us to assess how difficult the texts were.
- ③ A more sound psychometric analysis requires access to the items and texts description, but it was not possible due to legal restrictions.

Results

(Psychometric properties, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

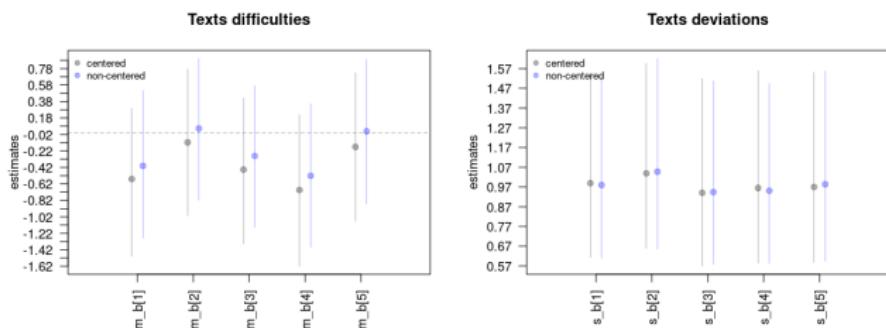
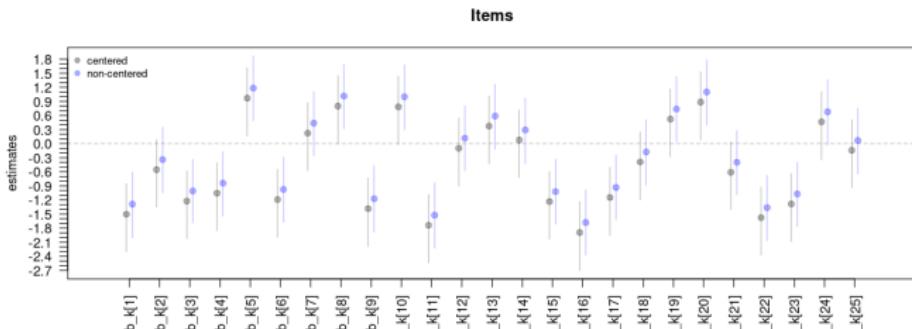
4. Simulation study

5. Application

6. Conclusions

Appendix

References



Results (Hypothesis testing)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

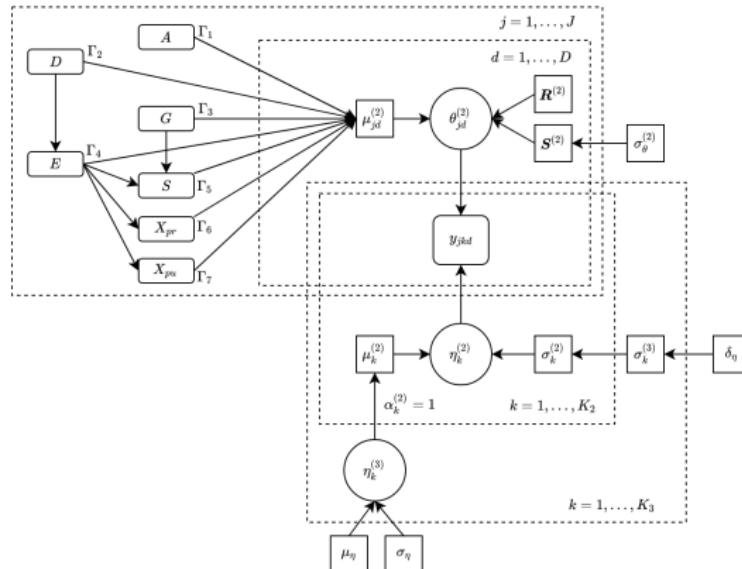


Figure: Directed Acyclic Graph (DAG). Application's first-order latent variable model (FOLV).

Results

(Hypothesis testing, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- ① Age (style of teaching proxy) explains negatively reading comprehension $-0.036[-0.04, -0.03]$.
- ② Disability also explained the reading comprehension sub-dimensions, although the results were mildly unexpected.
- ③ the statistical evidence seem to support the lower quality of training on pedagogical institutes.
- ④ Experience improved the reading comprehension abilities. Private experience effects were larger than the public. We observe also diminishing returns on abilities, as the years of experience increases.
- ⑤ Instructing students from secondary education improves reading comprehension.

Results

(Hypothesis testing, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

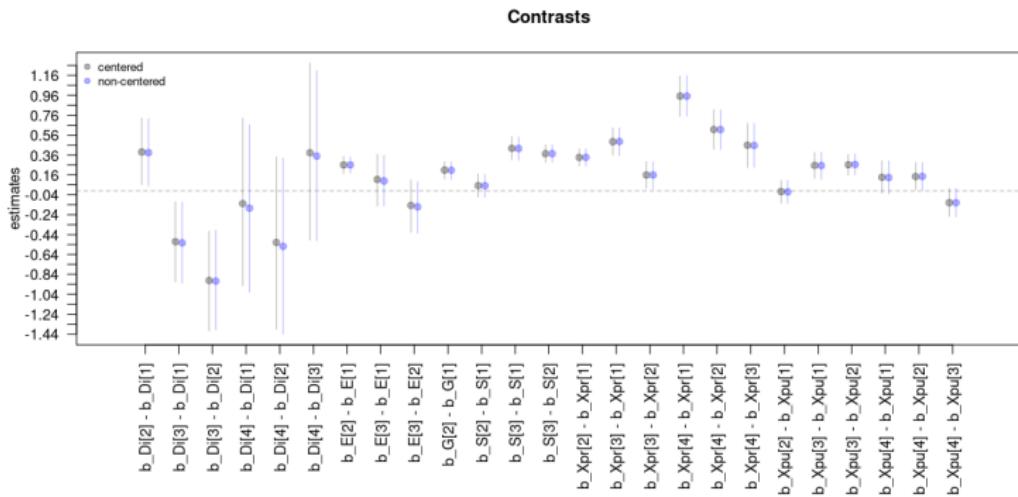


Figure: Application's first-order latent variable model. CP and NCP comparison plot.

