Modelagem de dados de resposta ao item sob efeito de speededness

Joelson da Cruz Campos

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística
Orientador: Prof^a. Dr^a. Márcia D'Elia Branco

Durante parte do desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, Abril de 2016

Modelagem de dados de resposta ao item sob efeito de speededness

Esta versão definitiva da tese contém as correções e alterações sugeridas pela comissão julgadora durante a defesa realizada por Joelson da Cruz Campos no dia 08/04/2016

Comissão Julgadora:

- Márcia D'Elia Branco (Orientadora) - IME-USP
- Júlia Maria Pavan Soler - IME-USP
- Jorge Luis Bazán Guzmán - ICMC-USP
- Caio Lucidius Naberezny Azevedo - IMECC-UNICAMP
- Cibele Queiroz da Silva - UNB

Agradecimentos

A Deus por ter me dado forças nos momentos mais difíceis e ter permitido que eu conseguisse concluir esse trabalho.

A professora Márcia pela orientação, amizade e apoio durante o desenvolvimento desse trabalho e por sempre ter acreditado que eu poderia concluí-lo.

A professora/amiga Michelli por toda a ajuda para que eu conseguisse ir a São Paulo e por toda confiança em mim depositada.

A UFCG e à todos os meus colegas de trabalho da UAEST pelo apoio e pela liberação para que eu conseguisse regressar a USP e concluir o doutorado.

A todos os amigos da república que proporcionaram um excelente ambiente em meu regresso a São Paulo.

A todos os colegas que comigo estudaram e compartilharam um pouco de seus conhecimentos. Entre eles agradeço especialmente ao Bruno, ao Gilberto (Tomtom) e ao Plínio que correspondem aos que mais estudaram comigo nesses últimos anos.

A Itailma (tatá) por todos os momentos comigo compartilhados ao longo dos últimos 10 anos.

A toda minha família por sempre ter acreditado em mim.

Resumo

Em testes nos quais uma quantidade considerável de indivíduos não dispõe de tempo suficiente para responder todos os itens temos o que é chamado de efeito de *Speededness*. O uso do modelo unidimensional da Teoria da Resposta ao Item (TRI) em testes com *speededness* pode nos levar a uma série de interpretações errôneas uma vez que nesse modelo é suposto que os respondentes possuem tempo suficiente para responder todos os itens. Nesse trabalho, desenvolvemos uma análise Bayesiana do modelo tri-dimensional da TRI proposto por [Wollack e Cohen \(2005\)](#) considerando uma estrutura de dependência entre as distribuições a priori dos traços latentes a qual modelamos com o uso de cópulas. Apresentamos um processo de estimação para o modelo proposto e fazemos um estudo de simulação comparativo com a análise realizada por [Bazan *et al.* \(2010\)](#) na qual foi utilizada distribuições *a priori* independentes para os traços latentes. Finalmente, fazemos uma análise de sensibilidade do modelo em estudo e apresentamos uma aplicação levando em conta um conjunto de dados reais proveniente de um subteste do EGRA, chamado de *Nonsense Words*, realizado no Peru em 2007. Nesse subteste os alunos são avaliados por via oral efetuando a leitura, sequencialmente, de 50 palavras sem sentidos em 60 segundos o que caracteriza a presença do efeito *speededness*.

Palavras-chave: Teoria da Resposta ao Item, Speededness, Cópulas, Gibbs Sampling.

Abstract

In tests where a reasonable amount of individuals does not have enough time to answer all items we observe what is called effect of Speededness. The use of a unidimensional model from Item Response Theory (IRT) in tests with speededness can lead us to erroneous interpretations, since this model assumes that the respondents have enough time to answer all items. In this work, we propose a Bayesian analysis of the three-dimensional item response models (IRT) proposed by Wollack and Cohen et al (2005) considering a dependency structure between the prior distributions of the latent traits which is modeled using Copulas. We propose and develop a MCMC algorithm for the estimation of the model. A simulation study comparing with the analysis in Bazan et al (2010), wherein an independent prior distribution assumption was presented. Finally, we apply our model in a set of real data from EGRA, called Nonsense Words, held in Peru in 2007, where students are evaluated for their performance in reading.

Keywords: Item theory response, Speededness, Copulas, Gibbs Sampling.

Sumário

Lista de Abreviaturas	xiii
Lista de Símbolos	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Modelos Clássicos da TRI para Respostas Dicotômicas	2
1.1.1 Interpretação dos modelos	3
1.1.2 Casos particulares dos modelos de 3 parâmetros	7
1.1.3 Suposições do modelo	7
1.2 Métodos de Estimação	8
1.3 Organização do Trabalho	11
2 Modelos da TRI que levam em consideração o <i>speededness</i>	13
2.1 Modelo Híbrido	13
2.2 Modelo Mistura de Rasch Multi-classe	15
2.3 Modelo de Resposta ao Item com Processo de Mudança de Resposta Gradual	18
2.4 Estimação Clássica no Modelo 3PLMG	20
3 Estimação Bayesiana do Modelo 3PLMG	23
3.1 Cópulas	24
3.2 Estimação MCMC da Matriz de Covariâncias	29
3.2.1 Caso da Matriz Irrestrita	29
3.2.2 Caso da Matriz de Covariância Restrita	32
3.3 Estimação Bayesiana no Modelo 3PLMG com Dependência entre as Distribuições <i>a Priori</i>	33
4 Estudo de Simulação	37
4.1 Modelo com distribuições <i>a priori</i> independentes	37
4.2 Modelo com distribuições <i>a priori</i> dependentes	44
4.3 Comparações dos modelos sem/com dependência entre as distribuições <i>a priori</i> e conclusões do estudo de simulação	48

5	Aplicação	51
5.1	Descrição dos dados	51
5.2	Análise de Sensibilidade	53
5.3	Ajuste do Modelo	55
5.4	Considerações Finais	65
A	Algumas Distribuições	67
A.1	Distribuição Categórica	67
A.2	Distribuição de Dirichlet	67
A.3	Distribuições Wishart e Wishart Inversa	67
B	Comparação de Modelos	69
B.1	Pseudo Fator de Bayes	69
B.2	$\bar{D}$ e DIC	70
C	Prova do Teorema 2	73
	Referências Bibliográficas	77

Lista de Abreviaturas

ENEM	Exame Nacional do Ensino Médio
CCI	Curva Característica do Item
EdData II	Education data for decision making
EGRA	Early Grade Reading Assessment
fda	Função de distribuição acumulada
MCMC	Markov Chain Monte Carlo
NWF	Nonsense Word Fluency
PISA	Programme for International Student Assessment
SARESP	Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
TAI	Testes Adaptativos Informatizados
TCM	Teoria Clássica das Medidas
TCT	Teoria Clássica dos testes
TOEFL	Teste of English as a Foreign Language
TRI	Teoria da Resposta ao Item
1PL	Um Parâmetro Logístico
2PL	Dois Parâmetros Logístico
3PL	Modelo Logístico de 3 Parâmetros
1PLMG	Um Parâmetros Logístico com Processo de Mudança Gradual
2PLMG	Dois Parâmetros Logístico com Processo de Mudança Gradual
3PLMG	Três Parâmetros Logístico com Processo de Mudança Gradual
DIC	<i>Deviance Information Criterion</i>
P	Fator de penalização
C	Cópula
N	Distribuição normal
LN	Distribuição log-normal
Dir	Distribuição de dirichlet
Cat	Distribuição categórica

Lista de Símbolos

ϕ	densidade da normal padrão
Φ	fda da distribuição normal padrão
Ψ	Função de Ligação
$\mathbb{P}$	Função de probabilidade
$\mathbb{F}$	Função de distribuição acumulada
$\sim$	Tem como distribuição
$\rightarrow$	Converge para

Lista de Figuras

1.1	CCI para o modelo 3PL considerando $D = 1, 7$ (linha) e para o modelo 3PN (tracejado).	4
1.2	CCI para o modelo 3PL considerando $D = 1, 7$ (linha) e para o modelo 3PN (tracejado).	5
1.3	CCI para o modelo 3PL para diferentes valores de a_i .	6
1.4	CCI para o modelo 3PL considerando $D = 1, 7$ (linha) e para o modelo 3ON (tracejado).	6
1.5	CCI considerando o parâmetro de discriminação negativo (esquerda) e positivo (direita).	8
2.1	Gráfico do fator de penalização para $\eta = 0, 5$ e diferentes valores de λ considerando um teste com 80 itens.	19
2.2	Gráfico do fator de penalização para diferentes valores de η e λ considerando um teste com 80 itens.	19
4.1	<i>Boxplots</i> dos valores reais dos parâmetros de discriminação, a_i , e das estimativas obtidas nas réplicas de 1 à 10 usando distribuições <i>a priori</i> independentes para os traços latentes. Consideramos que $a_i \sim N(1; 0, 5)I(0,)$.	38
4.2	<i>Boxplots</i> dos valores reais dos parâmetros de intercepto, b_i^* , e das estimativas obtidas para as 10 réplicas usando distribuições <i>a priori</i> independentes para os traços latentes. Consideramos que $b_i^* \sim N(0, 2)$.	39
4.3	<i>Boxplots</i> dos valores reais das habilidades, θ_j , e das estimativas obtidas para as réplica de 1 à 10 obtidas usando distribuições <i>a priori</i> independentes para os traços latentes. Foi considerado que $\theta_j \sim N(0, 1)$.	42
4.4	<i>Boxplots</i> dos valores reais de tolerância ao <i>speededness</i> , η_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições <i>a priori</i> independentes para os traços latentes. Consideramos que $\eta_j \sim Beta(1, 1)$.	42
4.5	<i>Boxplots</i> dos valores reais das propensões ao <i>speededness</i> , λ_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições <i>a priori</i> independentes para os traços latentes. Consideramos que $\lambda \sim LN(1, 1)$.	43
4.6	<i>Boxplots</i> dos valores reais dos parâmetros de discriminação, a_i , e das estimativas obtidas nas réplicas de 1 à 10 usando distribuições <i>a priori</i> dependentes para os traços latentes. Consideramos que $a_i \sim N(1; 0, 5)I(0,)$.	45

4.7	<i>Boxplots</i> dos valores reais dos parâmetros de intercepto, b_i^* , e das estimativas obtidas para as 10 réplicas usando distribuições <i>a priori</i> dependentes para os traços latentes. Consideramos que $b_i^* \sim N(0, 2)$	46
4.8	<i>Boxplots</i> dos valores reais das habilidades, θ_j , e das estimativas obtidas para as réplica de 1 à 10 obtidas usando distribuições <i>a priori</i> dependentes para os traços latentes. Foi considerado que $\theta_j \sim N(0, 1)$	46
4.9	<i>Boxplots</i> dos valores reais de tolerância ao <i>speededness</i> , η_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições <i>a priori</i> dependentes para os traços latentes. Consideramos que $\eta_j \sim Beta(1, 1)$	47
4.10	<i>Boxplots</i> dos valores reais das propensões ao <i>speededness</i> , λ_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições <i>a priori</i> dependentes para os traços latentes. Consideramos que $\lambda \sim LN(1, 1)$	47
4.11	Comparação das estimativas dos parâmetros de discriminação do modelo 2PL com os seus correspondentes valores reais.	49
4.12	Comparação das estimativas dos parâmetros de intercepto do modelo 2PL com os seus respectivos valores reais.	50
4.13	Comparação das estimativas dos θ 's do modelo 2PL com os valores reais das habilidades.	50
5.1	Proporção de respostas corretas - EGRA	52
5.2	Gráfico das estimativas das tolerâncias ao <i>speededness</i> , η , vs a propensão ao <i>speededness</i> , λ , para as distribuições <i>a priori</i> em M_1 e M_2	55
5.3	Comparação das estimativas dos parâmetros de discriminação, a_i , do Modelo 2PL e dos parâmetros de intercepto, b_i^* , estimados com diferentes métodos/programas.	57
5.4	Comparação das estimativas dos parâmetros de discriminação dos itens, a_i , e dos parâmetros de intercepto, b_i^* , do Modelo 2PL obtidas utilizando a abordagem totalmente Bayesiana e a a abordagem Clássica com diferentes programas/pacotes.	59
5.5	Comparação entre as estimativas dos parâmetros de discriminação, a_i , e de posição, b_i^* , dos modelo 2PL com o modelo 2PLMG com distribuições <i>a priori</i> independentes.	60
5.6	Intervalo com 95% de credibilidade (HPD) do parâmetro de discriminação, a_i , e gráfico da distribuição <i>a priori</i> dos respectivos parâmetros para o modelo 2PL com independência <i>a priori</i>	61
5.7	Intervalo com 95% credibilidade (HPD) do parâmetro de discriminação, a_i , e gráfico da distribuição <i>a priori</i> dos respectivos parâmetros para o modelo 2PLMG com independência <i>a priori</i>	61
5.8	Intervalo com 95% de credibilidade (HPD) do parâmetro de intercepto, b_i^* , e gráfico da distribuição <i>a priori</i> dos respectivos parâmetros para o modelo 2PL.	61
5.9	Intervalo com 95% de credibilidade (HPD) do parâmetro de intercepto, b_i^* , e gráfico da distribuição <i>a priori</i> dos respectivos parâmetro para o modelo 2PLMG com independência <i>a priori</i>	62
5.10	Matriz de correlação das habilidades, θ_j , estimadas com os modelos 2PL e 2PLMG com independência entre as distribuições <i>a priori</i> e o escore.	63
5.11	Correlações entre as estimativas das variáveis latentes (habilidades, θ_j , tolerância ao <i>speededness</i> , η_j e propensão ao <i>speededness</i> , λ_j) do modelo 2PLMG com independência <i>a priori</i>	64

Lista de Tabelas

2.1	Ilustração de classes latentes para o modelo mistura de rasch com <i>speededness</i> considerando 30 itens e efeito <i>speededness</i> definido para os 5 últimos itens	15
2.2	ilustração de classes latentes para o modelo mistura de rasch multiclasse com 6 classes e considerando 30 itens	17
4.1	No quadro a esquerda temos os fatores de penalização aplicados a cada item bem como a probabilidade de resposta correta para um indivíduo que teve $\eta = 0.006$ e $\lambda = 30.65$ e escore nulo enquanto no quadro ao lado direito temos um indivíduo que obteve escore perfeito e $\eta = 0.935$ e $\lambda = 1.525$	40
4.2	No quadro a esquerda temos os fatores de penalização aplicados a cada item bem como a probabilidade de resposta correta para um indivíduo que teve $\eta = 0.044$ e $\lambda = 2.35$, enquanto no quadro ao lado direito apresentamos os mesmos dados para um indivíduo que possui $\eta = 0.708$ e $\lambda = 2.665$	41
4.3	Resumo das correlações reais e estimadas	43
4.4	Tabela contendo as correlações reais entre os traços latentes e as correlações médias estimadas obtidas através das 10 réplicas	43
4.5	Percentual de cobertura dos hpds com 95% de credibilidade para o modelo que considera independência entre as distribuições <i>a priori</i> para os traços latentes	44
4.6	Tabela contendo as correlações reais entre os traços latentes e as correlações médias estimadas obtidas através das 10 réplicas considerando distribuições <i>a priori</i> dependentes para os traços latentes	45
4.7	Percentual de cobertura dos hpds com 95% de credibilidade para o modelo que considera dependência entre as distribuições <i>a priori</i> para os traços latentes	45
4.8	Comparativo das medidas resumo dos valores reais dos parâmetros conjuntamente com as medidas resumo obtidas das estimativas obtidas considerando distribuições <i>a priori</i> independentes para as 10 réplicas estudadas.	48
4.9	Comparativo das medidas resumo dos valores reais dos parâmetros dos itens e dos traços latentes conjuntamente com as medidas resumo obtidas das estimativas, das 10 réplicas, considerando distribuições <i>a priori</i> dependentes.	48
4.10	Diferenças médias entre as estimativas obtidas considerando o modelo com distribuições <i>a priori</i> dependentes com as do modelo considerando distribuições <i>a priori</i> independentes.	49

4.11	Comprimento médio dos intervalos para HPD's com 95% de credibilidade para os parâmetros do modelo com independência entre as distribuições <i>a priori</i> para os traços latentes quando analisado as 10 réplicas.	49
4.12	Comprimento médio dos intervalos para HPD's com 95% de credibilidade para os parâmetros do modelo com dependência entre as distribuições <i>a priori</i> para os traços latentes quando analisado as 10 réplicas.	49
5.1	Tabela contendo propostas de distribuições <i>a priori</i> para os traços latentes η_j e λ_j . . .	54
5.2	Medidas de seleção de modelos	55
5.3	Escores observados e medidas resumo para as estimativas Bayesianas das variáveis latentes dos indivíduos sob os modelos 2PL e 2PLMG com independência <i>a priori</i> . .	62
5.4	Tabela comparativa das estimativas dos traços latentes dos modelos 2PL e 2PLMG com o escore	64
5.5	Comparação dos modelos 2PL e 2PLMG para os dados do EGRA	65

