Bayesian inference for a skew-normal IRT model under the centred parameterization

Caio L. N. Azevedo, IMECC/Unicamp
Heleno Bolfarine, IME/USP
Dalton F. Andrade, INE/UFSC

10th Bayesian Statistics Brazilian Meeting

- Brief review about Item Response Theory.
- Two parameter IRT model for dichotomous responses.
- Latent traits.
- Skew normal distribution for the latent traits.
- Bayesian estimation.
- Simulation.
- Comments.

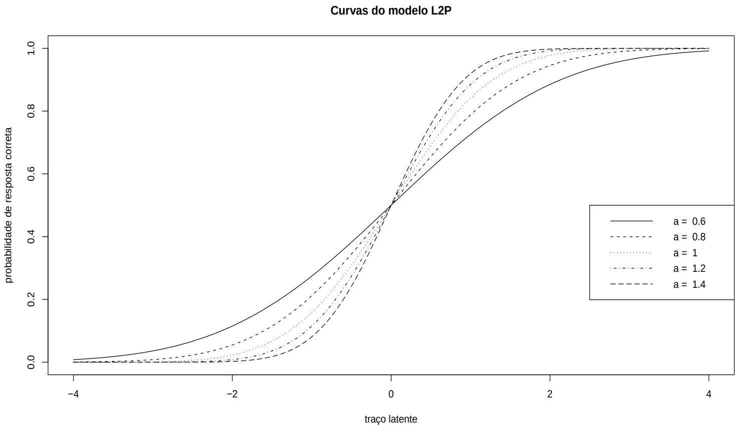
- Psychometric theory: a set of models which deals with latent variables (called latent traits).
- Item Response Models (IRM) : represent the probability of a examinee get a certain score in an item.
- Such probability is a function of the latent traits (examinees) and the item parameters (item).
- Large number of differents IRM : dichotomous, polytomous, one and multiple groups, multidimensionals, longitudinals.
- Applications in many fields: educational assessment, bilogical essays, marketing among other applications.
- First works due to Lord (1952) and to Rasch (1960).

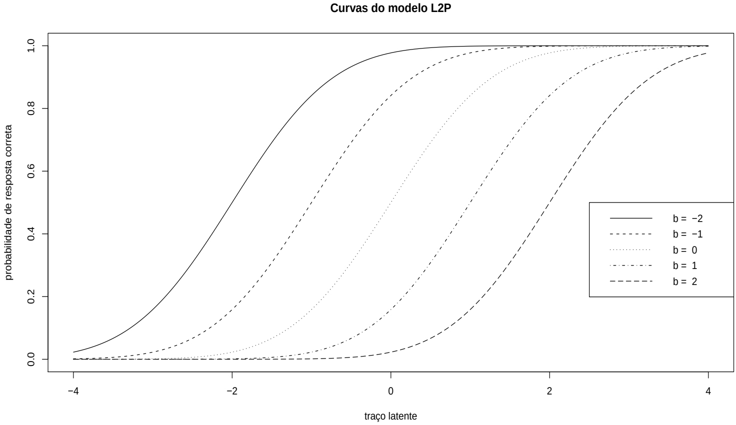
Item Response Function

$$P(Y_{ij} = 1|(\theta_j, \zeta_i)) \;=\; \Phi(a_i(\theta_j - b_i^*)) = \Phi(a_i\theta_j - b_i) \qquad (1)$$

$i = 1, \ldots, I$ (item), $j = 1, \ldots, n$ (indivíduo),

- Y_{ij} : is the answer of the examinee j to the item i. It is equals to 1 if the examinee answers the item i correctly and 0 otherwise.
- θ_j : is the latent trait (knowledge, “level of depression”, etc, of the examinee j.
- $\zeta_i = (a_i, b_i)^t$.
- a_i : is the discrimination parameter of the item i.
- b_i^* : is the difficulty parameter of the item i.
- $b_i = a_i b_i$: is the slope of the item i.


Curvas do modelo L2P
probabilidade de resposta correta
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
traço latente
a = 0.6
a = 0.8
a = 1
a = 1.2
a = 1.4



Curvas do modelo L2P
probabilidade de resposta correta
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
traço latente
b = −2
b = −1
b = 0
b = 1
b = 2

- Applications of the two parameter IRT model:
  - Cognitive tests with open items (corrected as right/wrong): basic level school.
  - Clinical assessment questionnaires.
  - Institucional assessment questionnaires.
    - Total quality management.
    - Schoolar management.
  - Test with multiple choice items where "guessing" is no possible.