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

6. Conclusions and further development

Conclusions

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- ➊ The NCP largely improved the performance of the MCMC chains, towards achieving ergodicity. True across models, simulated sample sizes, simulated replicas, and even under the real application, albeit with some caveats.
- ➋ Our proposed model was able to recover most of the simulated parameters with good precision. The model still had issues estimating the sub-dimensions correlations and loadings.
- ➌ The models managed to capture the traits of the data, while avoiding its exact replication. Consistent across models, simulated sample sizes and replicas, and even under the final educational application.

Conclusions (cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- ➊ The NCP was slightly faster than the CP, although the magnitudes of the differences in running time were not large.
- ➋ On the application, the model provided an extra benefit, to asses the psychometric properties of texts.
- ➌ In relation to our hypothesis of interest, the model produced sound statistical results, supported not only by the statistical application, but also from a DAG guiding our interpretation and causal assumptions.

Further investigation

Outline

1. Introduction
 2. GLLAMM dichotomous
 3. Bayesian estimation
 4. Simulation study
 5. Application
 6. Conclusions
- Appendix
- References

- ① Why the benefits of the NCP did not fully extend to correlations and loadings?
- ② Test other simulation scenarios to find why the CP (no ergodicity) managed recover the parameters as well as the NCP.
- ③ Related to the previous, test results for Variational Inference methods (VI)
- ④ The statistical evidence remains valid with a better sample scheme?
- ⑤ Test the model also with response explanatory variables, e.g. response time, number of alternatives, etc.
- ⑥ Test the model for measurement invariance.

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Appendix

Cluster effects

Outline

1. Introduction
 2. GLLAMM dichotomous
 3. Bayesian estimation
 4. Simulation study
 5. Application
 6. Conclusions
- Appendix
- References

Individual clustering involves the addition of more random effects to the linear predictor defined in equation (4):

$$\begin{aligned} v_{jkd} &= v_{jkd} + \sum_{c=1}^C \delta_c \\ &= v_{jkd} + \boldsymbol{\delta Z}_j \end{aligned} \tag{14}$$

Example (restrictions for IRT)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

we could set the restriction $\alpha^{(2)} = -\lambda^{(2)}$ where $\lambda^{(2)} > 0$. In that case we get a multidimensional generalization of the linear predictor observed in the archetypical Rasch [22], or 2PL [13] models, i.e. $\lambda_d^{(2)}(\theta_{jd}^{(2)} - \eta_k^{(2)})$

In addition,

$$\begin{aligned}\boldsymbol{\alpha}^{(3)} &= [\alpha_{11}^{(3)}, \dots, \alpha_{15}^{(3)}, \alpha_{21}^{(3)}, \dots, \alpha_{25}^{(3)}, \alpha_{31}^{(3)}, \dots, \alpha_{35}^{(3)}]^T \\ &= [1, \dots, 1]^T, \text{ indicating texts difficulties explain directly}\\ &\quad \text{the items difficulties at the lower level.}\end{aligned}$$

Example (restrictions for IRT)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Moreover, notice that because in the IRT framework η and θ should be orthogonal to each other by design, we can further decompose equation (6) in the following form:

$$\eta = \underset{(K \times K)(K \times 1)}{\Psi_\eta} \eta + \underset{(K \times Q)(Q \times 1)}{\Gamma_\eta} W_\eta + \underset{(K \times 1)}{\zeta_\eta} \quad (15)$$

$$\theta = \underset{(D \times S)(D \times 1)}{\Psi_\theta} \theta + \underset{(D \times Q)(Q \times 1)}{\Gamma_\theta} W_\theta + \underset{(D \times 1)}{\zeta_\theta} \quad (16)$$

Model Assumptions

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Following Skrondal and Rabe-Hesketh [25], the framework has two main assumptions:

(M1) Complete latent space.[9] In the GLLAMM representation, the space is complete if we consider all latent variables Θ at levels $l > 1$ and $m > 1$.

(M2) Local Independence. It assumes independence conditional on all the latent dimensions and covariates, at different hierarchical levels; effectively modeling all the observed dependencies.

Model Assumptions (cont.)

Outline

1. Introduction
2. GLLAMM
dichotomous
3. Bayesian
estimation
4. Simulation
study
5. Application
6. Conclusions

Appendix

References

Local Independence is defined as follows:

$$f(\mathbf{y} = \mathbf{1} \mid \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \prod_{j=1}^J \prod_{d=1}^D \prod_{k=1}^K f(y_{jkd} = 1 \mid \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (17)$$

Model Assumptions (cont.)

Outline

1. Introduction
2. GLLAMM
dichotomous
3. Bayesian
estimation
4. Simulation
study
5. Application
6. Conclusions

Appendix

References

but comes from:

① Local item independence,

$$f(y_{j..} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \prod_{d=1}^D \prod_{k=1}^K f(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (18)$$

② Local individual independence,

$$f(y_{.kd} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) = \prod_{j=1}^J f(y_{jkd} = 1 | \mathbf{X}, \mathbf{W}, \boldsymbol{\Omega}) \quad (19)$$

Why bayesian?