Capítulo 1

Introdução

A teoria da resposta ao item (TRI), consiste de um conjunto de modelos estatísticos baseados em pressupostos teóricos, associados à teoria dos testes, usados principalmente por psicólogos e educadores cujo objetivo é medir quantitativamente alguma(s) característica(s) do indivíduo que não pode(m) ser mensurada(s) diretamente. A esta(s) característica(s) damos o nome de traço(s) latente(s). Tais pressupostos motivaram a criação de modelos estatísticos hoje conhecidos como modelos da TRI. Atualmente, a TRI encontra-se amplamente difundida no país devido ao seu uso em provas como o ENEM (Exame Nacional do Ensino Médio) que serve para o ingresso no ensino superior e o SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo) que avalia os estudantes, do estado de São Paulo, do 3º, 5º, 7º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio avaliando seus conhecimentos por meio de provas com questões de Língua Portuguesa, Matemática, Ciências Humanas, Ciências da Natureza e redação com a finalidade de produzir um diagnóstico da situação da escolaridade básica paulista, visando orientar os gestores do ensino no monitoramento das políticas voltadas para a melhoria da qualidade educacional. Internacionalmente a TRI está difundida pela sua utilização no exame de proficiência em língua inglesa TOEFL (Test of English as a Foreign Language) e no PISA (Programme for International Student Assessment) programa internacional de avaliação de estudantes. A TRI surgiu por volta da década de 50 na área de psicometria como uma alternativa à Teoria Clássica dos Testes, TCT. Para mais detalhes sobre a TCT ver [Vianna \(1978\)](#) e [Van Belle e Arnold \(2000\)](#). No entanto, a TRI veio ganhar mais notoriedade recentemente devido aos avanços computacionais que permitiram ampliar sua utilização. Dentre as referências literárias mais importantes utilizadas nesse texto podemos citar os seguintes livros: [Lord \(1980\)](#), [Andrade *et al.* \(2000\)](#) e [Fox \(2010\)](#). Os modelos da TRI são utilizados não somente na área educacional mas também em outras áreas do conhecimento dentre as quais citamos duas à seguir:

Médica - Nessa área [Guewehr \(2007\)](#) comparou o desempenho da Teoria da Resposta ao Item com a TCT na avaliação de instrumentos da qualidade de vida aplicados a idosos. Em sua análise a autora fez uso do Modelo Logístico de 2 Parâmetros e concluiu que o mesmo é uma ferramenta útil e promissora na avaliação da qualidade de vida, como alternativa ou, tomada de forma conjunta com a Teoria Clássica de Medidas (TCM) e a Análise Fatorial (AF).

E-commerce - Nessa área [Moreira Junior *et al.* \(2013\)](#) propõem o uso de Testes Adaptativos Informatizados (TAI), que correspondem a testes que procuram estimar a habilidade do respondente através da aplicação de itens que sejam adequados a ele, como uma ferramenta alternativa as metodologias usuais utilizadas na avaliação da usabilidade de sites de *e-commerce*, desenvolvendo um algoritmo para um teste adaptativo informatizado baseado na TRI cujo objetivo é estimar o grau de usabilidade de sites de *e-commerce*.

Conforme veremos adiante os modelos usuais da TRI consideram que a probabilidade de um indivíduo responder corretamente um determinado item depende somente de seu traço latente (proficiência ou habilidade) e das características do item, em outras palavras, isso significa que esses

modelos assumem que as respostas atribuídas pelos indivíduos aos itens não são influenciadas pelo fator de tempo. De acordo com [Lu e Sireci \(2007\)](#) em testes nos quais uma quantidade considerável de indivíduos não possuem tempo suficiente para responder todos os itens temos o que chamamos de efeito de *speededness*. Na presença de *speededness* o uso dos modelos usuais da TRI (que não levam esse efeito em consideração) é inapropriado, pois as respostas dos indivíduos na presença da limitação do tempo podem acabar contaminando as estimativas das habilidades dos mesmos e as estimativas dos parâmetros dos itens, uma vez que os respondentes podem não conseguir responder corretamente determinados itens devido ao efeito de *speededness* ocasionado pela limitação do tempo e não pelo fato de não terem uma proficiência suficiente para respondê-los corretamente.

Nesse trabalho, concentraremos o nosso estudo em modelos da TRI que incluem o efeito de *speededness*. O modelo que será a base do nosso estudo é o modelo de resposta ao item com processo de mudança gradual proposto em [Wollack e Cohen \(2005\)](#). A vantagem do uso de tal modelo deve-se ao fato dele ser mais informativo que os modelos de três parâmetros usualmente utilizados em estudos de TRI. O modelo apresentado em [Wollack e Cohen \(2005\)](#) e posteriormente utilizado por [Goegebeur et al. \(2008\)](#) é mais informativo que os usuais devido ao fato do mesmo ser capaz de captar o ponto em que o indivíduo passa a sofrer o efeito da limitação de tempo, ou seja, o ponto a partir do qual o indivíduo começa a responder aos itens de maneira acelerada, o qual chamaremos de ponto de *speededness*, e a forma com que ele acelera a partir desse ponto, o que chamaremos de taxa de *speededness*.

No decorrer desse capítulo faremos então uma breve revisão dos principais modelos da TRI utilizados para modelar dados de respostas dicotômicas e dos principais métodos de estimação dos parâmetros desses modelos. O conteúdo que será apresentado no decorrer deste capítulo tem como intuito nos fundamentar para que possamos introduzir os modelos da TRI com *speededness* como uma extensão natural dos modelos sem *speededness*.

1.1 Modelos Clássicos da TRI para Respostas Dicotômicas

Os dados de respostas aos itens geralmente consistem de uma matriz, $\mathbf{Y}$, onde cada componente da matriz, denotado por y_{ij} , representa a resposta do j -ésimo indivíduo ao i -ésimo item de um teste, $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$. Essa resposta pode ser dicotômica (zeros e uns) ou politômica ($1, 2, \dots, K$), assumindo um caráter discreto ou categorizado de resposta. A cada elemento da matriz de respostas é associado uma probabilidade, $p_{ijk} = \mathbb{P}(Y_{ij} = k)$, que representa a probabilidade da resposta k ser atribuída ao item i pelo indivíduo j , com $\sum_{k=1}^K p_{ijk} = 1$.

A classificação dos modelos para respostas dicotômicas é feita de duas formas: primeiro de acordo com a quantidade de parâmetros associados aos itens 1, 2, 3, 4 ou 5 (ou abreviadamente 1P, 2P, 3P, 4P e 5P, respectivamente) os quais serão indexados por i e posteriormente, de acordo com a função de ligação em que se destacam a logística e a normal como sendo as mais utilizadas. Especificamente,

- (i) o modelo 1P considera somente a dificuldade do item;
- (ii) o modelo 2P considera a dificuldade e a discriminação; e
- (iii) o modelo 3P considera a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos que possuem baixa habilidade (acerto casual/chute).
- (iv) o modelo 4P considera além dos parâmetros abordado no modelo 3P um parâmetro que limita superiormente a probabilidade de acerto de um item.
- (v) o modelo 5P considera além dos parâmetros do 4P um parâmetro de assimetria na função de resposta do item.

No que segue, focaremos nos modelos de 1P, 2P e 3P uma vez que os modelos vistos nesse trabalho que levam em consideração o *speededness* podem ser vistos como adaptações dos mesmos.

Para isso apresentaremos inicialmente o modelo de 3 parâmetros uma vez que os demais modelos de 2 e 1 parâmetro serão vistos como casos particulares do mesmo. O modelo da TRI de três parâmetros é definido de modo que a probabilidade do j -ésimo indivíduo responder corretamente o item i dado sua habilidade, θ_j , e os respectivos parâmetros associados aos itens, a_i , b_i e c_i é dada por

$$\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i)\Psi(Da_i(\theta_j - b_i)) \quad 0 \leq c_i \leq 1 \quad a_i > 0, \quad b_i, \theta_j \in \mathbb{R} \quad \text{e} \quad (1.1)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$, em que Ψ geralmente corresponde à função de distribuição acumulada (fda) da normal padrão ou à fda da distribuição logística padrão.

No caso em que $D = 1$ e Ψ corresponde à fda da normal padrão, ou seja,

$$\Psi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad x \in \mathbb{R},$$

temos o modelo que na literatura é conhecido como modelo de ogiva normal de três parâmetros, que abreviaremos por 3PN. Esse modelo surgiu da necessidade de [Lord \(1952\)](#) em modelar a probabilidade de resposta correta diante da possibilidade de acertos casuais.

Quando Ψ corresponde a fda da distribuição logística padrão, ou seja,

$$\Psi(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}, \quad x \in \mathbb{R},$$

temos o modelo que na literatura é conhecido como modelo logístico de três parâmetros, o qual abreviaremos por 3PL ([Birnbaum, 1968](#)). O modelo logístico de três parâmetros corresponde ao modelo mais utilizado na TRI atualmente, sendo esse o modelo utilizado no ENEM de acordo com o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira).

A relação existente entre $\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i)$ e os parâmetros do modelo 1.1 é o que chamamos de *curva característica do item*, abreviada simplesmente por CCI. A representação gráfica da CCI possui sempre forma de “S” conforme veremos adiante em algumas ilustrações, e a mesma nos auxiliará a justificar os respectivos nomes atribuídos a cada um dos parâmetros.

No decorrer do texto consideramos $\mathbf{a} = (a_1, \dots, a_I)$, $\mathbf{b} = (b_1, \dots, b_I)$, $\mathbf{c} = (c_1, \dots, c_I)$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)$, ou seja, usaremos a letra em negrito para fazer referência ao vetor composto por todos os parâmetros dos correspondentes itens e das variáveis latentes, respectivamente. Além disso, usaremos $\mathbf{y}$ para nos referirmos à matriz de respostas observadas do teste.

1.1.1 Interpretação dos modelos

As interpretações feitas a seguir correspondem às interpretações do modelo 1.1 independentemente da função de ligação utilizada. No caso em que apresentarmos ilustrações de curvas características dos itens estaremos sempre apresentando as mesmas considerando a ligação logito.

Fator de escala

A variável D corresponde ao que chamamos de *fator de escala*, usualmente seu valor é 1, porém utiliza-se $D = 1,7$ no modelo 3PL quando deseja-se que o mesmo forneça resultados semelhantes aos do modelo de ogiva normal, pois neste caso a curva da logística aproxima-se da curva normal. A ilustração do que foi comentado encontra-se na Figura (1.1) onde para ambos os casos foi fixado que $a = 1$, $b = 1$ e $c = 0.2$.

Figura 1.1: CCI para o modelo 3PL considerando $D = 1,7$ (linha) e para o modelo 3PN (tracejado).

Traço latente (Proficiência ou Habilidade)

A variável θ_j , corresponde ao que chamamos de variável latente em estatística, ou seja, uma variável que não pode ser mensurada diretamente no indivíduo. É importante ressaltar que a variável θ_j está inteiramente ligada a variável resposta Y , por exemplo ao utilizarmos o modelo 1.1 em um questionário do tipo satisfeito ou insatisfeito com determinados produtos de uma empresa a variável θ_j mensura o grau de satisfação do j -ésimo usuário com os produtos da empresa, enquanto se utilizarmos o modelo em teste de múltipla escolha a variável θ_j irá mensurar a proficiência do indivíduo com relação ao assunto abordado no teste. Assim, ao utilizarmos o modelo 1.1 bem como os demais modelos da TRI citados, em testes sobre um determinado tema, a variável θ_j é comumente chamada de *proficiência* ou *habilidade* do j -ésimo indivíduo uma vez que a mesma mensura a habilidade do respondente.

Probabilidade de acerto Casual

É fácil ver que se $a_i > 0$ e $\theta_j \rightarrow \infty$ então $\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i) \rightarrow 1$, em outras palavras indivíduos com habilidade muito alta tendem a acertar o item. Além disso, se $a_i > 0$ e $\theta_j \rightarrow -\infty$ então $\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i) \rightarrow c_i$. Desse modo, c_i é um parâmetro que serve como limitante inferior da probabilidade de acerto do item i , por isso o mesmo é chamado de probabilidade de acerto casual, uma vez que representa a probabilidade de indivíduos com uma habilidade muito baixa responderem corretamente o item i (quando o modelo em questão está sendo usado para mensurar a proficiência de um indivíduo). Quando não é possível responder ao acaso considera-se $c_i = 0$. Na Figura (1.2) temos a ilustração gráfica das curvas características de dois itens, a curva tracejada, por exemplo, pode corresponder à probabilidade de acerto casual de um item que possui somente duas alternativas $c = 1/2 = 0,5$, enquanto a curva do tipo linha pode corresponder a probabilidade de acerto casual de um item que possui 5 alternativas $c = 1/5 = 0,2$.

Figura 1.2: CCI para o modelo 3PL considerando $D = 1,7$ (linha) e para o modelo 3PN (tracejado).

Parâmetro de discriminação

O valor do parâmetro a_i é proporcional à inclinação da CCI no ponto b_i (ver [Andrade et al. \(2000\)](#)) indicando que itens com valores maiores para a_i possuem maior poder de “discriminação”, uma vez que a CCI está mais inclinada, daí o fato de a_i ser chamado de parâmetro de inclinação do item ou parâmetro de discriminação. A Figura (1.3) (CCI pontilhada) mostra que, para $a = 1$, $b = 0$ e $c = 0.2$, se um indivíduo possui $\theta = -1$ e outro possui $\theta = 1$ então suas respectivas probabilidades de acerto são 0,41 e 0,78. Esses mesmos indivíduos sob as condições da CCI tracejada, ou seja, $a = 2$, $b = 0$ e $c = 0.2$, tem como probabilidades de respostas corretas 0,29 e 0,9, respectivamente. Notemos então que a CCI tracejada corresponde a uma curva que discrimina melhor as habilidades dos indivíduos que estão à direita do ponto b_i dos que estão à esquerda do ponto b_i em termos da probabilidade de acerto, uma vez que $0,78 - 0,41 = 0,37 < 0,51 = 0,9 - 0,29$. Em geral, itens com maiores valores de a_i terão maior poder de discriminação. Porém não é interessante que o a_i seja muito alto, pois teríamos uma CCI muito íngreme de modo a diferenciarmos os indivíduos em apenas dois grupos, os que possuem habilidade maior que b_i e os que possuem habilidade menor que b_i . Itens com o valor de a_i pequeno também não são interessantes do ponto de vista que teríamos a CCI pouco acentuada de modo que indivíduos com a habilidade muito distintas teriam praticamente a mesma probabilidade de acerto.

Figura 1.3: CCI para o modelo 3PL para diferentes valores de a_i .

Parâmetro de Dificuldade

A interpretação do parâmetro b_i , como dificuldade do item i , é mais simples. O parâmetro b_i no modelo 1.1 corresponde à habilidade necessária para uma probabilidade de acerto $(1 + c_i)/2$. Desse modo, quanto maior for o valor de b_i maior é a dificuldade do item i . Cabe ressaltar que diante da diferença $\theta_j - b_i$ presente no modelo temos que o parâmetro b_i na mesma escala (métrica) da variável latente θ_j . Na Figura (1.4) temos a ilustração gráfica da CCI de dois itens com parâmetros de dificuldade distintos. Observando a Figura (1.4) notamos que fixado um valor arbitrário para a probabilidade de acerto, o valor θ_j que originou esta probabilidade é sempre maior na CCI tracejada que possui $b = 1$, indicando que o item correspondente a esta CCI é mais difícil, ou seja, possui maior valor de b_i .

Figura 1.4: CCI para o modelo 3PL considerando $D = 1,7$ (linha) e para o modelo 3ON (tracejado).

1.1.2 Casos particulares dos modelos de 3 parâmetros

No modelo de ogiva normal (logístico) de três parâmetros (1.1), considerando que não existe a possibilidade de acerto casual, ou seja, $c_i = 0 \forall i$, obtemos o modelo de ogiva normal (logístico) de dois parâmetros, o qual é dado por

$$\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i) = \Psi(Da_i(\theta_j - b_i)), \quad (1.2)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$.

No conjunto de dados que trabalharemos adiante no Capítulo 5 faremos uso do modelo logístico de dois parâmetros o qual denotaremos por modelo 2PL.