- Latent traits:
  - (Intrinsic) Characteristics of the subjetcs (examinees).
  - Fixed effetcs: parameters.
  - Random effects: random variables.
  - An IRT models is (completely) characterized by its Item Response Function.
  - An usual assumption: the latent traits follow a suitable distribution (either a bayesian or a frequentist approach).
  - They incorporate information about the sampling process.
  - Prior information can be incoporated.
  - It is possible to make inference concerning not observed subjetcs.
  - It is helpfull to ensure the model identifiability.

- An usual assumption is: $\theta_j|\boldsymbol{\eta}_\theta \sim N(0,1)$, $\boldsymbol{\eta}_\theta = (\mu_\theta = 0, \psi_\theta = 1)$.
- This assumption can be unrealistic.
- Normality assumption does not hold: asymmetry, multimodality, heavy tails.
- Proposals in the literature.
  - Finite mixture of normal distributions: Mislevy (1984).
  - Beta-Binomial: Mislevy (1984).
  - Nonparametric approach (histogram): Mislevy (1984).
  - Multivariate t distribution with known degrees of freedom: Ghosh et al (2000).
  - Skew normal under the direct parameterization: Bazan, Branco & Bolfarine et al (2006).
  - Univariate t distribution with known degreees of reedom: Azevedo & Andrade (2007).

- Focus: asymmetry.
  - Selection of examinees: highest scores, lowest social - economic status.
  - Special teaching progam : longitudinal designs.
  - The nature of the latent traits distribution of the examinees.
- Alternative (skew-normal distribuion)

$$\begin{aligned} \theta_j|\boldsymbol{\eta}_\theta &\sim SN(0,1,\lambda_\theta), \\ \boldsymbol{\eta}_\theta &= (\mu_\theta = 0, \psi_\theta = 1, \lambda_\theta) \end{aligned} \tag{2}$$

- It is necessary to determine (to stablish) the latent trait scale.
- The results must be not only comparable but also be interpretable.
- Fact: under the assumption (2):

$$\mathcal{E}(\theta_j|\lambda_\theta) = h(\lambda_\theta) \tag{3}$$

$$\mathcal{V}(\theta_j|\lambda_\theta) = g(\lambda_\theta) \tag{4}$$

- This makes the model (1) be not identified.

Notice that:

$$
\begin{aligned}
P(Y_{ij}=1|\theta_j,\zeta_i) &= \Phi(a_i(\theta_j-b_i)) = \Phi\left(\frac{a_i}{\alpha}(\alpha\theta_j-\alpha b_i)\right)\\
&= \Phi\left(\frac{a_i}{\alpha}(\alpha\theta_j+\beta-\alpha b_i-\beta)\right)\\
&= \Phi\left(-a_i^*\left(\theta_j^*-b_i^*\right)\right) \qquad (5)
\end{aligned}
$$

where $\theta_j^*|\lambda_\theta \sim SN(\beta,\alpha^2,\lambda_\theta)$.
This occurs because the expected value and the variance of θ, that is, the metric, is not defined. This, in its turn, makes the model be not identified.

identifiability $\leftrightarrow$ metric is defined.

- Solution: the using of the centred parameterization defined by Azzalini (1989):

$$\theta_j^{(C)} = \frac{\theta_j - h(\lambda_\theta)}{\sqrt{g(\lambda_\theta)}} \tag{6}$$

- Therefore, $\theta_j^{(C)} \sim SN_C(0, 1, \gamma_\theta)$, where, $\forall\, \gamma_\theta \in (-0.99527, 0.99527)$:

$$\mathcal{E}(\theta_j^{(C)} | \lambda_\theta) = 0$$

$$\mathcal{V}(\theta_j^{(C)} | \lambda_\theta) = 1$$

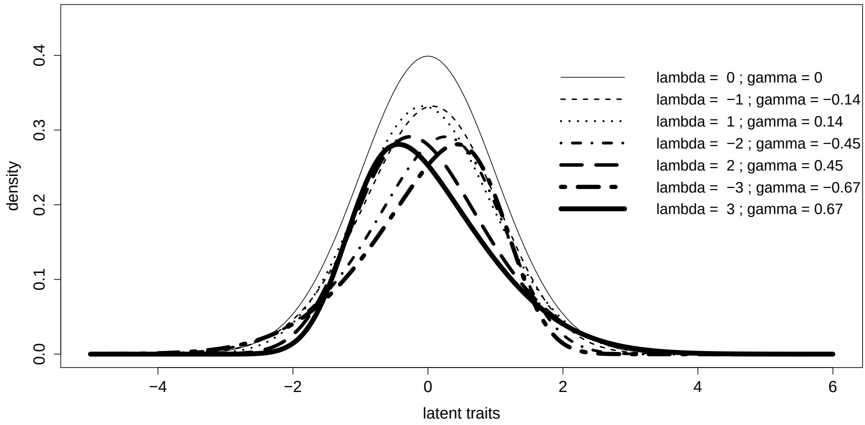
and, in this way, transformations as statetd in (5) are no longer possible and, therefore, the model is identified (the metric is defined).

- **Result**: The skew normal density under the CP is given by:

$$f(\theta_j|\gamma_\theta) \quad = \quad 2(\varsigma_\theta)^{-1/2}\phi\left(\frac{\theta_j-\alpha_\theta}{\sqrt{\varsigma_\theta}}\right)\Phi\left[\lambda_\theta\left(\frac{\theta_j-\alpha_\theta}{\sqrt{\varsigma_\theta}}\right)\right], \qquad (7)$$

where

$$\begin{aligned}
\alpha_\theta &\equiv \alpha_\theta(\gamma_\theta) = -s\gamma_\theta^{1/3}, \\
\varsigma_\theta &\equiv \varsigma_\theta(\gamma_\theta) = 1 + s^2\gamma_\theta^{2/3}, \\
\lambda_\theta &\equiv \lambda_\theta(\gamma_\theta) = \frac{s\gamma_\theta^{1/3}}{\sqrt{r^2 + s^2\gamma_\theta^{2/3}(r^2-1)}}, \\
r &= \sqrt{\frac{2}{\pi}}, s = \left(\frac{2}{4-\pi}\right)^{1/3}, \\
\gamma_\theta &= \sqrt{\frac{2}{\pi}}\delta_\theta^3\left[\frac{4}{\pi}-1\right]\left[1-\frac{2}{\pi}\delta_\theta^2\right]^{-3/2}, \gamma_\theta \in (-0.99527, 0.99527), \text{ and} \\
\delta_\theta &= \frac{\lambda_\theta}{\sqrt{1+\lambda_\theta^2}}.
\end{aligned}$$


density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
6
latent traits
lambda = 0 ; gamma = 0
lambda = −1 ; gamma = −0.14
lambda = 1 ; gamma = 0.14
lambda = −2 ; gamma = −0.45
lambda = 2 ; gamma = 0.45
lambda = −3 ; gamma = −0.67
lambda = 3 ; gamma = 0.67

- Augmented data likelihood

$$p(\boldsymbol{z}_{..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) = \left\{ \prod_{j=1}^{n} \prod_{i \in \mathcal{I}_j} p(z_{ij}|\theta_j, \boldsymbol{\zeta}_i) \right\},$$

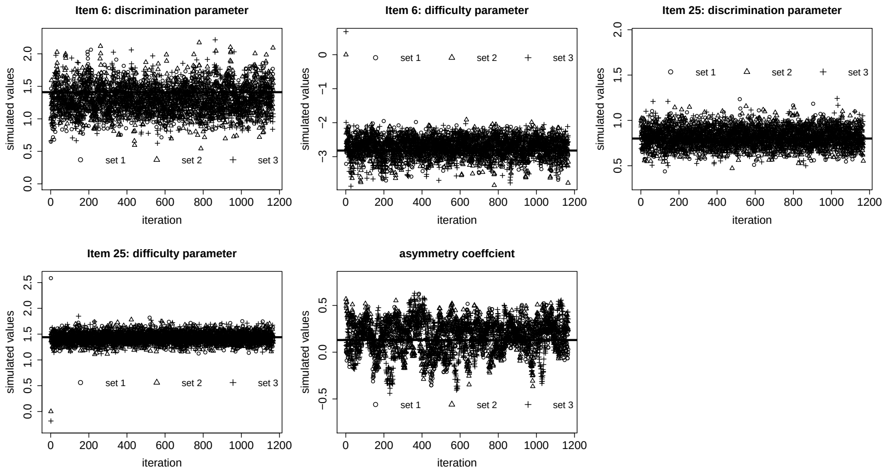
- Prior

$$p(\boldsymbol{\theta}, \boldsymbol{\zeta}, \gamma_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\gamma) = \left\{ \prod_{j=1}^{n} p(\theta_j|\gamma_\theta) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_\zeta) \right\} p(\gamma_\theta|\boldsymbol{\eta}_\gamma),$$