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- It is built on a simulation-based estimation method, therefore, it can handle all kinds of priors and data-generating processes [4].
- While the likelihood for the data and priors for the parameters are used to define the posterior sampling distributions, they can also be used in a generative way [15].
- Because the procedure integrates prior knowledge about the parameters, it can produce results even in scenarios where the Maximum Likelihood methods (ML) have issues of non-convergence or improper estimation [25, 4, 15]

There is nothing wrong?

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- It exposes the user to somewhat-arbitrary decisions about the running of the chains, in order to ensure a proper performance (**solution:** Hamiltonian Monte Carlo (HMC) [2]).
- The user can include all type of information through the priors distributions, making their elicitation convenient for manipulation (**solution:** prior predictive simulations and/or sensitivity analysis).
- Visual evaluation of performance, making it hard to assess if a proper posterior investigation have been made [8] (**solution:** help of Rhat, n_eff, and change of posterior sampling geometry).

There is nothing wrong? (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- The procedure makes it hard to discover parameters' lack of identification [25] (**solution:** regularizing priors).
- Oftentimes the posterior sampling geometry of the model makes it hard to find proper solutions for the parameter space [2] (**solution:** change the posterior sampling geometry).
- The greater the complexity of the model, the harder it is to communicate/share and takes more time (**solution:** No solution, but it is a small “price” to pay).

Prior elicitation

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

Un-informative priors are not the solution:

- ① Uninformative / weakly-informative latent prior:

$$\theta \sim N(0, 100) \quad \theta \sim N(0, 1)$$
$$\text{logit}(p) = \theta \quad \text{logit}(p) = \theta$$

- ② Uninformative / weakly-informative hierarchical latent prior:

$$v \sim \log N(0, 3) \quad v \sim \log N(0, 0.5)$$
$$\theta \sim N(0, v) \quad \theta \sim N(0, v)$$
$$\text{logit}(p) = \theta \quad \text{logit}(p) = \theta$$

Prior elicitation (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Un-informative priors are not the solution:

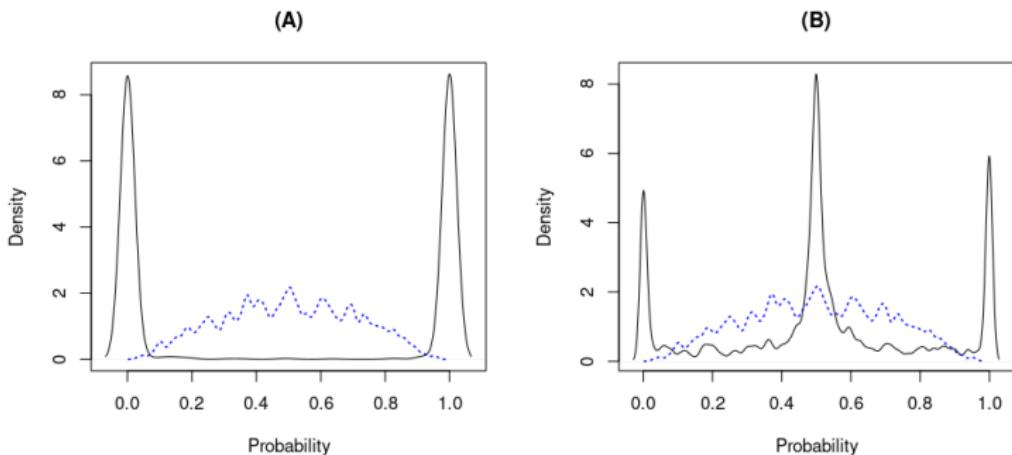


Figure: Prior predictive simulation. Examples of uninformative and mildly informative priors.

To center or not to center

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

Even the most simple hierarchical models present formidable pathologies, that no simple rotation/rescaling of the parameter can be performed to visit the posterior distribution properly [2].

Example, the devil's funnel [15]:

$$\begin{aligned} v &\sim N(0, 3) \\ \theta &\sim N(0, \exp(v)) \end{aligned} \tag{20}$$

Equation (20) describes a centered parametrization (CP)

To center or not to center (cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

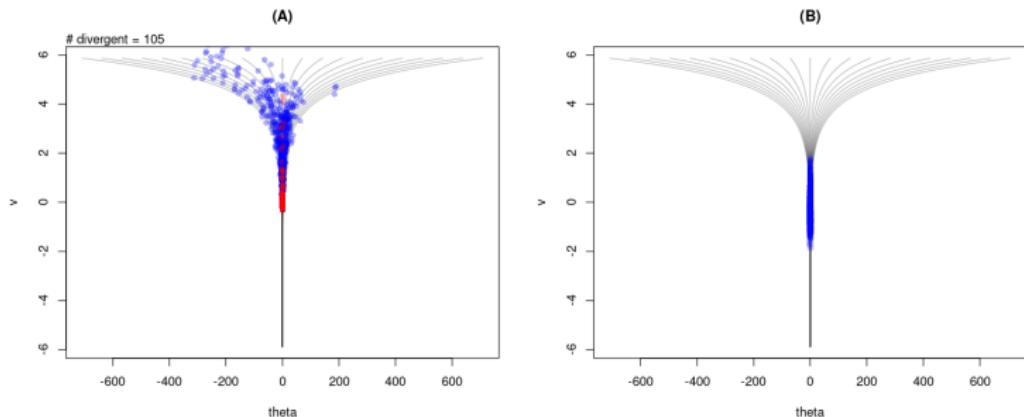


Figure: Posterior sampling geometry. Centered Parametrization.