No modelo de dois parâmetros (1.2) (ogiva normal ou logístico), considerando que todos os itens tem o mesmo poder de discriminação, ou seja, $a_i = 1 \forall i$, obtemos o modelo de um parâmetro (ogiva normal ou logístico, respectivamente), o qual é dado por

$$\mathbb{P}(Y_{ij} = 1 | \theta_j, b_i) = \Psi(D(\theta_j - b_i)), \quad (1.3)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$.

O modelo logístico de um parâmetro é conhecido na literatura como modelo 1PL ou modelo de Rasch em homenagem a George Rasch devido o mesmo ter sido o primeiro a fazer uso dele.

Uma medida bastante utilizada, juntamente com a CCI, é a *Função de Informação do Item*. Essa medida nos diz o quanto um item (ou teste) possui de informação para a medida da habilidade. Para o modelo de 3P da TRI ela é definida como:

$$IF_i(\theta) = \frac{\left[\frac{d}{d\theta} \mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i) \right]^2}{\mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i) [1 - \mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i)]}. \quad (1.4)$$

Para o modelo 3PL pode-se mostrar que é possível reescrever 1.4 da seguinte forma:

$$IF_i(\theta) = D^2 a_i^2 \frac{[1 - \mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i)]}{\mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i)} \left[\frac{\mathbb{P}(Y_{ij} = 1 | \theta, a_i, b_i, c_i) - c_i}{1 - c_i} \right]^2$$

onde podemos observar que a informação é maior quando

- (i) b_i se aproxima de θ ;
- (ii) quanto maior for a_i ;
- (iii) quanto mais próximo de zero c_i estiver.

Dada a independência entre os itens, a *função de informação do teste* é definida como sendo a soma das informações obtidas de cada item, ou seja,

$$IF(\theta) = \sum_{i=1}^I IF_i(\theta).$$

1.1.3 Suposições do modelo

Os modelos apresentados anteriormente foram construídos com base em algumas suposições as quais apresentaremos a seguir.

Os mesmos pressupõem que o conjunto de itens está medindo um único traço latente, essa suposição é chamada de *unidimensionalidade*. A princípio havia a necessidade de impor a suposição de *independência condicional*, também chamada de *independência local*, utilizada no processo de estimação dos parâmetros. Essa suposição diz que dado a habilidade do indivíduo e os demais parâmetros correspondentes aos itens, as respostas aos diferentes itens são independentes.

Hambleton *et al.* (1991) verificaram que para o modelo 1.1 a unidimensionalidade implica em independência condicional, tornando a última suposição desnecessária.

É natural esperar que o modelo proposto satisfaça o princípio de que indivíduos com maior habilidade tenham uma maior probabilidade de acerto. Sendo assim, itens com valores do parâmetro de discriminação negativo, $a_i < 0$, não são esperados, pois nesse caso a CCI associada a este item é monótona decrescente em θ_j . Por exemplo, na Figura (1.5 (a)) temos a CCI para $a = -1$, $b = 1$ e $c = 0,2$ onde notamos que a CCI é monótona decrescente, ou seja, indivíduos com maior habilidade possuem menor probabilidade de acerto. Isso ocorre sempre que $a_i < 0$. Logo, na área educacional devemos supor que os a_i 's são positivos como, por exemplo, na Figura (1.5 (b)).

Figura 1.5: CCI considerando o parâmetro de discriminação negativo (esquerda) e positivo (direita).

1.2 Métodos de Estimação

Conforme já mencionado os modelos da TRI, aqui apresentados, possuem dois tipos de parâmetros, os correspondentes aos itens ($\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$) e os correspondentes aos indivíduos (as variáveis latentes $\boldsymbol{\theta}$). No processo de estimação dos parâmetros há três casos a se considerar: (a) estimar as habilidades conhecendo os parâmetros dos itens; (b) estimar os parâmetros dos itens conhecendo os parâmetros dos indivíduos, e (c) estimar todos os parâmetros conjuntamente. Esse último caso corresponde ao que mais ocorre em situações práticas.

Conforme Andrade *et al.* (2000), em qualquer uma das situações (a) e (b) pode-se usar o método da máxima verossimilhança através da aplicação de algum processo iterativo, como o algoritmo de Newton-Raphson (Keller e Isaacson, 1966) ou “Scoring” de Fisher (Rao, 2009), sendo possível também fazer uso de alguns procedimentos Bayesianos (Mislevy, 1986).

O caso (c) em que desejamos estimar tanto os parâmetros dos itens quanto dos indivíduos é um pouco mais complicado. No primeiro caso em que temos os parâmetros dos itens, $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$, e queremos estimar as habilidades, $\boldsymbol{\theta}$, essas são estimadas na métrica (escala) dos parâmetros dos itens. No segundo caso, em que temos os parâmetros dos indivíduos, os parâmetros dos itens são estimados na métrica dos parâmetros dos indivíduos. No caso em que desejamos estimar todos os parâmetros nos deparamos com o problema chamado de *falta de identificabilidade*. Formalmente temos a seguinte definição de não-identificabilidade apresentada em Matos (2008).

Definição 1.1. (Parâmetro e verossimilhança não-identificável) Considere $\boldsymbol{\Psi} = (\phi_1, \phi_2) \in \Theta$ um vetor de parâmetros particionados em que ϕ_1 e ϕ_2 possivelmente também são vetores e $V(\boldsymbol{\Psi}; \mathbf{y}) = \mathbb{P}(Y = \mathbf{y}; \boldsymbol{\Psi})$ a função de verossimilhança de $\boldsymbol{\Psi}$. Dizemos que ϕ_1 é um parâmetro

identificável e que ϕ_2 é não-identificável se para todo conjunto de dados amostrais $\mathbf{y}$ tais que,

$$V(\Psi^{(1)}; \mathbf{y}) = V(\Psi^{(2)}; \mathbf{y}),$$

tivermos que $\phi_1^{(1)} = \phi_1^{(2)}$ e $\phi_2^{(1)} \neq \phi_2^{(2)}$, em que $\Psi^{(1)} = (\phi_1^{(1)}, \phi_2^{(1)})$ e $\Psi^{(2)} = (\phi_1^{(2)}, \phi_2^{(2)})$. Neste caso, dizemos ainda que a verossimilhança $V(\Psi; \mathbf{y})$ é não-identificável.

Observação: Não entraremos em detalhes sobre a distinção de identificabilidade do ponto de vista clássico e do ponto de vista Bayesiano. Conforme pode ser visto em [Matos \(2008\)](#), o conceito de não-identificabilidade de um modelo está relacionado somente à verossimilhança e não com a especificação das distribuições *a priori*, sendo assim, trabalharemos somente com essa definição.

A não-identificabilidade dos modelos 3P (logito e probito) ocorre no processo em que é necessário estimar todos os parâmetros porque mais de um conjunto de parâmetros produzem a mesma verossimilhança. De fato, se $\theta_j^* = \alpha\theta_j + \beta$, $b_i^* = \alpha b_i + \beta$, $a_i^* = a_i/\alpha$ e $c_i^* = c_i$, onde α e β são constantes reais com $\alpha > 0$, então

$$\begin{aligned} \mathbb{P}(Y_{ij} = 1 | \theta_j^*, a_i^*, b_i^*, c_i^*) &= c_i^* + (1 + c_i^*)\Psi(Da_i^*(\theta_j^* - b_i^*)) \\ &= c_i + (1 + c_i)\Psi\left(D\frac{a_i}{\alpha}(\alpha\theta_j + \beta - (\alpha b_i + \beta))\right) \\ &= c_i + (1 + c_i)\Psi(Da_i(\theta_j - b_i)) \\ &= \mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i) \end{aligned}$$

e conseqüentemente,

$$\mathbb{P}(Y_{ij} = 0 | \theta_j^*, a_i^*, b_i^*, c_i^*) = \mathbb{P}(Y_{ij} = 0 | \theta_j, a_i, b_i, c_i).$$

Segue então que,

$$\mathbb{P}(Y_{ij} = y_{ij} | \theta_j^*, a_i^*, b_i^*, c_i^*) = \mathbb{P}(Y_{ij} = y_{ij} | \theta_j, a_i, b_i, c_i) \quad \forall i = 1 \dots I \quad \text{e} \quad j = 1 \dots J.$$

Assim,

$$\begin{aligned} V(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*, \boldsymbol{\theta}^*; \mathbf{y}) &= \mathbb{P}(\mathbf{Y} = \mathbf{y} | \boldsymbol{\theta}^*, \mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*) \\ &= \prod_{j=1}^J \prod_{i=1}^I \mathbb{P}(Y_{ij} = y_{ij} | \theta_j^*, a_i^*, b_i^*, c_i^*) \\ &= \prod_{j=1}^J \prod_{i=1}^I \mathbb{P}(Y_{ij} = y_{ij} | \theta_j, a_i, b_i, c_i) \\ &= \mathbb{P}(\mathbf{Y} = \mathbf{y} | \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) \\ &= V(\mathbf{a}, \mathbf{b}, \mathbf{c}, \boldsymbol{\theta}; \mathbf{y}). \end{aligned}$$

Analogamente, é fácil verificar que o problema de falta de identificabilidade também ocorre nos modelos 2P ($\theta_j^* = \alpha\theta_j + \beta$, $b_i^* = \alpha b_i + \beta$) e 1P ($\theta_j^* = \alpha\theta_j + \beta$, e $a_i^* = a_i/\alpha$)

Dessa forma, devido ao problema de falta de identificabilidade do modelo é preciso definir primeiramente uma métrica para poder estimar todos os parâmetros simultaneamente. O método mais comum é especificar que os traços latentes, habilidades θ_j , seguem uma distribuição normal padrão. No entanto, outras abordagens podem ser utilizadas, por exemplo, [Fox \(2010\)](#) cita que a métrica pode ser identificada através da fixação de um parâmetro de discriminação e dificuldade ou adicionando restrições sobre os parâmetros dos itens. Uma possível restrição é fazer com que a soma das dificuldades dos itens e o produto dos valores desses parâmetros sejam iguais a, por exemplo, zero e um, respectivamente. Outra restrição seria, a fixação da média e da variância da distribuição das habilidades da população. A escolha da restrição geralmente é feita de modo a tornar mais

simples a implementação do algoritmo de estimação do modelo. Fox (2010) ressalta que a escolha das restrições de identificação pode levar a problemas numéricos nas estimativas dos parâmetros.

Há diversas metodologias possíveis a se seguir para realizar a estimação, mas basicamente podemos agrupá-las em duas classes. A primeira consiste do uso de métodos da inferência clássica e a outra que consiste do uso de métodos da inferência Bayesiana.

Conforme Bazan (2005), a estimação clássica é dominada pelo uso do método da máxima verossimilhança através de suas derivações que são: máxima verossimilhança conjunta (MVCJ), máxima verossimilhança marginal (MVM) e máxima verossimilhança condicional (MVC). Esses três métodos diferem entre si pela forma que abordam o problema dos θ 's serem desconhecidos. No método da MVCJ utilizam-se valores iniciais para as estimativas das habilidades tratando as mesmas como conhecidas (por exemplo, escores padronizados) e estima-se os parâmetros dos itens, posteriormente, conhecido os parâmetros dos itens, fixa-se os mesmos como verdadeiros na verossimilhança do modelo e estima-se as habilidades. O método da MVC modela a probabilidade dos vetores de respostas com base nos escores absolutos gerados por diversos padrões de respostas, em outras palavras ele condiciona a probabilidade de respostas corretas em função de estatísticas suficientes mínimas, mais detalhes sobre esse método bem como um exemplo prático pode ser encontrado em Andersen e Madsen (1977). Dentre essas três abordagens, a mais utilizada é a MVM de Bock e Lieberman (1970), uma vez que apresenta estimativas consistentes para os parâmetros dos itens. O MVM difere dos demais métodos citados por modelar a probabilidade dos vetores de respostas como vinda de uma população com distribuição para θ conhecida. O algoritmo de estimação dos parâmetros do modelo 1.1 usando o método da MVM está completamente descrito em Andrade *et al.* (2000) e encontra-se implementado em diversos *softwares* computacionais, dentre os quais podemos citar o BILOG, o R. PROJECT por meio de pacotes relacionados com a TRI, o MULTILOG e o SAS.

Por sua vez o processo de estimação usando a metodologia Bayesiana, assim como no método da Máxima Verossimilhança, pode ocorrer de duas formas, com ou sem o uso de métodos de Monte Carlo (MC). O uso de métodos Bayesianos apresenta algumas vantagens quando comparado ao método clássico. No método clássico há a possibilidade das estimativas dos parâmetros dos itens caírem fora dos intervalos desejados, como por exemplo, $a_i < 0$ ou $c_i \notin [0, 1]$, bem como termos problemas para estimar os parâmetros de itens que foram respondidos incorretamente por todos os indivíduos e estimar a habilidade dos indivíduos que responderam incorretamente ou corretamente a todos os itens. Esses problemas não ocorrem com o uso da abordagem Bayesiana.

Com relação ao uso da abordagem Bayesiana sem a utilização de técnicas de MC o método mais utilizado é o da estimação Bayesiana Marginal, o qual foi proposto por Mislevy (1986) e que corresponde à generalização da proposta de Bock e Aitkin (1981). Lembrando que, por se tratar de uma generalização da MVM (mesmo fazendo uso de EM), esse método também consiste de duas etapas. Além disso, devido à falta de identificabilidade do modelo de 3P a métrica deve ser especificada previamente. Nesse método, são especificadas distribuições *a priori* para os parâmetros de interesse e em seguida é obtida a distribuição *a posteriori* conjunta. Posteriormente, para estimarmos os parâmetros dos itens marginalizamos a verossimilhança conjunta com relação ao traço latente e usamos como estimativas para os mesmos a moda *a posteriori* que pode ser obtida via métodos de otimização.

As estimativas das habilidades nesse método são obtidas com base na distribuição *a posteriori* marginal dos θ 's condicional aos parâmetros dos itens (estimados anteriormente), considerando como estimador a média ou a moda *a posteriori*.

Com o recente avanço computacional o uso de técnicas de Monte Carlo fazendo uso de cadeias de Markov, ou MCMC abreviadamente, tem se tornado cada vez mais comum. Nessa metodologia pode ser feito uso de Dados Aumentados com amostrador de Gibbs (DAGS) para estimar os parâmetros do modelo 3P ver, por exemplo, Fu *et al.* (2009) ou fazer uso de Metropolis-Hasting dentro do Amostrado de Gibbs Patz e Junker (1999a) e Patz e Junker (1999b) o qual usaremos adiante em nosso estudo.

Tanto a abordagem clássica quanto a abordagem Bayesiana possuem suas vantagens e limitações,

a seguir apresentamos um resumo das mesmas conforme [Andrade *et al.* \(2000\)](#).

Estimação dos parâmetros dos itens

Ao utilizarmos a abordagem clássica por meio da MVM obtemos estimativas consistentes para os parâmetros dos itens, porém as mesmas não estão definidos para os itens que forem respondidos corretamente ou incorretamente por todos os indivíduos. Além disso, a MVM é bastante trabalhosa computacionalmente e pode apresentar problemas nas estimativas dos parâmetros dos c_i 's caso não haja uma quantidade de indivíduos suficientemente grande respondendo o teste. O uso da abordagem Bayesiana por sua vez tem a vantagem dos parâmetros dos itens estarem definidos para qualquer tipo de padrão de resposta, porém é mais trabalhoso computacionalmente que a MVM e tem a necessidade da especificação das distribuições *a priori* para os parâmetros desconhecidos.

Estimação dos traços latente ou habilidades

A estimação dos traços latentes usando a máxima verossimilhança tem a vantagem de que para testes, relativamente, longos as estimativas são não viciadas, porém para indivíduos que acertam ou erram todos os itens as estimativas das habilidades dos mesmos não estão definidas, ao contrário do uso da metodologia Bayesiana (estimando via média ou moda *a posteriori*), que apesar de produzir estimadores que são viciados, as habilidades podem ser estimadas para qualquer padrão de respostas. A metodologia Bayesiana também tem as desvantagens de requerer cálculos mais complexos e a necessidade de especificação de uma distribuição *a priori* para θ .