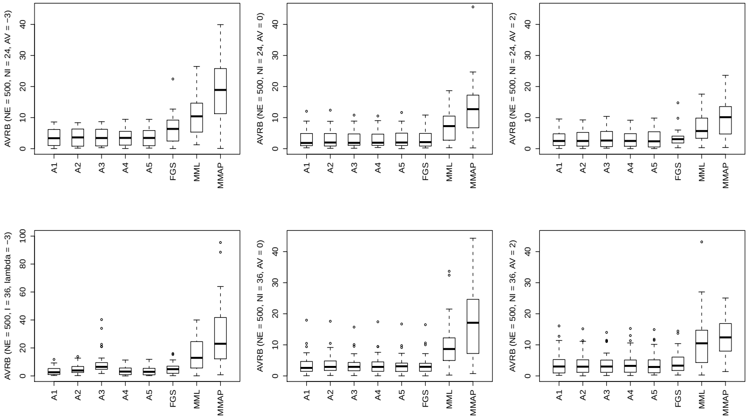
$$
\begin{aligned}
p(\boldsymbol{z}_{..}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \gamma_\theta | \boldsymbol{y}_{..}, \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\gamma) &\propto p(\boldsymbol{z}_{..} | \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{y}_{..}) p(\boldsymbol{\theta}|\gamma_\theta) p(\boldsymbol{\zeta}|\boldsymbol{\eta}_\zeta) p(\gamma_\theta|\boldsymbol{\eta}_\gamma) \\
&= \left\{ \prod_{j=1}^{n} \prod_{i \in \mathcal{I}_j} p(z_{ij}|\theta_j, \boldsymbol{\zeta}_i, y_{ij}) \right\} \left\{ \prod_{j=1}^{n} p(\theta_j|\gamma_\theta) \right\} \\
&\times \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_\zeta) \right\} \left\{ p(\gamma_\theta|\boldsymbol{\eta}_\gamma) \right\} \\
&\propto \prod_{j=1}^{n} \prod_{i \in \mathcal{I}_j} \exp\left[-0.5 \left(z_{ij} - a_i\theta_j + b_i \right)^2 \right] I_{(y_{ijk}, z_{ijk})} \\
&\times \left\{ \prod_{j=1}^{n} \phi\left(\frac{\theta_j - \alpha_\theta}{\sqrt{\varsigma_\theta}} \right) \right\} \left\{ \prod_{j=1}^{n} \Phi\left[\lambda_\theta \left(\frac{\theta_j - \alpha_\theta}{\sqrt{\varsigma_\theta}} \right) \right] \right\} \\
&\times \left\{ \prod_{i=1}^{I} \exp\left[\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta \right)^t \boldsymbol{\Psi}_\zeta^{-1} \left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta \right) \right] I_{(a_i > 0)} \right\} \\
&\times (\varsigma_\theta)^{-n/2} I_{\left(\gamma_\theta \in A_{\gamma_\theta}\right)}(\gamma_\theta),
\end{aligned}
$$

- The posterior has an intractable form, however a Metropolis-Hastings within Gibbs sampling algorithm is feasible (MHWGS).

- Several situations were simulated concerning: number of examinees (NE) (500,1000), number of items (NI) (24,36) and asymmetry values (AV) $\lambda_\theta \in$ (-3,-2,-1,0,1,2,3) ($\gamma_\theta \in (-0.67, -0.45, -0.14, 0.00, 0.14, 0.45, 0.67)$).
- Were generated R $=$ 10 replicas.
- Sensitivity to the choice of the hyperparameter of the kernel dentisty of γ_θ.
- Comparison of the MHWGS algorithm with other estimation methods.
- More details: Azevedo, Bolfarine & Andrade (2010).
- The proposed algorithm (MHWGS) recoveries all parameters properly.
- Best results are obtained by MHWGS.
- Burn-in: 5000; total sample $=$ 35000; thin : 30; valid sample $=$ 1000.