To center or not to center (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

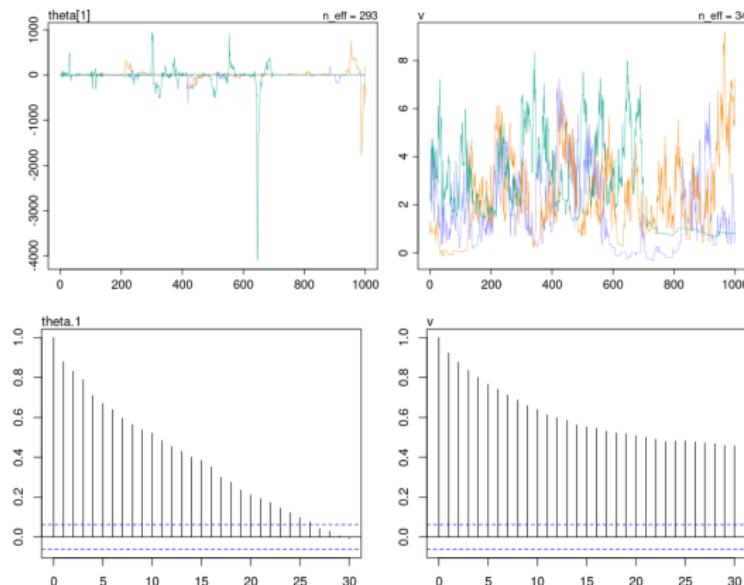


Figure: The Devil's funnel. Centered Parametrization. Trace and AFC plots.

To center or not to center (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

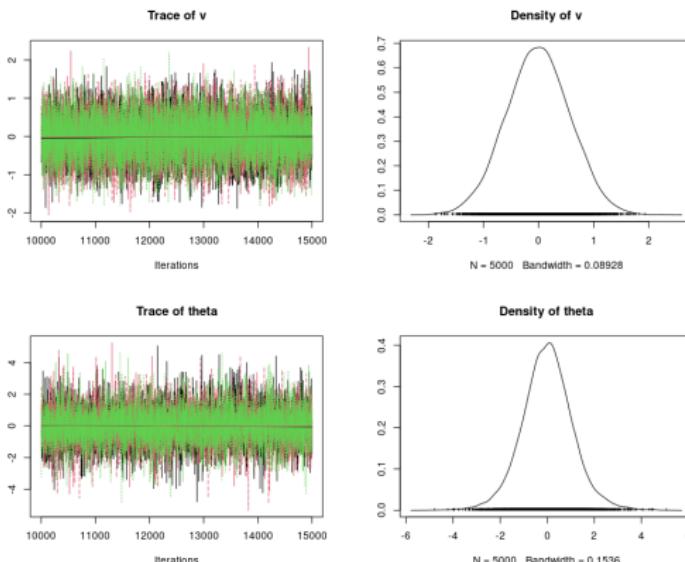


Figure: The Devil's funnel. Centered Parametrization implemented in JAGS.

To center or not to center (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

How can we solve this?:

- ① Adapt HMC warm-up (`adapt_delta= 0.99`).
- ② Use **regularizing priors**
- ③ Use the non-centered parametrization.

To center or not to center (regularizing priors)

Outline

1. Introduction
2. GLLAMM
dichotomous
3. Bayesian
estimation
4. Simulation
study
5. Application
6. Conclusions

Appendix

References

Consider the use of a regularizing prior on equation (20):

$$\begin{aligned} v &\sim N(0, 1) \\ \theta &\sim N(0, \exp(v)) \end{aligned} \tag{21}$$

To center or not to center (regularizing priors, cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

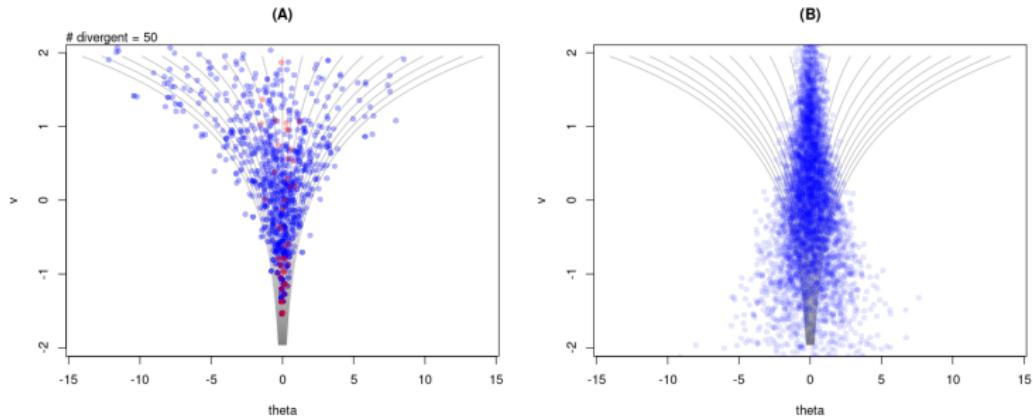


Figure: Posterior sampling geometry. Centered Parametrization with mildly informative priors.