1.3 Organização do Trabalho

No que segue o trabalho será organizado da seguinte maneira. No Capítulo 2 faremos uma breve revisão sobre os modelos da TRI para modelar dados na presença da restrição de tempo, bem como apresentamos alguns textos que servem de referência em estudos envolvendo a limitação de tempo. Entre os modelos que apresentados estarão o Modelo Híbrido de [Yamamoto e Everson \(1997\)](#), o Modelo Mistura de Rasch Multi-Classe proposto por [Mroch *et al.* \(2005\)](#) e o Modelo de resposta ao item com processo de mudança de resposta gradual proposto por [Wollack e Cohen \(2005\)](#). Este último será o objeto principal de estudo deste trabalho. Ainda no Capítulo 2, apresentamos o processo de estimação clássico proposto por [Goegebeur *et al.* \(2008\)](#) para estimar os parâmetros do modelo em questão. No Capítulo 3 construímos o processo de estimação dos parâmetros do modelo com processo de mudança de resposta gradual usando a abordagem totalmente Bayesiana, mas antes, faremos um breve estudo de cópulas e mostraremos como realizar a estimação da matriz de covariância do modelo normal na presença de restrições. No Capítulo 4 faremos um breve estudo de simulação onde estimamos os parâmetros do modelo proposto por [Wollack e Cohen \(2005\)](#) considerando inicialmente que as distribuições dos traços latentes são independentes e posteriormente estimamos os parâmetros considerando que os mesmos possuem uma estrutura de dependência entre as distribuições *a priori* dos traços latentes e finalizando fazendo uma comparação entre os dois métodos utilizados. No Capítulo 5 faremos então uma aplicação do conteúdo estudado nos capítulos anteriores num conjunto de dados reais provenientes de um exame realizado no Perú em 2007 e 2008.

Capítulo 2

Modelos da TRI que levam em consideração o *speededness*

Os modelos (1.1), (1.2) e (1.3) correspondem aos modelos usuais da TRI para respostas dicotômicas. Os mesmos possuem uma importante característica em comum, tal característica corresponde ao fato desses modelos assumirem que os respondentes possuem tempo suficiente para responder todos os itens no período de aplicação do teste, o que pode não ocorrer em situações reais. Na prática ao realizarmos um teste dispomos de uma quantidade limitada de tempo, que pode ser insuficiente, de modo a vir alterar a probabilidade de respostas corretas dos respondentes para os itens próximo ao fim do teste. Por exemplo, é natural que alguns respondentes venham a não responder, ou responder ao acaso, itens próximo do fim do teste, mesmo sabendo respondê-los, devido à ausência de tempo. A ocorrência de tais situações não é levada em conta nos modelos já apresentados, fazendo com que o uso desses modelos venham a ser questionado. Nessas situações podemos estar subestimando as habilidades de alguns indivíduos ou superestimando a dificuldade de alguns itens conforme citam Rost (1990), Yamamoto e Everson (1997) e Bazan *et al.* (2010). Schmitt *et al.* (2010) mencionam que a presença de *speededness* faz com que seja violada, nos modelos usuais, a hipótese de independência entre as respostas aos diferentes itens. Diante disso é necessário que o resultado não seja baseado exclusivamente no indicador de habilidade, mas que contenha também informações sobre a influência do efeito *speededness* nas respostas atribuídas. Nesse contexto, alguns autores, dentre os quais citamos Bolt *et al.* (2002), Yamamoto e Everson (1997) e Goegebeur *et al.* (2008), tem proposto o uso de modelos alternativos que levam em conta o efeito de *speededness* ocasionado pela limitação de tempo. A seguir apresentamos alguns desses modelos. Os modelos que citaremos e que levam em conta o efeito de *speededness* são baseados na suposição de que os itens são respondidos **sequencialmente**, por isso usualmente esses modelos são aplicados em testes adaptativos informatizados (TAI) também conhecidos como CAT (*computerized adaptive testing*).

2.1 Modelo Híbrido

De acordo com Yamamoto e Everson (1997), quando testes possuem uma restrição de tempo (o que ocorre tanto no TOEFL como nos demais testes em geral), padrões de respostas são frequentemente observados no final dos testes. Isso ocorre quando os indivíduos que respondem os itens de forma mais lenta, que os demais, ficam sem tempo perto do fim do teste e começam a responder ao acaso os itens remanescentes. Um exemplo de padrão de respostas é selecionar uma única alternativa para todos os itens a partir da presença do efeito de *speededness*. A menos que o algoritmo do padrão das respostas seja óbvio, tal como na situação mencionada, é muito difícil saber determinar se as respostas seguem algum padrão. O modelo híbrido de Yamamoto e Everson (1997) busca incluir no modelo a possibilidade de que, a partir de um determinado item, as probabilidades de acertos sejam modeladas via acerto casual.

O modelo proposto pelos autores é tal que a probabilidade de um indivíduo j responder corretamente o item i é dada por

$$\mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i, k_j) = (1 + \exp(-a_i(\theta_j - b_i)))^m c_i^{m+1}, \quad (2.1)$$

onde k_j indica o último item respondido pelo indivíduo j sem efeito de aceleração (via modelo logístico de dois parâmetros), $m = -1$ quando $i \leq k_j$ e $m = 0$ quando $i > k_j$ (modelo de resposta casual). y_{ij} é a resposta dicotômica (0=errada, 1=correta) do j -ésimo indivíduo ao item i , a_i , b_i e c_i representam, respectivamente, os parâmetros de discriminação, dificuldade e probabilidade de acerto casual do item i , enquanto θ_j corresponde ao traço latente do j -ésimo respondente.

Nesse modelo assume-se que:

- 1) sob uma estratégia padrão de resposta, a probabilidade condicional de uma resposta correta independe da habilidade;
- 2) a resposta de cada indivíduo a um item pode ser caracterizada ou por um modelo 2PL ($m = -1$) ou por um modelo de respostas ao acaso ($m = 0$); e
- 3) dado o traço latente do respondente, os parâmetros dos itens e a estratégia de resposta à independência condicional está assegurada, ou seja, a resposta aos diferentes itens são condicionalmente independentes.

Basicamente esse modelo nos diz que a partir de um item k_j o indivíduo j passa a sofrer efeito da restrição de tempo deixando de usar a sua habilidade e passando a responder somente via um certo padrão de respostas (ou estratégia adotada pelo mesmo, a qual é modelada via acerto casual).

Com relação ao processo de estimação conjunta, nota-se que esse modelo também possui problemas de identificabilidade sendo necessário a fixação de uma métrica, no caso, assumindo que $\theta_j \sim N(0, 1)$. O método de estimação utilizado por Yamamoto e Everson (1997) para estimar os parâmetros do modelo híbrido foi o de máxima verossimilhança marginal dado por Bock e Aitkin (1981) o qual descreveremos brevemente a seguir.

Considere $\mathbf{y}_j = (y_{1j}, \dots, y_{Ij})$ como sendo o vetor de respostas do j -ésimo indivíduo assim,

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{a}, \mathbf{b}, \mathbf{c}, k_j) &= \prod_{i=1}^{k_j} \mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, k_j)^{y_{ij}} (1 - \mathbb{P}(Y_{ij} = 1 | \theta_j, a_i, b_i, k_j))^{1-y_{ij}} \\ &\times \prod_{i=k_j+1}^I c_i^{y_{ij}} (1 - c_i)^{1-y_{ij}}. \end{aligned}$$

Segue então que a probabilidade marginal do vetor de respostas do j -ésimo indivíduo, $\mathbf{y}_j$, dado os parâmetros dos itens é

$$\mathbb{P}(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{a}, \mathbf{b}, \mathbf{c}) = \sum_{k_j} \int_{\theta_j} \mathbb{P}(\mathbf{Y}_j = \mathbf{y}_j | \theta_j, \mathbf{a}, \mathbf{b}, \mathbf{c}, k_j) f(\theta_j | k_j) d\theta_j f(k_j),$$

em que $f(\theta_j | k)$ é a probabilidade condicional de θ_j dado o ponto de mudança k_j , e $f(k_j)$ é a distribuição marginal do ponto de mudança de estratégia do indivíduo j da **população**. Desse modo a verossimilhança marginal dos parâmetros dos itens dado a matriz de respostas observadas $\mathbf{y}$ dos J respondentes aos I itens é

$$V(\mathbf{a}, \mathbf{b}, \mathbf{c}; \mathbf{y}) = \prod_{j=1}^J \mathbb{P}(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{a}, \mathbf{b}, \mathbf{c}). \quad (2.2)$$

De posse da verossimilhança marginal dos parâmetros dos itens, as estimativas dos mesmos são obtidas maximizando (2.2). As integrais envolvidas são resolvidas via método da quadratura e a maximização é feita via Newton Raphson uma vez que as integrais envolvidas em (2.2) não

possuem forma fechada. Obtido as estimativas dos parâmetros dos itens, [Yamamoto e Everson \(1997\)](#) propõem a estimação dos parâmetros associados aos indivíduos via uso da média *a posteriori*.

Um método de estimação usando a abordagem totalmente Bayesiana para o modelo híbrido é proposto em [Mroch et al. \(2005\)](#). Considerando que haja r possíveis pontos de *speededness* no teste (ou seja, pontos a partir dos quais os indivíduos passam a responder via acerto casual) os mesmos propõem a criação de $1 < r \leq I$ classes. Denotando por π_* a proporção de indivíduos na classe $*$ onde, naturalmente, $\sum_{i=1}^r \pi_i = 1$, os mesmos propõem uma estimação Bayesiana especificando as seguintes distribuições *a priori*.

$$\begin{aligned} b_i &\sim N(0, 1) \\ \theta_j &\sim N(0, 1) \\ c_j &\sim \text{Cat}(\pi_1, \dots, \pi_r), \text{ e} \\ (\pi_1, \dots, \pi_r) &\sim \text{Dir}(1, \dots, 1), \end{aligned}$$

em que *Cat*¹ e *Dir*² correspondem, respectivamente, às distribuições Categórica e Dirichlet.

No entanto, conforme [Goegebeur et al. \(2008\)](#), é improvável que modelar dados de respostas na presença da limitação do tempo seja tão simples assim, ou seja, que cada indivíduo passe diretamente a responder ao acaso após um determinado item. Desse modo, alternativas mais realísticas devem ser consideradas. A seguir veremos algumas alternativas ao modelo de [Yamamoto e Everson \(1997\)](#).

2.2 Modelo Mistura de Rasch Multi-classe

[Bolt et al. \(2002\)](#) adaptaram o modelo mistura de Rasch proposto por [Rost \(1990\)](#) no intuito de diferenciar os indivíduos que sofreram o efeito *speededness* dos que não sofreram. Para isso o modelo proposto pelos mesmos separa os indivíduos em duas classes, chamadas de classes latentes, de acordo com os padrões de respostas. Uma classe é composta pelos indivíduos que não sofreram o efeito *speededness* e a outra pelos indivíduos que sofreram o efeito *speededness* a partir do item k . Cabe ressaltar que, diferentemente do modelo proposto por [Yamamoto e Everson \(1997\)](#), o k nesse modelo será único para todos os indivíduos que responderem na presença de *speededness*.

Para melhor compreendermos as classes latentes considere o exemplo a seguir. Em um teste contendo 30 itens observou-se que um certo conjunto de indivíduos sofreu efeito de *speededness* nos 5 últimos itens. Nesse contexto, pode-se criar duas classes de modo que, $g = 1$ indica a classe com ausência de *speededness* e $g = 2$ indicará a classe de indivíduos que sofreram efeito de *speededness* conforme ilustrado na Tabela 2.1.

Tabela 2.1: Ilustração de classes latentes para o modelo mistura de rasch com *speededness* considerando 30 itens e efeito *speededness* definido para os 5 últimos itens

	Itens													
Classes Latentes	1	2	3	...	22	23	24	25	26	27	28	29	30	
1	0	0	0	...	0	0	0	0	0	0	0	0	0	
2	0	0	0	...	0	0	0	0	1	1	1	1	1	

0 = Item sem *speededness*

1 = Item com *speededness*

No modelo proposto por [Bolt et al. \(2002\)](#) são impostas restrições ordinais sobre os parâmetros de dificuldade dos itens, b_i 's, nas classes latentes de modo a diferenciar a classe que sofreu efeito da limitação de tempo da que não sofreu esse efeito. Basicamente, para os itens iniciais do teste as dificuldades são iguais nas duas classes. Entretanto, para os itens na classe dos indivíduos que sofrem o efeito da limitação de tempo, a partir do surgimento desse efeito a dificuldade dos mesmos

¹Ver Apêndice A.1

²Ver Apêndice A.2

são limitadas a serem maiores do que na classe que não sofreu o efeito da limitação. Considerando k o primeiro item em que os respondentes experimentam o efeito da limitação do tempo, o modelo mistura de Rasch estendido é definido de modo que

$$\mathbb{P}(Y_{ij} = 1 | \theta_j^{(g)}, b_i^{(g)}) = \frac{\exp(\theta_j^{(g)} - b_i^{(g)})}{1 + \exp(\theta_j^{(g)} - b_i^{(g)})}, \quad (2.3)$$

em que,

$$\begin{aligned} b_i^{(1)} &= b_i^{(2)}, \quad \text{para } i < k, \\ b_i^{(1)} &< b_i^{(2)}, \quad \text{para } i \geq k \end{aligned}$$

e $g \in \{1, 2\}$.

Assim, de acordo com Bolt *et al.* (2002) as estimativas de dificuldade dos itens obtidas na classe de indivíduos que não sofreram o efeito da limitação do tempo, $b_i^{(1)}$ são estimativas mais adequadas para os itens de fim de teste do que as estimativas obtidas usando todos os respondentes quando considerado modelo de Rasch. Naturalmente observamos no modelo proposto que para itens que possam ser respondidos possivelmente na presença de *speededness* temos dois parâmetros de dificuldade associados ao mesmo, enquanto que para os demais temos apenas um parâmetro de dificuldade. Como a princípio não sabemos em quais classes os indivíduos estão alocados definimos $\pi^{(g)}$ como sendo a proporção de indivíduos na classe (g), em que $\sum_g \pi^{(g)} = 1$. Por fim, cada classe é associada com um parâmetro μ_{θ_g} o qual corresponde à habilidade média dos indivíduos daquela classe, ou seja, à média de $\theta^{(g)}$ (a variância de $\theta^{(g)}$ é definida como sendo 1). No estudo de simulação feito por Mroch *et al.* (2005) foi imposta ainda uma restrição sobre $\mu_{\theta}^{(g)}$, tal restrição consistiu em considerar $\mu_{\theta}^{(g)}$ como uma função relativa das dificuldades dos itens nas classes que sofreram e que não sofreram o efeito de *speededness*, isto é,

$$\mu_{\theta}^{(g)} = \mu_b - \mu_b^{(g)}, \quad (2.4)$$

em que μ_b é a dificuldade média dos itens na classe sem efeito de *speededness* e $\mu_b^{(g)}$ é a média das dificuldades dos itens na classe, g , que sofreu efeito de *speededness*.

É fácil ver que esse modelo possui problemas de identificabilidade. Afim de evitá-los a alternativa adotada em Mroch *et al.* (2005) foi impor uma restrição sobre os parâmetros de dificuldade dos itens dentro de cada classe, tal restrição é chamada de *condição normalizadora* e consiste em impor que $\sum_i b_i^{(g)} = 0$, o uso da mesma foi feito seguindo um raciocínio análogo ao do modelo estudado em Rost (1990). O uso da condição normalizadora é consistente com as aplicações típicas de modelos e permite que as classes sejam definidas pelos perfis de respostas ao invés da quantidade de acertos.

Esse modelo possui dois pontos falhos. O primeiro ponto falho consiste no fato dele não permitir respondentes distintos sentirem o efeito da limitação do tempo em momentos distintos do teste. O segundo ponto falho está na especificação inicial das classes.

Mroch *et al.* (2005) propuseram então uma extensão do modelo (2.3) criando r classes de indivíduos, uma com indivíduos que não sofreram o efeito da limitação do tempo ($g = 1$) e $r - 1$ classes com indivíduos que sofreram o efeito da limitação do tempo, $g = \{2, \dots, r\}$. A esse modelo os mesmos deram o nome de modelo mistura de Rasch multi-classe. O modelo mistura de Rasch multi-classe é dado por

$$\mathbb{P}(Y_{ij} = 1 | \theta_j^{(g)}, b_i^{(g)}) = \frac{\exp(\theta_j^{(g)} - b_i^{(g)})}{1 + \exp(\theta_j^{(g)} - b_i^{(g)})}, \quad g = 1, 2, \dots, r. \quad (2.5)$$

Nesse modelo restrições também devem ser impostas nos parâmetros dos itens, de modo que nas classes em que o efeito da limitação do tempo surge, devemos ter valores maiores para os parâmetros de dificuldade a partir da presença desse efeito quando comparado com classe sem o efeito dessa

limitação. Ou seja,

$$\begin{aligned} b_i^{(1)} &= b_i^{(2)}, \text{ para } i < k_1 \\ b_i^{(1)} &< b_i^{(2)}, \text{ para } i \geq k_1 \\ &\vdots \\ b_i^{(1)} &= b_i^{(r)}, \text{ para } i < k_r \\ b_i^{(1)} &< b_i^{(r)}, \text{ para } i \geq k_r \end{aligned}$$

Note que o modelo (2.3) corresponde ao modelo mistura de Rasch multi-classe quando consideramos apenas duas classes. O modelo mistura de Rasch multi-classe esbarra nos seguintes problemas: (i) definir quantas classes devemos tomar e (ii) como definir as r classes.