Item 6: discrimination parameter
Item 6: difficulty parameter
Item 25: discrimination parameter
Item 25: difficulty parameter
asymmetry coeffcient
simulated values
iteration
set 1
set 2
set 3
0
200
400
600
800
1000
1200

NE	NI	AV	ADMHWGS					FGS	MML	MMAP
			1	2	3	4	5			
500	24	-3	0.154	0.154	0.154	0.153	0.155	0.161	0.237	0.326
500	36	-3	0.205	0.210	0.303	0.205	0.204	0.212	0.507	0.774
1000	24	-3	0.103	0.105	0.118	0.103	0.153	0.132	0.237	0.292
1000	36	-3	0.160	0.190	0.286	0.161	0.284	0.193	0.554	0.688
500	24	-2	0.135	0.136	0.138	0.138	0.138	0.152	0.233	0.342
500	36	-2	0.214	0.209	0.214	0.214	0.214	0.219	0.555	0.801
1000	24	-2	0.127	0.131	0.127	0.131	0.125	0.133	0.217	0.265
1000	36	-2	0.151	0.155	0.155	0.156	0.154	0.168	0.522	0.666
500	24	-1	0.167	0.171	0.165	0.169	0.172	0.170	0.299	0.409
500	36	-1	0.207	0.203	0.201	0.207	0.205	0.204	0.503	0.758
1000	24	-1	0.120	0.117	0.124	0.120	0.122	0.117	0.223	0.269
1000	36	-1	0.165	0.169	0.166	0.165	0.168	0.164	0.517	0.661
500	24	0	0.149	0.150	0.150	0.149	0.150	0.143	0.239	0.317
500	36	0	0.195	0.194	0.189	0.189	0.192	0.194	0.425	0.649
1000	24	0	0.108	0.107	0.108	0.108	0.108	0.108	0.203	0.236
1000	36	0	0.156	0.157	0.158	0.155	0.158	0.154	0.498	0.608
500	24	1	0.167	0.171	0.165	0.169	0.172	0.170	0.256	0.315
500	36	1	0.191	0.189	0.190	0.188	0.190	0.190	0.525	0.728
1000	24	1	0.110	0.110	0.107	0.108	0.107	0.100	0.193	0.220
1000	36	1	0.186	0.183	0.185	0.182	0.184	0.185	0.505	0.619
500	24	2	0.147	0.145	0.146	0.145	0.145	0.144	0.209	0.261
500	36	2	0.196	0.194	0.197	0.196	0.195	0.199	0.550	1.037
1000	24	2	0.129	0.127	0.127	0.125	0.126	0.132	0.226	0.248
1000	36	2	0.163	0.160	0.161	0.160	0.161	0.175	0.522	0.640
500	24	3	0.151	0.148	0.149	0.154	0.151	0.162	0.250	0.329
500	36	3	0.210	0.209	0.209	0.209	0.212	0.218	0.441	0.618
1000	24	3	0.230	0.300	0.299	0.120	0.302	0.158	0.257	0.286
1000	36	3	0.153	0.152	0.158	0.155	0.152	0.184	0.524	0.636


AVRB (NE = 500, NI = 24, AV = −3)
AVRB (NE = 500, NI = 24, AV = 0)
AVRB (NE = 500, NI = 24, AV = 2)
AVRB (NE = 500, I = 36, lambda = −3)
AVRB (NE = 500, NI = 36, AV = 0)
AVRB (NE = 500, NI = 36, AV = 2)
A1
A2
A3
A4
A5
FGS
MML
MMAP

- The CP ensures the model identifiability under the skew normality of the latent traits distribution.
- The MHWGS algorithm recoveries all parameters properly.
- Biased results are obtained when the asymmetry of the latent traits distribution is not considered.
- In the real data analysis (not showed) we found that the latent traits distributions presents negative asymmetry.

- Azzalini, A. (1985), *A class of distribution which includes the normal ones*, Scandinavian journal of statistics. 12, pp. 171-178.
- Azevedo, C. L. N., Bolfarine, H. & Andrade. D. F. (2010). *Bayesian inference for a skew-normal IRT model under the centred parameterization*. Compputational Staistics & Data Analysis, under revision.
- Andrade, D.F., Tavares, H.R., Cunha, R.V. (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: Associação Brasileira de Estatística. Bazán, J. L., Branco, M. D. and Bolfarine, H. (2006), *A Skew Item Response Model*, Bayesian analysis, pp. 861-892.
- Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale: Lawrence Erlbaum Associates

Muito obrigado !

Thank you very much!