To center or not to center (regularizing priors, cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

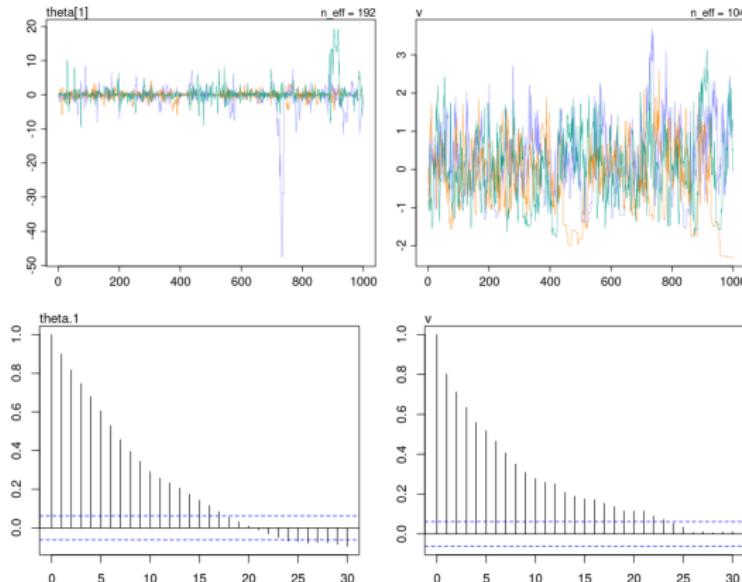


Figure: The Devil's funnel. Centered Parametrization with mildly informative priors. Trace and AFC plots.

To center or not to center (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

How can we solve this?:

- ① Adapt HMC warm-up (`adapt_delta= 0.99`).
- ② Use regularizing priors
- ③ Use the **non-centered parametrization**.

To center or not to center (non-centered parametrization)

Outline

1. Introduction
 2. GLLAMM dichotomous
 3. Bayesian estimation
 4. Simulation study
 5. Application
 6. Conclusions
- Appendix
- References

Changing the posterior sampling geometry means to modify equation (20) in the following way:

$$\begin{aligned} v &\sim N(0, 3) \\ z &\sim N(0, 1) \\ \theta &= \exp(v) z \end{aligned} \tag{22}$$

To center or not to center (non-centered parametrization, cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

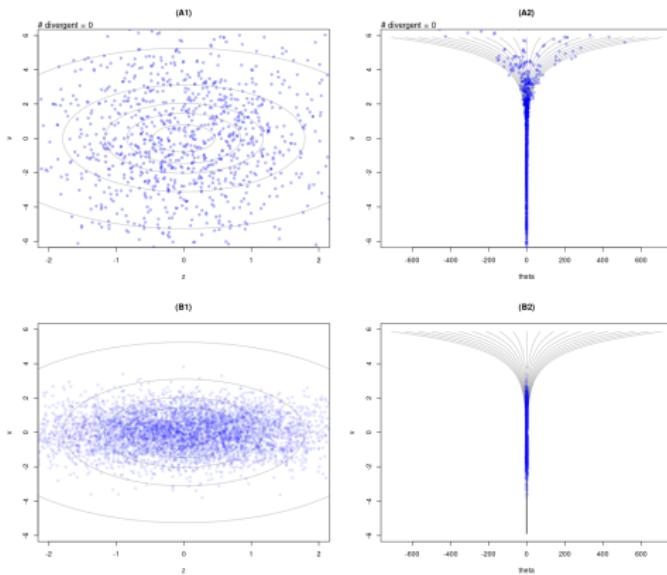


Figure: Posterior sampling geometry. Non-Centered Parametrization.

To center or not to center (non-centered parametrization, cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

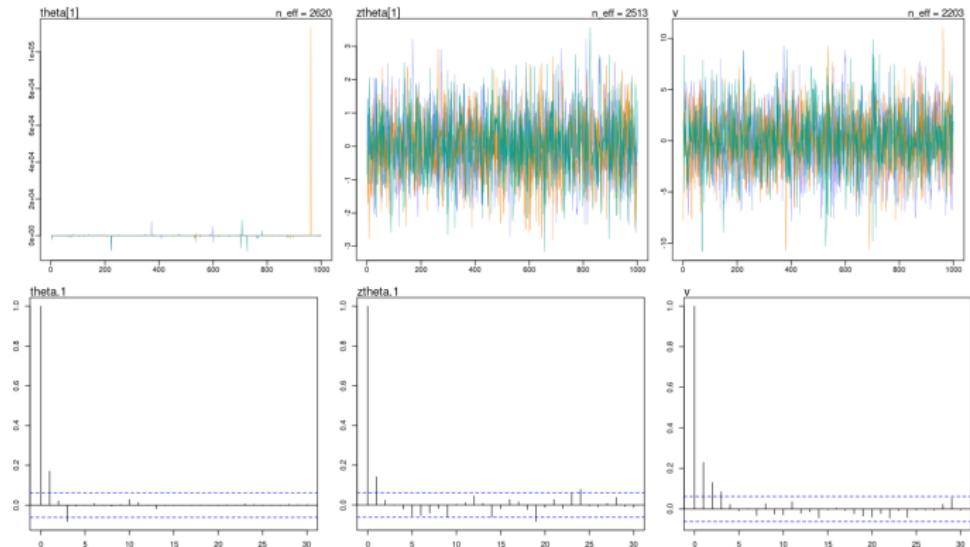


Figure: The Devil's funnel. Non-centered Parametrization. Trace and AFC plots.

Simulation study design

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

A $3 \times 2 \times 2$ fractional factorial design:

- Three different samples sizes to generate the data under analysis: 500, 250, and 100.
- Two parametrization of the models: CP and NCP.
- Two models of interest: the first- and second-order latent variable model.