Na Tabela 2.2 temos um exemplo de um teste com 30 itens, similar ao anterior, no qual definimos 6 classes latentes. A primeira obviamente corresponde à classe sem *speededness* enquanto as demais correspondem às classes com *speededness* a partir de itens distintos.

Tabela 2.2: ilustração de classes latentes para o modelo mistura de rasch multiclasse com 6 classes e considerando 30 itens

Classes latentes	Itens												
	1	2	3	...	22	23	24	25	26	27	28	29	30
1	0	0	0	...	0	0	0	0	0	0	0	0	0
2	0	0	0	...	0	0	0	0	0	0	0	0	1
3	0	0	0	...	0	0	0	0	0	0	0	1	1
4	0	0	0	...	0	0	0	0	0	0	1	1	1
5	0	0	0	...	0	0	0	0	0	1	1	1	1
6	0	0	0	...	0	0	0	0	1	1	1	1	1

0 = Item sem *speededness*

1 = Item com *speededness*

Com relação à estimação dos parâmetros vamos fazer referência apenas ao modelo mistura de Rasch multi-classe, uma vez que o modelo em (2.3) é um caso particular do modelo em (2.5). Para realizar a estimação dos parâmetros do modelo (2.5), Mroch *et al.* (2005) fizeram uso de algoritmos *Markov Chain Monte Carlo* (MCMC) e utilizaram o *software* WinBUGS. As distribuições *a priori* especificadas pelos mesmos foram:

$$b_i^{(g)} \sim Normal(0, 1)$$

$$\theta_j^{(g)} \sim N(\mu^{(g)}, 1)$$

$$\mu_{\theta}^{(1)} \sim Normal(0, 1) \text{ (para classe sem speededness) e}$$

$$\mu_{\theta}^{(2)}, \dots, \mu_{\theta}^{(r)} \text{ são funções dos parâmetros dos itens conforme (2.4)}$$

$$c_j \sim Cat(\pi_1, \dots, \pi_r), \text{ onde } c_j = \{1, 2, \dots, r\} \text{ é um parâmetro de associação de classe, e}$$

$$(\pi_1, \dots, \pi_k) \sim Dir(1, \dots, 1).$$

No que segue veremos o modelo de resposta ao item com processo de mudança de resposta gradual. Este modelo busca suprir as falhas mencionadas no modelo mistura de Rasch multi-classe não sendo necessário a criação de classes *speededness* previamente e, além disso, também permite uma aceleração gradativa no processo de respostas dos itens. Ou seja, o modelo que estudaremos agora corresponde de certa forma a uma extensão tanto do modelo híbrido quanto do modelo mistura de Rasch multi-classe.

2.3 Modelo de Resposta ao Item com Processo de Mudança de Resposta Gradual

Proposto por [Wollack e Cohen \(2005\)](#), o modelo de resposta ao item com processo de mudança gradual que apresentaremos a seguir foi utilizado inicialmente pelos mesmos como um gerador de respostas de indivíduos na presença de *speededness*. Posteriormente, [Goegebeur et al. \(2008\)](#) fizeram uso do modelo para estimar os traços latentes de um conjunto de indivíduos que realizaram um teste na presença da limitação do tempo. O modelo de resposta ao item com processo de mudança de resposta gradual (3PLMG) considera que os indivíduos respondem o exame sequencialmente até que, em certo ponto, passam a sofrer os efeitos da limitação do tempo; o que faz com que o respondente não consiga usar toda a sua habilidade e, conseqüentemente, a partir desse ponto cada vez ele usará menos a habilidade que possui.

Considerando j como índice do respondente, $j = 1, \dots, J$ e i como índice do item, $i = 1, \dots, I$. O modelo 3PLMG é definido como

$$Y_{ij}|\theta_j, \eta_j, \lambda_j, a_i, b_i, c_i \sim \text{Bern}(\pi_{ij}), \quad (2.6)$$

em que

$$\pi_{ij} = \mathbb{P}(Y_{ij} = 1|\theta_j, \eta_j, \lambda_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))} P_i(\eta_j, \lambda_j). \quad (2.7)$$

$P_i(\eta_j, \lambda_j) = \min \left\{ 1, \left[1 - \left(\frac{i}{I} - \eta_j \right) \right]^{\lambda_j} \right\}$ corresponde ao que chamamos de fator de penalização da probabilidade de acerto na presença de *speededness*. O parâmetro c_i corresponde a probabilidade de acerto casual, assim por se tratar de uma probabilidade temos que $0 \leq c_i \leq 1$. A variável latente η_j representa o ponto de aceleração e será chamado de ponto, ou tolerância ao, *speededness*, o mesmo é expressado como uma fração do número de itens onde o respondente j primeiro experimenta o efeito de *speededness*, por isso $(\eta_j \in [0, 1])$. A variável latente λ_j por sua vez corresponde à taxa de aceleração do examinando j , ou seja, esse parâmetro controla a forma com a qual o indivíduo se comporta na presença do *speededness*, assumimos que $\lambda_j \geq 0$. No decorrer do texto chamaremos λ_j de taxa ou propensão ao *speededness*. Enquanto a_i e b_i representam, respectivamente, os parâmetros de discriminação e dificuldade do item i . A variável latente θ_j nesse modelo corresponde à habilidade de um indivíduo que possui uma tolerância η_j ao *speededness* e que na presença de *speededness* acelera numa taxa λ_j .

Sobre a variável latente η_j notemos que para os itens com $i \leq \eta_j I$ não há efeito da limitação de tempo. Uma vez que o examinando passa seu ponto de aceleração então $i/I - \eta_j > 0$, o que resulta numa diminuição (penalização) na probabilidade de resposta correta ocasionada pelo fator de penalização dado por $\min \left\{ 1, \left[1 - \left(\frac{i}{I} - \eta_j \right) \right]^{\lambda_j} \right\}$. A taxa de diminuição do fator de penalização é controlada pelo parâmetro λ_j . Valores maiores de λ_j resultam num decréscimo mais rápido na probabilidade de acerto, ou seja, numa maior penalização. Note então que, nesse novo modelo não temos incorporado apenas um traço latente associado ao indivíduo e sim três: η_j , θ_j e λ_j . Em que, η_j é o ponto de aceleração, θ_j é a habilidade e λ_j é a taxa de aceleração do j -ésimo indivíduo, ou seja, esse modelo corresponde a um modelo tridimensional da TRI.

Esse modelo leva em conta a restrição do tempo e, de acordo com o mesmo, as respostas para os itens no início do teste são regidas por um modelo 3PL. Porém, ocorre que a probabilidade de sucesso diminui gradualmente, podendo eventualmente reduzir-se apenas à probabilidade de acerto casual (o chute) sendo, portanto, uma clara extensão do modelo híbrido de [Yamamoto e Everson \(1997\)](#). Podemos obter também o que chamamos de modelo 2PLMG considerando que não há possibilidade de acerto casual e 1PLMG considerando que além de não haver a possibilidade de acerto casual todos os itens possuem o mesmo poder de discriminação. Em analogia aos modelos 1PL, 2PL e 3PL mencionados anteriormente.

Para uma interpretação dos parâmetros adicionais oriundos do fator de penalização incluído ao

modelo 3PL, suponha que tenhamos um teste com 80 itens o qual será respondido por um conjunto de indivíduos. Assim, de acordo com a Figura 2.1 considerando $\eta = 0,5$ podemos ver que quanto maior o valor de λ mais rapidamente o fator de penalização decresce impondo portanto uma maior penalização na probabilidade de acerto.

Figura 2.1: Gráfico do fator de penalização para $\eta = 0,5$ e diferentes valores de λ considerando um teste com 80 itens.

Para uma melhor visualização consideremos então a Figura 2.2 que apresenta uma visão bidimensional do fator de penalização para diferentes valores de η e λ .

Figura 2.2: Gráfico do fator de penalização para diferentes valores de η e λ considerando um teste com 80 itens.

De acordo com a mesma podemos observar que:

- Se um indivíduo possui $\eta = 0.5$ e $\lambda = 15$ (curva tipo linha) então a partir de 50% do teste o mesmo começa a sofrer o efeito da restrição do tempo. Notamos assim que, até a metade da prova a probabilidade de resposta correta desse indivíduo segue o modelo 3PL e, posteriormente, o mesmo começa a sofrer efeito da limitação do tempo sendo que, por volta do item 60 (como o valor do mínimo é praticamente nulo), a sua probabilidade de acerto passa a ser modelada via um modelo de acerto casual. Notamos então que o modelo híbrido é de certa forma um caso particular do modelo 3PLMG.

- Se outro indivíduo que também possui $\eta = 0.5$ possui $\lambda = 5$, (curva tracejada) notamos que o decaimento da probabilidade de acerto é menor em relação ao decaimento da mesma probabilidade relativa ao indivíduo anterior. Isso deve-se ao fato de λ ter assumido um valor maior para o indivíduo anterior, caso o λ seja menor a probabilidade de acerto diminuirá de forma mais gradativa (lenta), o que justifica o fato de chamarmos esse parâmetro (λ) de taxa de aceleração.
- Por fim, se um indivíduo possui $\lambda = 5$ e $\eta = 0.7$ (curva pontilhada) notamos que a partir de 70% do teste a probabilidade de acerto do mesmo passe a decair, ou seja, o mesmo passa a sofrer o efeito da limitação do tempo.

Algumas considerações sob os traços latentes e o fator de penalização

- Se para todo j tal que $i/I > \eta_j$ tivermos que $[1 - (\frac{i}{I} - \eta_j)]^{\lambda_j} = 0$ (caso este que ocorre quando $\lambda_j \rightarrow +\infty$ ou quando o mesmo é suficientemente grande) então, o modelo proposto reduz-se ao modelo híbrido de Yamamoto e Everson (1997). Nota-se então que esse modelo estende o modelo híbrido no sentido de que os examinandos não necessariamente mudam imediatamente para um processo de acerto casual após a presença de *speededness* uma vez que isso ocorre gradualmente.
- Notemos também que o modelo 2.6 também generaliza o modelo mistura de Rasch uma vez que permite examinandos distintos sentirem o efeito da limitação do tempo em momentos distintos. Além disso, no modelo 2.6 não é necessário a especificação, inicialmente, de classes de *speededness*, sendo que essas surgem naturalmente devido a presença do traço latente η_j .
- Se para algum indivíduo tivermos $\eta_j = 0$ e $\lambda_j > 0$, segue que as respostas do mesmo são modeladas sob efeito *speededness* do início ao fim do teste.
- Se $\eta_j = 1$ a probabilidade de resposta correta é modelada via um modelo 3PL, independentemente do valor de λ_j . Mas se $\lambda_j = 0$ também temos que a probabilidade de resposta correta é modelada via um modelo logístico de três parâmetros independentemente de η_j . Notamos, nessas situações, o que chamamos de falta de identificabilidade do modelo uma vez que valores distintos dos parâmetros levam à mesma função de verossimilhança. Para resolver esses problemas de falta de identificabilidade ocasionado pelos parâmetros η_j e λ_j propomos considerar as seguintes restrições $0 \leq \eta_j < 1$ e $\lambda_j > 0$. Além disso, também estabelecemos uma métrica para θ_j impondo que $\theta_j \sim N(0, 1)$ conforme feito para o modelo 3PL no Capítulo 1.

2.4 Estimação Clássica no Modelo 3PLMG

Para obtenção das estimativas dos parâmetros do modelo 3PLMG, Goegebeur *et al.* (2008) utilizaram o método da máxima verossimilhança marginal. Para isso a verossimilhança conjunta é integrada sobre os traços latentes θ_j , η_j e λ_j e, posteriormente maximizada nos parâmetros remanescentes. Para evitar problemas de identificabilidade, assim como no modelos de 3P usuais, é necessário especificar uma métrica para a habilidade, isso foi feito considerando uma distribuição normal padrão para as habilidades, ou seja, foi assumido que os θ_j são oriundos uma variável θ em que $\theta \sim N(0, 1)$. O parâmetro η_j por representar uma proporção (percentual do teste em que o indivíduo não sofreu efeito de *speededness*) foi modelado por uma variável η a qual segue uma distribuição Beta com parâmetros α e β , denotada por $\eta \sim \text{Beta}(\alpha, \beta)$. Os λ_j , por sua vez, por corresponderem a uma taxa de aceleração foram modelados por uma variável λ a qual segue uma distribuição log-normal com parâmetros μ_λ e σ_λ , denotada por $\lambda \sim \text{LN}(\mu_\lambda, \sigma_\lambda)$.

Denotando por G a distribuição conjunta θ , η e λ e por $\rho_{x,y}$ a correlação entre as variáveis x e y , segue que a função de verossimilhança marginal do vetor de parâmetros desconhecidos $\xi = (\mathbf{a}, \mathbf{b}, \mathbf{c}, \mu_\lambda, \sigma_\lambda, \alpha, \beta, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\lambda\eta})$ é dada por

$$V(\boldsymbol{\xi}; \mathbf{y}) = \prod_{j=1}^J \int_{\mathbb{R}} \int_0^1 \int_0^\infty \prod_{i=1}^I \mathbb{P}(Y_{ij} = y_{ij} | \theta_j, \eta_j, \lambda_j) dG(\theta_j, \eta_j, \lambda_j). \quad (2.8)$$

Como essas integrais não podem ser resolvidas analiticamente torna-se necessário a utilização de algum método numérico. Uma solução aproximada da integral tripla pode ser obtida via método da quadratura e a maximização pode ser feita usando o algoritmo de Newton Raphson. A *procedure* NL MIXED do SAS ajusta modelos mistos não lineares com distribuições normal multivariada para os efeitos aleatórios. Portanto, desde que em (2.8) G seja caracterizada por uma estrutura de dependência normal (cópula) a *procedure* NL MIXED poderá ser usada. Isso é válido para qualquer que seja a forma funcional (contínua) da função de distribuição marginal de $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ e $\boldsymbol{\lambda}$. Uma crítica ao uso dessa *procedure* é que não se sabe a princípio quantos pontos de quadratura considerar. Por exemplo, para modelos da TRI Sheu *et al.* (2005) recomendam o uso de 20 a 30 pontos de quadratura. Uma vez que o tempo de estimação durante o uso dessa *procedure* depende da quantidade de itens, da quantidade de respondentes e da complexidade do modelo, a utilização da mesma requer uma enorme exigência computacional tornando inviável o seu uso. Além disso, mesmo considerando de 20 a 30 pontos de quadratura os autores também recomendam um estudo de simulação com o aumento da quantidade de pontos de quadratura até que as estimativas dos parâmetros estejam estabilizadas tornando-se inviável qualquer análise mais detalhada. Também cabe ressaltar que para o conjunto de dados utilizado na aplicação no Capítulo (5) o uso da abordagem clássica não é aconselhável, uma vez que temos um grande conjunto de indivíduos que responderam todos os itens corretamente ou incorretamente.

Capítulo 3

Estimação Bayesiana do Modelo 3PLMG

Uma alternativa à metodologia de estimação clássica é usar a metodologia Bayesiana. Nessa abordagem são definidas distribuições *a priori* para todos os parâmetros. Bazan *et al.* (2010) realiza a estimação dos parâmetros do modelo 3PLMG considerando distribuições *a priori* que são independentes, para o conjunto de parâmetros. Esse procedimento de supor independência *a priori* entre as variáveis latentes visa facilitar o procedimento uma vez que pode ser difícil estabelecer a estrutura de correlação *a priori* entre os mesmos. Além disso, os autores acreditam que a dependência pode ser obtida *a posteriori*.

Neste trabalho, propomos a estimação Bayesiana considerando que os parâmetros dos indivíduos possuem uma dependência *a priori*. Essa dependência entre as distribuições *a priori* será modelada através do uso de uma cópula a qual induzirá o surgimento das covariâncias (correlações) *a priori*. A suposição de tal dependência torna-se natural uma vez que é de se esperar que as variáveis latentes possuam algum tipo de associação.

No que segue, para estimação dos parâmetros do modelo 3PLMG consideramos em (2.6) a reparametrização $a_i\theta_j - b_i^*$ ao invés de $a_i(\theta_j - b_i)$, ou seja, $b_i^* = a_ib_i$. O uso dessa reparametrização é usual na TRI sob a abordagem Bayesiana pois reduz bastante o tempo computacional, ver por exemplo Patz e Junker (1999b), Bazan *et al.* (2010) e Sahu (2002). De agora em diante como usaremos apenas a abordagem Bayesiana, quando nos referirmos ao modelo 3PLMG ou 3PL estaremos sempre usando $a_i\theta_j - b_i^*$ ao invés de $a_i(\theta_j - b_i)$.

Um ponto importante no processo de estimação Bayesiano está na especificação das distribuições *a priori*. Como é usual em TRI, consideramos que $\theta_j \sim N(0, 1)$, $j = 1 \dots J$, com isso fica estabelecida uma métrica para as estimativas das habilidades. A escolha da normal padrão é bastante comum e pode ser vista em Patz e Junker (1999b), Kim *et al.* (1994) e Kim (2001). Para os parâmetros dos itens há diversas opções de especificação das distribuições *a priori*. Para o modelo 3PL Patz e Junker (1999b) propõem que $a_i \sim LN(1, 0.5)$, $b_i^* \sim N(0, 2)$ e $c_i \sim Beta(12.5, 37, 5)$; Para o modelo 2PL Albert e Ghosh (2000) consideram que $a_i \sim N(0, 1)$ e $b_i^* \sim N(0, 1)$; Fu *et al.* (2009) por sua vez considera que $a_i \sim NT(0, 1)$, $b_i^* \sim N(0, 1)$ e $c_i \sim Beta(5, 17)$, em que LN e NT correspondem, respectivamente, às distribuições log-normal e normal truncada. Para os parâmetros novos dos indivíduos consideramos que $\eta_j \sim Beta(\alpha, \beta)$, uma vez que representa uma proporção, e $\lambda_j \sim LN(\mu_\lambda, \sigma_\lambda)$, assumindo somente valores positivos por controlar uma taxa de aceleração. Cabe ressaltar que em nossa aplicação α , β , μ_λ e σ_λ serão previamente especificados onde justificaremos a escolha dos mesmos através de uma análise de sensibilidade. As escolhas dessas distribuições *a priori* para os novos parâmetros foi feita seguindo Bazan *et al.* (2010) e embasada principalmente pela escolha de Goegebeur *et al.* (2008) que primeiro realizaram a estimação no modelo em estudo. É importante notar também que existem diversos valores possíveis para atribuímos aos parâmetros das distribuições *a priori* o que pode ser feito considerando, por exemplo, informações prévias sobre o ponto de aceleração dos examinandos ou considerando que conhecemos a média e a variância das respectivas distribuições *a priori*. No que segue para os parâmetros dos itens iremos adotar as distribuições *a priori* utilizadas em Bazan *et al.* (2006) as quais apresentaram melhores resultados na análise de sensibilidade feita por eles. Para tornar as contas que faremos adiante mais gerais

consideramos então que $a_i \sim NT(\mu_a, \sigma_a^2)I(a_i > 0)$, $b_i^* \sim N(\mu_b, \sigma_b^2)$ e $c_i \sim Beta(e_1, e_2)$, em que $\mu_a, \sigma_a^2, \mu_b, \sigma_b^2, e_1$ e e_2 são supostamente conhecidos, ou seja, não são hiperparâmetros do modelo em questão.

Conforme mencionado anteriormente, iremos induzir a estrutura de dependência entre os traços latentes através do uso de cópulas. Sendo assim, na Seção 3.1 apresentamos algumas definições e desenvolvemos alguns resultados que serão úteis para estimarmos os parâmetros do modelo 3PLMG. Para um aprofundamento maior na teoria de cópulas recomenda-se a leitura de Anjos *et al.* (2004). Porém, veremos que o uso da cópula conduzirá a uma restrição na matriz de covariâncias *a priori* dos parâmetros θ , η e λ . Assim, teremos a necessidade de estudar uma proposta de estimação MCMC da matriz de covariâncias. Na Subseção 3.2.1 apresentamos o processo MCMC para estimação da matriz de covariâncias no caso sem restrições. Por fim, na Subseção 3.2.2 apresentamos o processo de estimação para o caso com restrição.

3.1 Cópulas

Em estatística temos que em diversas situações-problema está intrínseca a dependência entre as variáveis. No entanto, por simplicidade muitas vezes é considerada a hipótese de independência. O uso da suposição de independência quando a mesma não é válida faz com que possamos perder informação sobre os parâmetros do modelo e num eventual processo de estimação pode nos conduzir a conclusões errôneas sobre o problema em questão. Sendo assim, é aconselhável buscar formas de incluir a dependência entre as variáveis em análises estatísticas. Uma alternativa é usar uma distribuição multivariada para incluir essa dependência de modo a preservar o conhecimento *a priori* das distribuições marginais, e isso é feito usando cópulas. Cópulas são funções que ligam uma função de distribuição multivariada com suas respectivas funções de distribuições marginais univariadas. A vantagem do uso de cópulas em relação à alternativa de se trabalhar com distribuições conjuntas multivariadas, está no fato de ser possível impor distribuições marginais específicas e com diferentes estruturas para uma distribuição multivariada conforme veremos no decorrer do trabalho.

A próxima proposição apresenta um resultado que corresponde a um dos resultados mais importante para realização de simulações em estatística, esse resultado pode ser encontrado em diversos livros de probabilidade dentre os quais podemos citar James (2006) e Magalhães (2006) e sua importância está no fato dele fundamentar o tradicional método de simulação de Monte Carlo.

Proposição 1. *Seja X uma variável aleatória absolutamente contínua com função de distribuição $\mathbb{F}$. Então, $Y = \mathbb{F}(X)$ terá distribuição uniforme contínua em $[0, 1]$. Vale a recíproca, sendo $U \sim Unif[0, 1]$ então $X = \mathbb{F}^{-1}(U)$ terá função de distribuição $\mathbb{F}$.*

A importância desse resultado é que o mesmo permite gerar valores da variável aleatória X a partir da geração de valores de variável aleatória U , em que $U \sim Unif(0, 1)$. Para isso, basta gerarmos um valor u de U para obtermos o valor $x = \mathbb{F}^{-1}(u)$ da variável aleatória X .

Note que, quando temos um vetor de n variáveis aleatórias $\mathbf{X} = (X_1, \dots, X_n)$ a dependência entre as variáveis fica completamente especificada pela função de distribuição conjunta, ou seja, por

$$\mathbb{F}_{\mathbf{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Sendo $X_1, \dots, X_n$ variáveis aleatórias absolutamente contínuas vamos definir $C(\cdot)$ como sendo a função de distribuição do vetor aleatório $(\mathbb{F}_{X_1}(X_1), \dots, \mathbb{F}_{X_n}(X_n))$. De acordo com a Proposição 1, temos que $U_i = \mathbb{F}_{X_i}(X_i) \sim U(0, 1) \forall i = 1, \dots, n$. Segue então que,

$$\begin{aligned} C(u_1, \dots, u_n) &= \mathbb{P}(\mathbb{F}_{X_1}(X_1) \leq u_1, \dots, \mathbb{F}_{X_n}(X_n) \leq u_n) \\ &= \mathbb{P}(X_1 \leq \mathbb{F}_{X_1}^{-1}(u_1), \dots, X_n \leq \mathbb{F}_{X_n}^{-1}(u_n)) \\ &= \mathbb{F}_{\mathbf{X}}(\mathbb{F}_{X_1}^{-1}(u_1), \dots, \mathbb{F}_{X_n}^{-1}(u_n)) \end{aligned}$$

como $u_i = \mathbb{F}_{X_i}(x_i)$, temos

$$\mathbb{F}_{\mathbf{X}}(x_1, \dots, x_n) = C(\mathbb{F}_{X_1}(x_1), \dots, \mathbb{F}_{X_n}(x_n)). \quad (3.1)$$

A função $C(\cdot)$ que satisfaz a Equação (3.1) é o que chamamos de cópula do vetor aleatório $(X_1, \dots, X_n)$ ou cópula associada a $\mathbb{F}_{\mathbf{X}}$. No que segue faremos uma modelagem multivariada na qual a estrutura de dependência entre as variáveis é modelada por uma cópula. A seguir apresentamos duas definições formais de cópula.

Definição 3.1. *Uma cópula é definida como a função de distribuição conjunta*

$$C(u_1, \dots, u_n) = \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n), \quad 0 \leq u_i \leq 1$$

em que $U_i \sim \text{Unif}(0, 1)$, $i = 1, \dots, n$.

Definição 3.2. *Uma cópula é qualquer função $C : [0, 1]^n \rightarrow [0, 1]$ que satisfaz as seguintes propriedades:*

- (i) $C(x_1, \dots, x_n)$ é crescente em cada componente x_i ;
- (ii) $C(1, \dots, 1, x_i, 1, \dots, 1) = x_i \in [0, 1]$ para todo $i = 1, \dots, n$;
- (iii) Para todos $(a_1, \dots, a_n), (b_1, \dots, b_n) \in [0, 1]^n$ com $a_i \leq b_i$ tem-se

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1+\dots+i_n} C(x_{1i_1}, \dots, x_{ni_n}) \geq 0$$

com $x_{j1} = a_j$ e $x_{j2} = b_j$, $j = 1, \dots, n$.

[Embrechts et al. \(2002\)](#) mostraram que a Definição (3.1) e a Definição (3.2) são equivalentes.

A seguir temos o resultado mais importante da teoria de cópulas que corresponde ao teorema de Sklar ([Sklar, 1959](#)), o mesmo assegura a existência e a unicidade da cópula $C(\cdot)$. Como trabalharemos apenas com distribuições contínuas enunciaremos uma versão do Teorema suprimindo a parte discreta. No que segue apenas enunciaremos o teorema de Sklar e apresentaremos alguns resultados sobre cópulas omitindo suas demonstrações por não se tratarem do objetivo desse trabalho. Aos interessados, uma demonstração da versão bidimensional do teorema de Sklar pode ser encontrada em [Nelsen \(2013\)](#), enquanto as demonstrações do Corolário 1 e das Proposições 2 e 3 podem ser encontradas em [Anjos et al. \(2004\)](#). Uma demonstração formal, usando uma abordagem topológica, para o caso n -dimensional é apresentada em [Durante et al. \(2013\)](#).

Teorema 1. *(Teorema de Sklar) Seja $\mathbb{F}(\cdot)$ uma função de distribuição conjunta absolutamente contínua com distribuições marginais $F_{X_1}, \dots, F_{X_n}$. Então, existe uma cópula n -dimensional $C(\cdot)$ tal que*

$$\mathbb{F}(x_1, \dots, x_n) = C(\mathbb{F}_{X_1}(x_1), \dots, \mathbb{F}_{X_n}(x_n)).$$

Além disso, se $\mathbb{F}_{X_1}, \dots, \mathbb{F}_{X_n}$ são todas contínuas, então $C(\cdot)$ é única. Reciprocamente, se $C(\cdot)$ é uma cópula n -dimensional e $\mathbb{F}_{X_1}, \dots, \mathbb{F}_{X_n}$ são funções de distribuição, então a função $\mathbb{F}(\cdot)$ é a função de distribuição conjunta das variáveis $X_1, \dots, X_n$ e suas distribuições marginais são $\mathbb{F}_{X_1}(x_1), \dots, \mathbb{F}_{X_n}(x_n)$.

A importância do Teorema de Sklar na modelagem estatística está em sua recíproca, pois de acordo com a mesma podemos acoplar distribuições univariadas de qualquer tipo, com qualquer cópula, e teremos definido uma distribuição conjunta para o vetor aleatório $(X_1, \dots, X_n)$.

O corolário a seguir mostra que podemos extrair uma cópula de qualquer distribuição multivariada e usá-la independentemente das distribuições marginais originais.

Corolário 1. *Seja $\mathbb{F}(\cdot)$ uma função de distribuição conjunta com distribuições marginais $\mathbb{F}_{X_1}, \dots, \mathbb{F}_{X_n}$ e $\mathbb{F}_{X_1}^{-1}, \dots, \mathbb{F}_{X_n}^{-1}$ suas respectivas funções inversas. Então, para qualquer $u_i \in [0, 1]$ existe uma cópula n -dimensional $C(\cdot)$ tal que*

$$C(u_1, \dots, u_n) = \mathbb{F}(\mathbb{F}_{X_1}^{-1}(u_1), \dots, \mathbb{F}_{X_n}^{-1}(u_n)).$$

A interpretação desse resultado é que podemos estudar o fenômeno de dependência entre as variáveis sem fixar as distribuições.

Exemplo 1. *Se a distribuição conjunta de (X, Y) é dada por*

$$\mathbb{F}(x, y) = [1 + \exp(-x) + \exp(-y)]^{-1}, \quad x, y \in R$$

temos que

$$\mathbb{F}_X(x) = \lim_{y \rightarrow \infty} \mathbb{F}(x, y) = \frac{1}{1 + \exp(-x)}$$

e

$$\mathbb{F}_Y(y) = \lim_{x \rightarrow \infty} \mathbb{F}(x, y) = \frac{1}{1 + \exp(-y)}.$$

As funções inversas são dadas por $\mathbb{F}_X^{-1}(u) = -\ln(u^{-1} - 1)$ para $u \in (0, 1)$ e $\mathbb{F}_Y^{-1}(v) = -\ln(v^{-1} - 1)$ para $v \in (0, 1)$. De acordo com o corolário, a cópula $C(u, v)$ associada com $\mathbb{F}(x, y)$ é

$$C(u, v) = \mathbb{F}(\mathbb{F}_X^{-1}(u), \mathbb{F}_Y^{-1}(v)) = (u^{-1} + v^{-1} - 1)^{-1}.$$

A seguir veremos um importante resultado do estudo de cópulas cuja demonstração pode ser vista em [Anjos et al. \(2004\)](#). Tal resultado garante que se $X_1, \dots, X_n$ possuem conjuntamente uma cópula $C(\cdot)$, então ao aplicarmos transformações estritamente crescentes nas variáveis aleatórias em questão a cópula associada às mesmas não é alterada.

Proposição 2. *Seja $(X_1, \dots, X_n)$ um vetor de variáveis aleatórias contínuas com cópula $C(\cdot)$. Se $\alpha_1(X_1), \dots, \alpha_n(X_n)$ são funções estritamente crescentes nas imagens das variáveis aleatórias $X_1, \dots, X_n$, ou seja, em $Im(X_1), \dots, Im(X_n)$, então o vetor $(\alpha_1(X_1), \dots, \alpha_n(X_n))$ também tem cópula $C(\cdot)$*

Uma consequência imediata dessa proposição é que se a localização e a escala das distribuições forem alteradas, a cópula associada às mesmas não é alterada. Diante disso, por convenção, definimos a cópula normal considerando $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma}$ como sendo uma matriz de correlações, isto é, os elementos da diagonal principal são iguais a um e os demais elementos da matriz pertencem ao intervalo $[-1, 1]$.

Definição 3.3. *Seja $\Phi(\cdot)$ a função distribuição normal padrão e seja $\Phi_{\boldsymbol{\Sigma}}^n(\cdot)$ a função distribuição normal multivariada com vetor de médias nulo e matriz de correlações $\boldsymbol{\Sigma}$, então*

$$C(u_1, \dots, u_n) = \Phi_{\boldsymbol{\Sigma}}^n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

é denominada cópula normal n -variada.

Proposição 3. *Considere um vetor aleatório n -dimensional $\mathbf{X} = (X_1, \dots, X_n)$ com função de distribuição conjunta $\mathbb{F}(\cdot)$ e funções de distribuições marginais contínuas $\mathbb{F}_1, \dots, \mathbb{F}_n$. Assuma que $\mathbb{F}(\cdot)$ é caracterizada por uma função de dependência normal C (cópula), isto é,*

$$\mathbb{F}_{\mathbf{X}}(x_1, \dots, x_n) = C(\mathbb{F}_1(x_1), \dots, \mathbb{F}_n(x_n))$$

com

$$C(u_1, \dots, u_n) = \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_n)} \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(1/2) \mathbf{z}' \boldsymbol{\Sigma}^{-1} \mathbf{z}} d\mathbf{z}$$

em que Σ denota uma matriz (positiva definida) de correlações e Φ^{-1} é a inversa da função de distribuição normal padrão. Então, as variáveis aleatórias

$$Y_i = \Phi^{-1}(\mathbb{F}_i(X_i)), \quad i = 1, \dots, n$$

são tais que $(Y_1, \dots, Y_n) \sim N(\mathbf{0}, \Sigma)$.

Esse resultado é interessante pois, dado uma sequência de variáveis aleatórias $X_1, \dots, X_n$, o mesmo permite-nos gerar valores da distribuição conjunta dessa sequência de variáveis aleatórias, se especificarmos uma cópula normal multivariada como na Definição (3.3).