Ten (10) data sets were generated for each study condition. Each data set resembled responses to 25 binary scored items, conforming to the SOLV model defined in figure 12. The model was motivated by the hypothesized structure of the reading comprehension sub-test, from the Peruvian public teaching career national assessment

Simulation study design (cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

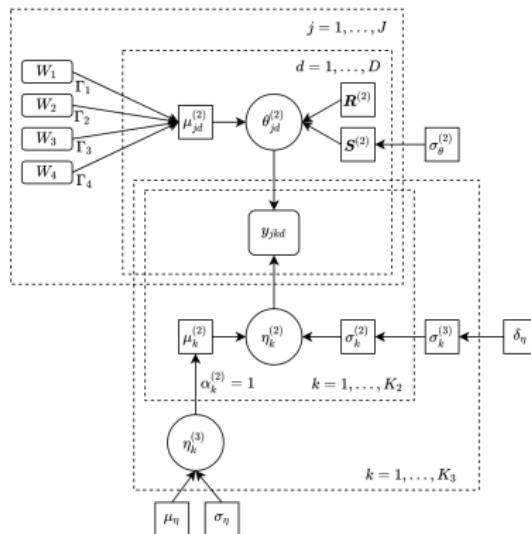


Figure: Directed Acyclic Graph (DAG). First-order latent variable model (SOLV).

Simulation study design (cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

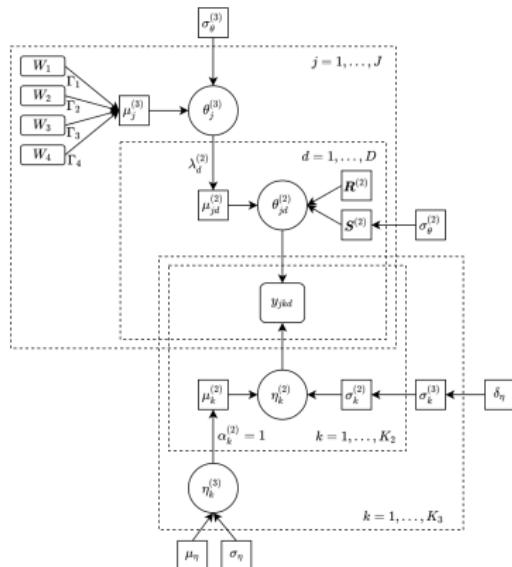


Figure: Directed Acyclic Graph (DAG). Second-order latent variable model (SOLV).

Likelihood, priors and hyper-priors (centered parametrization)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

$$y_{jkd} \sim \text{Bernoulli}(\pi_{jkd}) \quad (23)$$

$$\text{logit}(\pi_{jkd}) = v_{jkd} \quad (24)$$

$$v_{jkd} = \theta_{jd}^{(2)} - \eta_k^{(2)} \quad (25)$$

$$\boldsymbol{\theta}_j^{(2)} = [\theta_{j1}^{(2)}, \theta_{j2}^{(2)}, \theta_{j3}^{(2)}] \quad (26)$$

$$\boldsymbol{\theta}_j^{(2)} \sim \text{MVNormal} \left(\boldsymbol{\mu}_j^{(2)}, \boldsymbol{\Sigma}^{(2)} \right) \quad (27)$$

$$\boldsymbol{\Sigma}^{(2)} = \mathbf{S}^{(2)} \cdot \mathbf{R}^{(2)} \cdot \mathbf{S}^{(2)} \quad (28)$$

$$\mathbf{S}^{(2)} = \boldsymbol{\sigma}_{\theta}^{(2)} \mathbf{I} \quad (29)$$

Likelihood, priors and hyper-priors (centered parametrization, cont.)

Outline

1. Introduction
 2. GLLAMM
dichotomous
 3. Bayesian
estimation
 4. Simulation
study
 5. Application
 6. Conclusions
- Appendix
- References

For the FOLV model:

$$\boldsymbol{\mu}_j^{(2)} = \left[\mu_{j1}^{(2)}, \mu_{j2}^{(2)}, \mu_{j3}^{(2)} \right] \quad (30)$$

$$\mu_{jd}^{(2)} = \Gamma_0 + \Gamma_1 W_{1j} + \Gamma_2 (W_{2j} - W_{2\min}) + \Gamma_3 W_{3j} + \Gamma_4 W_{4j} \quad (31)$$

Likelihood, priors and hyper-priors (centered parametrization, cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

For the SOLV model:

$$\boldsymbol{\mu}_j^{(2)} = \left[\mu_{j1}^{(2)}, \mu_{j2}^{(2)}, \mu_{j3}^{(2)} \right] \quad (32)$$

$$\boldsymbol{\lambda}^{(2)} = \left[\lambda_1^{(2)}, \lambda_2^{(2)}, \lambda_3^{(2)} \right] \quad (33)$$

$$\mu_{jd}^{(2)} = \lambda_d^{(2)} \theta_j^{(3)} \quad (34)$$

$$\theta_j^{(3)} \sim \text{Normal} \left(\mu_j^{(3)}, \sigma_\theta^{(3)} \right) \quad (35)$$

$$\mu_j^{(3)} = \Gamma_0 + \Gamma_1 W_{1j} + \Gamma_2 (W_{2j} - W_{2\min}) + \Gamma_3 W_{3j} + \Gamma_4 W_{4j} \quad (36)$$

Likelihood, priors and hyper-priors (centered parametrization, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