Vejam agora uma proposição que será de extrema importância no processo de estimação do modelo 3PLMG. Através dessa proposição uma vez especificada a cópula como sendo normal multivariada e as distribuições marginais *a priori* de θ , η e λ , obtemos a densidade conjunta de θ , η e λ . Além disso, também vemos nessa proposição como gerar da distribuição conjunta de θ , η e λ .

Proposição 4. *Seja $\mathbb{F}(\theta, \eta, \lambda) = C(\mathbb{F}_1(\theta), \mathbb{F}_2(\eta), \mathbb{F}_3(\lambda))$, em que $\mathbb{F}_1$, $\mathbb{F}_2$ e $\mathbb{F}_3$ correspondem, respectivamente, as distribuições acumuladas das seguintes distribuições: $N(0, 1)$, $Beta(\alpha, \beta)$ e $LN(\mu_\lambda, \sigma_\lambda^2)$. Se C corresponde a uma cópula normal multivariada com matriz de correlação Σ então,*

- (i) *Para gerarmos valores da distribuição conjunta de θ , η e λ basta gerarmos y_1 , y_2 e y_3 de uma cópula normal multivariada e fazermos $\theta = y_1$, $\eta = \mathbb{F}_2^{-1}(\Phi(y_2))$ (como essa expressão não possui uma forma fechada mantivemos esse formato, observando que a mesma corresponde a um quantil da distribuição Beta podemos obter seu valor computacionalmente) e $\lambda = \exp(\sigma_\lambda y_3 + \mu_\lambda)$.*

- (ii) *A densidade conjunta de θ , η e λ é dada por*

$$f(\theta, \eta, \lambda) = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2} \sigma_\lambda \lambda} \frac{f_2(\eta)}{\phi(\Phi^{-1}(\mathbb{F}_2(\eta)))} \exp(-\gamma' \Sigma^{-1} \gamma / 2) \quad (3.2)$$

$$\text{em que } \gamma' = \left(\theta, \Phi^{-1}(\mathbb{F}_2(\eta)), \frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda} \right).$$

Demonstração.

- (i) Segue da Proposição 3 que

$$Y_1 = \Phi^{-1}(\mathbb{F}_1(\theta)) = \Phi^{-1}(\Phi(\theta)) = \theta,$$

$$Y_2 = \Phi^{-1}(\mathbb{F}_2(\eta)) \iff \eta = \mathbb{F}_2^{-1}(\Phi(y_2))$$

e $Y_3 = \Phi^{-1}(\mathbb{F}_3(\lambda))$, são tais que $\mathbf{Y} = (Y_1, Y_2, Y_3)$ possui uma distribuição normal multivariada com matriz de correlação Σ .

Para vermos que $\lambda = \exp(\sigma_\lambda y_3 + \mu_\lambda)$ lembremos inicialmente que $\lambda \sim LN(\mu_\lambda, \sigma_\lambda^2)$ se, e somente se, $\ln(\lambda) \sim N(\mu_\lambda, \sigma_\lambda^2)$ ou, equivalentemente, $\frac{\ln(\lambda) - \mu_\lambda}{\sigma_\lambda} \equiv Z \sim N(0, 1)$. Assim, como por hipótese $Y_3 = \Phi^{-1}(\mathbb{F}_3(\lambda))$ e

$$\begin{aligned} \mathbb{F}_3(\lambda) &= \mathbb{P}(\lambda \leq y) \\ &= \mathbb{P}\left(\frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda} \leq \frac{\ln y - \mu_\lambda}{\sigma_\lambda}\right) \\ &= \mathbb{P}\left(Z \leq \frac{\ln y - \mu_\lambda}{\sigma_\lambda}\right) \\ &= \Phi\left(\frac{\ln y - \mu_\lambda}{\sigma_\lambda}\right), \end{aligned} \quad (3.3)$$

obtemos que,

$$Y_3 = \Phi^{-1}(\mathbb{F}_3(\lambda)) = \Phi^{-1}\left(\Phi\left(\frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda}\right)\right) = \frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda}, \quad (3.4)$$

e, conseqüentemente, concluímos que, $\lambda = \exp(\sigma_\lambda y_3 + \mu_\lambda)$.

(ii) Como foi visto em (i) temos que

$$Y_1 = \theta, \quad (3.5)$$

$$Y_2 = \Phi^{-1}(\mathbb{F}_2(\eta)), \quad (3.6)$$

$$Y_3 = \frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda}. \quad (3.7)$$

Conseqüentemente,

$$\frac{\partial Y_1}{\partial \theta} = 1, \quad \frac{\partial Y_2}{\partial \eta} = \frac{f_2(\eta)}{\phi(\Phi^{-1}(\mathbb{F}_2(\eta)))} \quad \text{e} \quad \frac{\partial Y_3}{\partial \lambda} = \frac{1}{\lambda \sigma_\lambda},$$

em que $f_2(\eta) = \frac{d\mathbb{F}_2(\eta)}{d\eta}$, ou seja, a função densidade de probabilidade de η .

Assim,

$$\begin{aligned} |J| &= \begin{vmatrix} 1 & 0 & 0 \\ 0 & \frac{f_2(\eta)}{\phi(\Phi^{-1}(\mathbb{F}_2(\eta)))} & 0 \\ 0 & 0 & \frac{1}{\lambda \sigma_\lambda} \end{vmatrix} \\ &= \frac{f_2(\eta)}{\lambda \sigma_\lambda \phi(\Phi^{-1}(\mathbb{F}_2(\eta)))}. \end{aligned} \quad (3.8)$$

Logo, utilizando (3.8) e o método do Jacobiano concluímos que

$$\begin{aligned} f(\theta, \eta, \lambda) &= f_{(Y_1, Y_2, Y_3)}((y_1, y_2, y_3)) |J| \\ &= \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2} \sigma_\lambda \lambda} \frac{f_2(\eta)}{\phi(\Phi^{-1}(\mathbb{F}_2(\eta)))} \exp(-\gamma' \Sigma^{-1} \gamma / 2) \end{aligned}$$

em que, $\gamma' = \left(\theta, \Phi^{-1}(\mathbb{F}_2(\eta)), \frac{\ln \lambda - \mu_\lambda}{\sigma_\lambda} \right)$. ■

De acordo com a Proposição 3 nota-se que uma vez imposta uma cópula normal multivariada, independentemente das distribuições marginais estabelecidas *a priori*, ficam completamente especificados o vetor de médias, $\boldsymbol{\mu} = \mathbf{0}$, e a diagonal da matriz de covariâncias, $\boldsymbol{\Sigma}$, uma vez que a mesma corresponde a uma matriz de correlações. Assim, sendo a diagonal principal de $\boldsymbol{\Sigma}$ composta por elementos conhecidos (fixos e iguais a 1) devemos estimar os demais elementos da matriz de correlações da cópula sob essa restrição. Chan e Jeliazkov (2009) propõem uma maneira interessante para a estimação da matriz de covariâncias na presença de restrições. Nas próximas duas subseções faremos um breve resumo desta alternativa.

3.2 Estimação MCMC da Matriz de Covariâncias

É comum em modelos lineares termos uma variável resposta normal multivariada com matriz de covariância desconhecida e, conseqüentemente, termos o interesse em estimar a mesma. Usualmente, quando temos todos os valores da matriz de covariâncias desconhecidos, uma candidata a distribuição *a priori* para esta matriz é a Wishart Inversa ³ (Kadane, 2011). No entanto, Chan e Jeliazkov (2009) citam que podem surgir restrições nos elementos da matriz de covariância, como os elementos da diagonal da matriz de covariâncias serem 1, em estudos envolvendo cópulas normais. Problemas semelhantes, com restrições nos elementos da matriz de covariância, ocorrem em modelos com truncamento incidentais ou indicadores de tratamento endógenos. Nesses casos não é possível atribuir diretamente uma distribuição *a priori* Wishart Inversa para a matriz de covariâncias sendo, portanto, necessário usar alguma metodologia alternativa. Para desenvolver essa metodologia os autores consideram inicialmente a inexistência de restrições na matriz de covariâncias e desenvolvem o algoritmo de *Gibbs* para esse caso, posteriormente eles generalizam para o caso restrito. Na próxima subseção apresentamos o desenvolvimento do *Gibbs Sampling* proposto em Chan e Jeliazkov (2009) para o caso irrestrito.

3.2.1 Caso da Matriz Irrestrita

Suponha que temos N observações independentes e identicamente distribuídas de uma distribuição normal p -variada, $N_p(\mathbf{0}, \mathbf{\Sigma})$, e desejamos estimar $\mathbf{\Sigma}$. Notamos primeiramente que $\mathbf{\Sigma}$ é positiva definida assim, existem únicas matrizes $\mathbf{L}$ e $\mathbf{D}$ tais que $\mathbf{L}\mathbf{\Sigma}\mathbf{L}' = \mathbf{D}$ ou, equivalentemente, $\mathbf{\Sigma}^{-1} = \mathbf{L}'\mathbf{D}^{-1}\mathbf{L}$ em que $\mathbf{L}$ é uma matriz triangular inferior com uns na diagonal e $\mathbf{D}$ é uma matriz diagonal com elementos positivos (Golub e Van Loan, 2012). Assim,

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_p \end{pmatrix} \text{ e } \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & l_{p3} & \dots & 1 \end{pmatrix}.$$

Considere $\mathbf{d} = (d_1, \dots, d_p)'$, $\mathbf{l}_k = (l_{k1}, l_{k2}, \dots, l_{k(k-1)})'$ e $k = 2, \dots, p$, $\mathbf{l} = (\mathbf{l}'_2, \dots, \mathbf{l}'_p)'$. Como $\mathbf{\Sigma}^{-1} = \mathbf{L}'\mathbf{D}^{-1}\mathbf{L}$ e a densidade de $\mathbf{\Sigma}$, $f(\mathbf{\Sigma})$, por ser expressa como $f(\mathbf{\Sigma}) = f(\mathbf{l}, \mathbf{d})$ é evidente que o conhecimento de $\mathbf{L}$ e $\mathbf{D}$, nos proporciona o conhecimento de $\mathbf{\Sigma}$. Por outro lado, $f(\mathbf{l}, \mathbf{d}) = f(\mathbf{d})f(\mathbf{l}|\mathbf{d})$, ou seja, $\mathbf{\Sigma}$ é totalmente especificada em termos das densidades de $\mathbf{d}$ e de $\mathbf{l}|\mathbf{d}$.

Conforme já mencionamos uma candidata natural a ser uma distribuição *a priori* para $\mathbf{\Sigma}^{-1}$ é a distribuição Wishart ⁴. À seguir apresentamos um resultado que consiste de uma relação de equivalência entre a distribuição Wishart e as distribuições Gama Inversa e Normal. De acordo com o mesmo $\mathbf{\Sigma}^{-1} \sim W(\nu, \mathbf{R})$ se, e somente se,

$$d_k \sim IG(\nu_{k0}/2, \delta_{k0}/2), \quad k = 1, \dots, p, \quad (3.9)$$

e

$$\mathbf{l}|\mathbf{d} \sim N(\mathbf{l}_0, \mathbf{L}_0) \quad (3.10)$$

onde os elementos de $\mathbf{l}_0$ não são necessariamente nulos, e os de $\mathbf{L}_0$ podem variar livremente e possivelmente depender de $\mathbf{d}$. Formalmente, temos:

Teorema 2. *As distribuições a priori em (3.9) e (3.10) induzem uma distribuição Wishart para a matriz $\mathbf{\Sigma}^{-1} \equiv \mathbf{L}'\mathbf{D}^{-1}\mathbf{L}$ de modo que $\mathbf{\Sigma}^{-1} \sim W(\nu, \mathbf{R})$, em que $\mathbf{R}$ é uma matriz simétrica $p \times p$ positiva definida. Além disso, a recíproca também é verdadeira, se $\mathbf{\Sigma}^{-1} \sim W(\nu, \mathbf{R})$, então as distribuições*

³ver Apêndice A.3

⁴Ver Apêndice A.3

a priori induzidas em d_k , $k = 1, \dots, p$ e $\mathbf{l}|\mathbf{d}$, $1 \leq j < k \leq p$ são as apresentadas em (3.9) e (3.10), respectivamente.

Demonstração. Apêndice C. ■

Uma observação importante é que, em princípio, pode-se definir distribuições *a priori* não conjugadas para $\mathbf{d}$ e $\mathbf{l}$ de modo a obter-se distribuições ainda mais flexíveis para a matriz de covariâncias. Como distribuições *a priori* possíveis para $\mathbf{d}$ podemos citar as distribuições gama, lognormal ou log-t (as duas últimas permitem também a correlação entre os elementos de $\mathbf{d}$) enquanto $\mathbf{l}$ pode ser modelado mais flexivelmente por uma mistura na escala de normais, que incluem as distribuições t-student e logística, entre outras. No que segue, vamos trabalhar com distribuições *a priori* conjugadas. O resultado enunciado a seguir apresenta as condicionais completas necessárias para a construção do algoritmo de *Gibbs* através do qual podemos estimar os elementos da matriz Σ no caso sem restrição.

Proposição 5. *Sob as condições do Teorema 2, se $\mathbf{u}_1, \dots, \mathbf{u}_N$ compõem uma amostra aleatória da variável aleatória $\mathbf{u} \sim N_p(\mathbf{0}, \Sigma)$ e $\Sigma^{-1} \sim W(\nu, \mathbf{R})$ então,*

(i)

$$d_k|\mathbf{u}, \mathbf{l} \stackrel{ind}{\sim} IG\left(\frac{\nu_{k0} + N}{2}, \frac{\delta_{k0} + s_k}{2}\right),$$

em que s_k é o elemento (k, k) de $\sum_{i=1}^N \boldsymbol{\omega}_i \boldsymbol{\omega}_i'$.

(ii)

$$\mathbf{l}_k|\mathbf{u}, d_k \stackrel{ind}{\sim} N_k(\hat{\mathbf{l}}_k, \hat{\mathbf{L}}_k), \quad k = 2, \dots, p,$$

em que, $\mathbf{L}_{k0}$ pode, possivelmente, depender de d_k e

$$\hat{\mathbf{L}}_k = (\mathbf{L}_{k0}^{-1} + d_k^{-1} \mathbf{X}_k' \mathbf{X}_k)^{-1}, \quad \hat{\mathbf{l}}_k = \hat{\mathbf{L}}_k (\mathbf{L}_{k0}^{-1} \mathbf{l}_{k0} + d_k^{-1} \mathbf{X}_k' \mathbf{z}_k)$$

$$\mathbf{z}_k = (u_{1k}, \dots, u_{Nk})' \quad e \quad \mathbf{X}_k = [\mathbf{z}_1 : \dots : \mathbf{z}_{k-1}].$$

Demonstração. Inicialmente vamos obter a função de verossimilhança, a qual denotaremos por $V(\mathbf{u}; \Sigma)$. Definindo $\boldsymbol{\omega}_i \equiv \mathbf{L}\mathbf{u}_i$ e $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_N)'$ e notando que

$$|\Sigma^{-1}| = |\mathbf{L}'||\mathbf{D}^{-1}||\mathbf{L}| = |\mathbf{D}^{-1}| = \prod_{k=1}^p d_k^{-1},$$

a verossimilhança pode ser escrita como

$$\begin{aligned} V(\mathbf{u}; \Sigma) &\propto |\Sigma|^{-N/2} \exp \left\{ \sum_{i=1}^N -\frac{1}{2} \mathbf{u}'_i \Sigma^{-1} \mathbf{u}_i \right\} \\ &= \left(\prod_{k=1}^p d_k^{-N/2} \right) \exp \left\{ -\frac{1}{2} \sum_{i=1}^N \boldsymbol{\omega}'_i \mathbf{D}^{-1} \boldsymbol{\omega}_i \right\} \\ &= \left(\prod_{k=1}^p d_k^{-N/2} \right) \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{D}^{-1} \sum_{i=1}^N \boldsymbol{\omega}_i \boldsymbol{\omega}'_i) \right\} \\ &= \prod_{k=1}^p d_k^{-N/2} \exp \left\{ -\frac{s_k}{2d_k} \right\}, \end{aligned} \tag{3.11}$$

onde s_k é o elemento (k, k) de $\sum_{i=1}^N \boldsymbol{\omega}_i \boldsymbol{\omega}_i'$. Uma vez que temos a verossimilhança podemos demonstrar (i) e (ii).