For the items:

$$\eta_k^{(2)} \sim \text{Normal} \left(\mu_k^{(2)}, \sigma_k^{(2)} \right) \quad (37)$$

$$\mu_k^{(2)} = \boldsymbol{\eta}^{(3)} \mathbf{A} \quad (38)$$

$$\sigma_k^{(2)} = \boldsymbol{\sigma}^{(3)} \mathbf{A} \quad (39)$$

$$\boldsymbol{\eta}^{(3)} = [\eta_1^{(3)}, \eta_2^{(3)}, \eta_3^{(3)}, \eta_4^{(3)}, \eta_5^{(3)}] \quad (40)$$

$$\boldsymbol{\sigma}^{(3)} = [\sigma_1^{(3)}, \sigma_2^{(3)}, \sigma_3^{(3)}, \sigma_4^{(3)}, \sigma_5^{(3)}] \quad (41)$$

$$\eta_k^{(3)} \sim \text{Normal} (\mu_\eta, \sigma_\eta) \quad (42)$$

$$\sigma_k^{(3)} \sim \text{Exponential} (\delta_\eta) \quad (43)$$

Likelihood, priors and hyper-priors (centered parametrization, cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Remaining priors and hyper-priors

$$\mathbf{R}^{(2)} \sim \text{LkjCorrelation}(2) \quad (44)$$

$$\Gamma_{1c} \sim \text{Normal}(0, 0.5) \quad (45)$$

$$\Gamma_2 \sim \text{Normal}(0, 0.5) \quad (46)$$

$$\Gamma_{3c} \sim \text{Normal}(0, 1) \quad (47)$$

$$\Gamma_{4c} \sim \text{Normal}(0, 0.5) \quad (48)$$

Likelihood, priors and hyper-priors (Non-centered parametrization)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

Under the NCP, equation (37) was re-defined as follows:

$$\eta_k^{(2)} = \mu_k^{(2)} + \sigma_k^{(2)} z_k^{(2)} \quad (49)$$

$$z_k^{(2)} \sim \text{Normal}(0, 1) \quad (50)$$

Equation (27) was re-defined as follows:

$$\boldsymbol{\theta}_j^{(2)} = \boldsymbol{\mu}_j^{(2)} + \mathbf{S}^{(2)} \cdot \mathbf{L}_{\Sigma}^{(2)} \cdot (\mathbf{z}_j \mathbf{I}) \quad (51)$$

$$\mathbf{z}_j = [z_{j1}, \dots, z_{jd}]^T \quad (52)$$

$$z_{jd} \sim \text{Normal}(0, 1) \quad (53)$$

$$\mathbf{L}_{\Sigma}^{(2)} \sim \text{LKJCorrelationCholesky}(2) \quad (54)$$

Likelihood, priors and hyper-priors (Non-centered parametrization, cont.)

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

Finally, equation (35) was re-defined as follows:

$$\theta_j^{(3)} = \mu_j^{(3)} + \sigma_\theta^{(3)} z_j \quad (55)$$

$$z_j \sim \text{Normal}(0, 1) \quad (56)$$

Identification

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

We used the unit variance identification scheme (UVI), that is, to set the scale of the higher-order dimension and sub-dimensions to one:

- $\sigma_{\theta}^{(3)} = 1$
- $S^{(2)} = \boldsymbol{\sigma}_{\theta}^{(2)} I$ with $\boldsymbol{\sigma}_{\theta}^{(2)} = [1, 1, 1]^T$
- $\mu_{\eta} = 0, \sigma_{\eta} = 1, \delta_{\eta} = 2$

The second turned the covariance matrix into a correlation, i.e. $\Sigma^{(2)} = \mathbf{R}^{(2)}$.

Prior predictive investigation

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

From two perspectives:

- the IRT perspective
- the outcome perspective

Prior predictive investigation (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

From the IRT perspective:

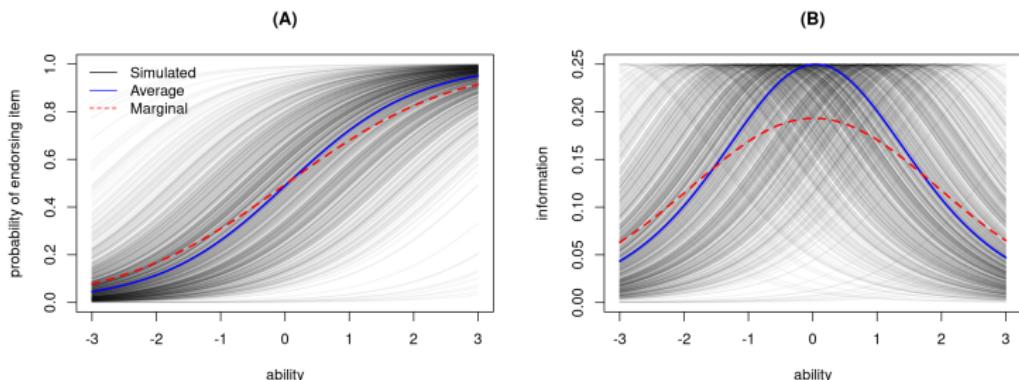


Figure: First-order latent variable model (FOLV). (A) Item Characteristics Curve, ICC. (B) Item Information Function, IIF.

Prior predictive investigation (cont.)

Outline

1. Introduction

2. GLLAMM dichotomous

3. Bayesian estimation

4. Simulation study

5. Application

6. Conclusions

Appendix

References

From the outcome perspective:

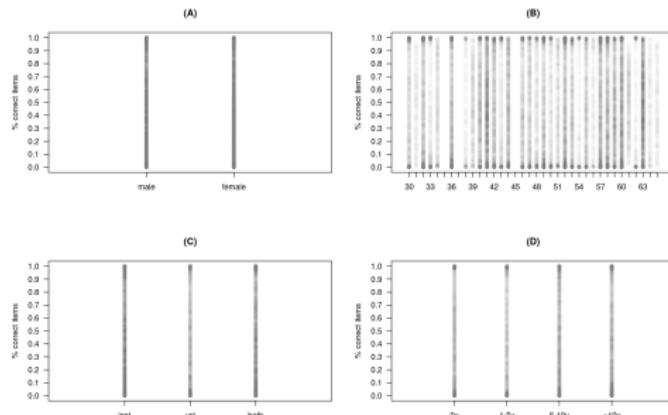


Figure: First-order latent variable model (FOLV). Aggregated endorsement rate per simulated covariate: (A) gender, (B) age, (C) education, and (D) experience.

Running time

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

	Parametrization	Sample	Time (min.)		
			mean	min	max
1	CP	100	4.80	1.89	7.56
2	CP	250	7.85	5.60	16.57
3	CP	500	19.15	17.20	22.53
4	NCP	100	1.94	1.74	2.23
5	NCP	250	7.18	6.78	8.40
6	NCP	500	22.38	20.12	28.62

Table: First-order latent variable model (FOLV). Running time statistics.

Running time (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

	Parametrization	Sample	Time (min.)		
			mean	min	max
1	CP	100	4.04	1.39	8.55
2	CP	250	7.13	5.22	11.49
3	CP	500	16.14	14.45	19.73
4	NCP	100	1.87	1.62	2.28
5	NCP	250	5.92	5.20	6.79
6	NCP	500	16.58	13.37	18.26

Table: Second-order latent variable model (SOLV). Running time statistics.

Application model fit

Outline

1. Introduction
 2. GLLAMM dichotomous
 3. Bayesian estimation
 4. Simulation study
 5. Application
 6. Conclusions
- Appendix
- References

They try to approximate the out-of-sample KL-divergence [12]:

	Model	Param.	WAIC	Ippd	penalty
1	FOLV	CP	54,632.4	-25,429.7	1,886.5
2	FOLV	NCP	54,631.7	-25,427.6	1,888.3
3	SOLV	CP	54,610.3	-25,348.4	1,956.7
4	SOLV	NCP	54,614.7	-25,337.9	1,969.4

Table: Model fit. Widely Applicable Information Criterion (WAIC).

Application model fit (cont.)

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

	Model	Param.	PSIS	lppd	penalty
1	FOLV	CP	54,657.4	-27,328.7	1,904.4
2	FOLV	NCP	54,656.9	-27,328.5	1,898.1
3	SOLV	CP	54,627.3	-27,313.7	1,940.1
4	SOLV	NCP	54,642.5	-27,321.2	1,990.2

Table: Model fit. Pareto-smoothed importance sampling cross-validation (PSIS).

References I

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- [1] Baker, F. [2001]. The basic of item response theory, *Technical report*, ERIC Clearinghouse on Assessment and Evaluation.
- [2] Betancourt, M. and Girolami, M. [2012]. Hamiltonian monte carlo for hierarchical models.
url: arxiv.org/abs/1312.0906v1.
- [3] 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>.

References II

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [4] 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.
- [5] 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>.
- [6] 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.

References III

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [7] 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.
- [8] 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>.
- [9] Hambleton, R. and Swaminathan, H. [1991]. *Item Response Theory*, Evaluation in Education and Human Services series, Springer Science+Business Media, LLC.

References IV

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- [10] Hambleton, R., Swaminathan, H. and Rogers, H. [1991]. *Fundamentals of Item Response Theory*, SAGE Publications Inc.
- [11] Jiao, H., Kamata, A., Wang, S. and Jin, Y. [2012]. A multilevel testlet model for dual local dependence, *Journal of Educational Measurement* **49**(1): 82–100.
doi:
<https://doi.org/10.1111/j.1745-3984.2011.00161.x>.
- [12] Kullback, S., . L. R. [1951]. On information and sufficiency, *The Annals of Mathematical Statistics* **22**(1): 79–86.
url: <http://www.jstor.org/stable/2236703>.

References V

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [13] Lord, F. and Novik, M. [2008]. *Statistical Theories of Mental Test Scores*, Information Age Publishing.
- [14] McCullagh, P. and Nelder, J. [1989]. *Generalized Linear Models*, Monographs on Statistics Applied Probability, Chapman Hall/CRC Press.
- [15] 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>.

References VI

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [16] 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>.
- [17] 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>.

References VII

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- [18] 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>.
- [19] Patz, R. J. and Junker, B. W. [1999]. A straightforward approach to markov chain monte carlo methods for item response models, *Journal of Educational and Behavioral Statistics* **24**(2): 146–178.
doi: 10.3102/10769986024002146.

References VIII

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [20] 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>.
- [21] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004b]. 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/S0304407004001708>

References IX

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- [22] Rasch, G. [1980]. *Probabilistic Models for Some Intelligence and Attainment Tests*, University of Chicago Press.
- [23] 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.
- [24] Reckase, M. [2009]. *Multidimensional Item Response Theory*, Statistics for Social and Behavioral Sciences, Springer Science+Business Media, LLC.

References X

Outline

1. Introduction
2. GLLAMM dichotomous
3. Bayesian estimation
4. Simulation study
5. Application
6. Conclusions

Appendix

References

- [25] Skrondal, A. and Rabe-Hesketh, S. [2004]. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Interdisciplinary Statistics, Chapman Hall/CRC Press.
- [26] Stan Development Team. [2021]. *Stan Modeling Language Users Guide and Reference Manual, version 2.26*, Vienna, Austria.
url: <https://mc-stan.org>.
- [27] Wainer, H., Bradlow, E. and Wang, X. [2007]. *Testlet response theory and its applications*, Cambridge University Press.

References XI

Outline

1. Introduction

2. GLLAMM
dichotomous

3. Bayesian
estimation

4. Simulation
study

5. Application

6. Conclusions

Appendix

References

- [28] 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>.