(i) Considerando então que os d_k possuem distribuições *a priori* gama inversa como especificadas em (3.9), para $k = 1, \dots, p$, e usando (3.11) segue que

$$\begin{aligned}\pi(d_k|\mathbf{u}, \mathbf{l}) &\propto d_k^{-N/2} \exp\left(\frac{-s_k}{2d_k}\right) d_k^{-\frac{\nu_{k0}}{2}-1} \\ &= d_k^{-\frac{N+\nu_{k0}}{2}-1} \exp\left(\frac{-s_k}{2d_k}\right).\end{aligned}$$

Além disso, pode-se verificar facilmente que os $d_k|\mathbf{u}, \mathbf{l}$ são independentes, basta notar que

$$\pi(\mathbf{d}|\mathbf{u}, \mathbf{l}) = \prod_{k=1}^p \pi(d_k|\mathbf{u}, \mathbf{l}).$$

Resumidamente concluímos que

$$d_k|\mathbf{u}, \mathbf{l} \stackrel{ind}{\sim} IG\left(\frac{\nu_{k0} + N}{2}, \frac{\delta_{k0} + s_k}{2}\right). \quad (3.12)$$

(ii) Para obtermos a distribuição de $\mathbf{l}_k|\mathbf{u}, d_k$, vamos inicialmente reescrever a verossimilhança apresentada em (3.11) de uma maneira mais conveniente.

Observe primeiramente que os elementos l_{ij} entram na verossimilhança como coeficientes nas regressões de u_{ij} nos valores negativos de $\{u_{ik}\}_{k < j}$, para $j < p$. Para ver isso note que,

$$\mathbf{L}\mathbf{u}_i = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & l_{p3} & \dots & 1 \end{pmatrix} \begin{pmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \\ \vdots \\ u_{ip} \end{pmatrix} = \begin{pmatrix} u_{i1} \\ u_{i2} + l_{21}u_{i1} \\ u_{i3} + l_{31}u_{i1} + l_{32}u_{i2} \\ \vdots \\ u_{ip} + \sum_{k=1}^p l_{pk}u_{ik} \end{pmatrix}.$$

Podemos então reescrever a verossimilhança como

$$\begin{aligned}V(\mathbf{u}; \mathbf{\Sigma}) &\propto \left(\prod_{k=1}^p d_k^{-N/2}\right) \exp\left\{-\frac{1}{2} \sum_{i=1}^N (\mathbf{L}\mathbf{u}_i)' \mathbf{D}^{-1} (\mathbf{L}\mathbf{u}_i)\right\} \\ &= \left(\prod_{k=1}^p d_k^{-N/2}\right) \exp\left\{\frac{1}{2} \sum_{i=1}^N \frac{u_{i1}^2}{d_1}\right\} \exp\left\{-\frac{1}{2} \sum_{i=1}^N \frac{(u_{i2} + l_{21}u_{i1})^2}{d_2}\right\} \dots \\ &\dots \exp\left\{-\frac{1}{2} \sum_{i=1}^N \frac{(u_{ip} + \sum_{k=1}^p l_{pk}u_{ik})^2}{d_p}\right\}.\end{aligned} \quad (3.13)$$

Por sua vez, (3.13) pode ser reescrita de uma forma mais conhecida com $\mathbf{l}$ entrando como um vetor de coeficientes de regressão. Assim,

$$V(\mathbf{u}; \mathbf{\Sigma}) \propto \left(\prod_{k=1}^p d_k^{-N/2}\right) \exp\left\{-\frac{1}{2} \sum_{i=1}^N (\mathbf{u}_i - \mathbf{U}_i \mathbf{l})' \mathbf{D}^{-1} (\mathbf{u}_i - \mathbf{U}_i \mathbf{l})\right\}, \quad (3.14)$$

em que

$$\mathbf{U}_i = - \begin{pmatrix} 0 & \dots & & & & & & 0 \\ u_{i1} & 0 & \dots & & & & & \vdots \\ 0 & u_{i1} & u_{i2} & 0 & \dots & & & 0 \\ 0 & \dots & 0 & u_{i1} & u_{i2} & u_{i3} & 0 & \dots & 0 \\ \vdots & \dots & & \ddots & & & \ddots & \dots & 0 \\ 0 & \dots & & 0 & \dots & 0 & u_{i1} & \dots & u_{ip} \end{pmatrix}.$$

Portanto, dado a distribuição *a priori* em (3.10), a distribuição condicional completa obtida é

$$\mathbf{l}|\mathbf{u}, \mathbf{d} \sim N(\hat{\mathbf{l}}, \hat{\mathbf{L}}), \quad (3.15)$$

com $\hat{\mathbf{L}} = (\mathbf{L}_0^{-1} + \sum_{i=1}^N \mathbf{U}_i \mathbf{D}^{-1} \mathbf{U}_i)^{-1}$ e $\hat{\mathbf{l}} = \hat{\mathbf{L}}(\mathbf{L}_0^{-1} \mathbf{l}_0 + \sum_{i=1}^N \mathbf{U}_i \mathbf{D}^{-1} \mathbf{u}_i)$.

Como $\mathbf{D}$ é diagonal, se acontecer de $\mathbf{L}_0$ ser diagonal ou bloco diagonal por linhas de $\mathbf{L}$, isto é,

$$\mathbf{l}_k \stackrel{ind}{\sim} N(\mathbf{l}_{k0}, \mathbf{L}_{k0}), \quad k = 2, \dots, p, \quad (3.16)$$

as derivações são simplificadas e os elementos de $\mathbf{l}$ podem ser atualizados em uma série de passos independentes

$$\mathbf{l}_k|\mathbf{u}, d_k \stackrel{ind}{\sim} N_k(\hat{\mathbf{l}}_k, \hat{\mathbf{L}}_k), \quad k = 2, \dots, p, \quad (3.17)$$

onde, tendo em mente que $\mathbf{L}_{k0}$ pode possivelmente depender de d_k , temos

$$\hat{\mathbf{L}}_k = (\mathbf{L}_{k0}^{-1} + d_k^{-1} \mathbf{X}'_k \mathbf{X}_k)^{-1}, \quad \hat{\mathbf{l}}_k = \hat{\mathbf{L}}_k(\mathbf{L}_{k0}^{-1} \mathbf{l}_{k0} + d_k^{-1} \mathbf{X}'_k \mathbf{z}_k)$$

$$\mathbf{z}_k = (u_{1k}, \dots, u_{Nk})', \quad \mathbf{X}_k = [\mathbf{z}_1 : \dots : \mathbf{z}_{k-1}].$$

■

De posse das condicionais completas obtidas na Proposição 5, para estimarmos $\mathbf{\Sigma}$ basta implementarmos o *Gibbs Sampling* para estimar os elementos de $\mathbf{L}$ e $\mathbf{D}$ via média *a posteriori*, por exemplo, e lembrarmos que $\mathbf{\Sigma} = (\mathbf{L}' \mathbf{D}^{-1} \mathbf{L})^{-1}$.

Na próxima subseção usaremos o método de estimação apresentado para $\mathbf{\Sigma}$ no caso irrestrito e adaptaremos o mesmo para o caso restrito. Lembrando que a restrição de nosso interesse é que os elementos da diagonal sejam fixos e iguais a um.

3.2.2 Caso da Matriz de Covariância Restrita

Uma importante vantagem em modelar a matriz de covariâncias como feito na subseção anterior, deve-se ao fato da possibilidade de se adaptar a mesma para o caso em que temos algum tipo de restrição. Nessa subseção apresentamos a modelagem no caso em que a restrição consiste em ter todos os elementos da diagonal de $\mathbf{\Sigma}$ iguais a 1. Para mais restrições ver [Chan e Jeliazkov \(2009\)](#).

Note que os elementos de $\mathbf{\Sigma}$ podem ser expressos em termos dos elementos de $\mathbf{D}$ e de $\mathbf{L}$, uma vez que $\mathbf{\Sigma} = \mathbf{L}^{-1} \mathbf{D} (\mathbf{L}^{-1})'$. Considerando l^{kj} os elementos abaixo da diagonal de $\mathbf{L}^{-1}$ e σ_{kj} o elemento (k, j) de $\mathbf{\Sigma}$, temos que

$$\sigma_{kk} = d_k + \sum_{j=1}^N d_j (l^{kj})^2, \quad k = 1, \dots, p, \quad (3.18)$$

$$\sigma_{kj} = l^{kj} d_j + \sum_{h=1}^{j-1} l^{kh} l^{jh} d_h, \quad 1 \leq j < k \leq p. \quad (3.19)$$

É fácil ver, a partir de (3.18) e (3.19), que impor que os elementos da diagonal de $\mathbf{\Sigma}$ sejam iguais a 1 equivale a dizer que

$$d_1 = 1, \quad (3.20)$$

$$d_k = 1 - \sum_{j=1}^{k-1} (l^{kj})^2 d_j, \quad k = 2, \dots, p. \quad (3.21)$$

As equações (3.20) e (3.21) nos dizem que sob a restrição $\sigma_{11} = \dots = \sigma_{pp} = 1$, $\mathbf{d}$ é uma função determinística de $\mathbf{l}$ e seu papel é puramente notacional, tratando-se simplesmente de um parâmetro

que será usado para manter a notação concisa e consistente. Formalmente, a distribuição *a priori* conjunta para $\mathbf{d}$ e $\mathbf{l}$ é da forma

$$f(\mathbf{d}, \mathbf{l}) \propto f(\mathbf{l})f(\mathbf{d}|\mathbf{l})\mathcal{I}(\mathbf{d} > 0), \quad (3.22)$$

onde $f(\mathbf{l})$ é a densidade implícita em (3.10) ou (3.16) dependendo da especificação *a priori*, $f(\mathbf{d}|\mathbf{l})$ assume os valores definidos de forma recursiva em (3.20) e (3.21) com probabilidade 1 e $\mathcal{I}(\cdot)$ corresponde à função indicadora. Essa representação garante que os elementos de $\mathbf{d}$ sejam positivos e satisfaçam (3.20) e (3.21), de modo que Σ é tanto positiva definida quanto em forma de correlação (ou seja, possui os elementos da diagonal igual a um e os demais estão no intervalo $[-1, 1]$).

Diante da restrição $\sigma_{11} = \dots = \sigma_{pp} = 1$ como os d_k são funções determinísticas dos l^{kj} segue que o conhecimento de Σ decorre exclusivamente do conhecimento de $\mathbf{l}$. A ideia então é construir um algoritmo para estimar os elementos de $\mathbf{l}$. Como a densidade condicional de $\mathbf{l}|\mathbf{u}$ não pertence a uma família conhecida utilizamos o algoritmo de *Metropolis-Hastings* (MH) para gerarmos uma amostra dessa distribuição *a posteriori*. Consideramos então um passeio aleatório normal centralizando a densidade proposta em $\boldsymbol{\mu} = \mathbf{l}^{(t-1)}$, onde $\mathbf{l}^{(t-1)}$ é o último passo disponível na cadeia de Markov. Fazendo uso de (3.22) uma amostra da distribuição *a posteriori* de $\mathbf{l}$ (e portanto de Σ) pode ser obtida com um passo de MH, definindo $\mathbf{l}^{(t)} = \mathbf{l}^c$ com probabilidade

$$\min \left\{ 1, \frac{V(\mathbf{u}; \Sigma^c)f(\mathbf{l}^c)\mathcal{I}(\mathbf{d}^c > 0)}{V(\mathbf{u}; \Sigma^{(t-1)})f(\mathbf{l}^{(t-1)})\mathcal{I}(\mathbf{d}^{(t-1)} > 0)} \right\} \quad (3.23)$$

e, retornando $\mathbf{l}^{(t)} = \mathbf{l}^{(t-1)}$, caso contrário.

Com isso, o processo de estimação da matriz de covariâncias pode ser resumido da seguinte maneira.

- 1) Defina $\mathbf{L}^{(0)}$ correspondente aos valores iniciais da cadeia e obtenha $\mathbf{D}^{(0)}$.
- 2) Especifique $\mathbf{l}_0$ e $\mathbf{L}_0$ que correspondem aos parâmetros da distribuição *a priori* em (3.10).
- 3) Faça $t = 1$.
- 4) Calcule $\hat{\mathbf{L}}$ e $\hat{\mathbf{l}}$, e em seguida, obtenha uma amostra candidata denotada por $\mathbf{l}^c$.
- 5) Obtenha “uma amostra candidata” de $\mathbf{d}$ seguindo as restrições dadas em (3.20) e (3.21). Denotaremos essa amostra por $\mathbf{d}^c$.
- 6) Gere q de uma distribuição uniforme.
- 7) Se $\min \left\{ 1, \frac{V(\mathbf{u}; \Sigma^c)p(\mathbf{l}^c)\mathcal{I}(\mathbf{d}^c > 0)}{V(\mathbf{u}; \Sigma^{(t-1)})p(\mathbf{l}^{(t-1)})\mathcal{I}(\mathbf{d}^{(t-1)} > 0)} \right\} > q$, faça $\mathbf{D}^{(t)} = \mathbf{D}^c$, $\mathbf{L}^{(t)} = \mathbf{L}^c$ e $t = t + 1$.
- 8) Volte ao passo 4). Considere então um *burn-in* de tamanho n_1 e saltos de tamanho s . Repita o procedimento até obter uma amostra de tamanho n_2 .

3.3 Estimação Bayesiana no Modelo 3PLMG com Dependência entre as Distribuições *a Priori*

Com base no que foi apresentado o modelo proposto nesse trabalho o qual leva em consideração uma estrutura de dependência entre as distribuições *a priori* é definido do seguinte modo:

$$\begin{aligned} Y_{ij}|\theta_j, \eta_j, \lambda_j, a_i, b_i, c_i &\sim \text{Bern}(\pi_{ij}), \\ (\mathbb{F}_{\theta_j}(\theta_j), \mathbb{F}_{\eta_j}(\eta_j), \mathbb{F}_{\lambda_j}(\lambda_j)) &\sim C(\cdot), \end{aligned}$$

em que

$$\pi_{ij} = \mathbb{P}(Y_{ij} = 1 | \theta_j, \eta_j, \lambda_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))} P_i(\eta_j, \lambda_j),$$

$$P_i(\eta_j, \lambda_j) = \min \left\{ 1, \left[1 - \left(\frac{i}{I} - \eta_j \right) \right]^{\lambda_j} \right\}$$

e $C(\cdot)$ corresponde a uma cópula normal com vetor de médias nulo e matriz de covariâncias Σ dada por

$$\Sigma = \begin{bmatrix} 1 & \rho_{\theta\eta} & \rho_{\theta\lambda} \\ \rho_{\theta\eta} & 1 & \rho_{\eta\lambda} \\ \rho_{\theta\lambda} & \rho_{\eta\lambda} & 1 \end{bmatrix}.$$

Assim, de acordo com o que foi estudado temos então condições de realizar a estimação Bayesiana no modelo 3PLMG com dependência *a priori* usando um algoritmo MCMC. Considerando o modelo acima apresentado e que a dependência entre os traços latentes é realmente modelada por uma cópula normal multivariada temos que o vetor de parâmetros desconhecidos é $\zeta = (\mathbf{a}, \mathbf{b}, \mathbf{c}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda}, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\eta\lambda})$. Sob essas condições, e especificando as distribuições *a priori* para os parâmetros, temos a seguinte estrutura hierárquica.

$$\begin{aligned} y_{ij} &\sim \text{Bern}(\pi_{ij}) \\ a_i &\sim N(\mu_a, \sigma_a^2) I(a_i > 0) \\ b_i &\sim N(\mu_b, \sigma_b^2) \\ c_i &\sim \text{Beta}(e_1, e_2) \\ \theta_j &\sim N(0, 1) \\ \eta_j &\sim \text{Beta}(\alpha, \beta) \\ \lambda_j &\sim \text{LN}(\mu_\lambda, \sigma_\lambda^2) \\ (\mathbb{F}_{\theta_j}(\theta_j), \mathbb{F}_{\eta_j}(\eta_j), \mathbb{F}_{\lambda_j}(\lambda_j)) &\sim C(\cdot), \end{aligned}$$

em que $C(\cdot)$ é uma cópula normal multivariada com média zero e matriz de covariâncias Σ . A posteriori conjunta para o modelo em questão é dada por,

$$\begin{aligned} \pi(\zeta | \mathbf{y}) &\propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \\ &\times \prod_{i=1}^I \exp(-(a_i - \mu_a)^2 / 2\sigma_a^2) I(a_i > 0) \\ &\times \prod_{i=1}^I \exp(-(b_i - \mu_b)^2 / 2\sigma_b^2) \\ &\times \prod_{i=1}^I c_i^{e_1-1} (1 - c_i)^{e_2-1} \\ &\stackrel{\text{Prop. (4)}}{\times} \prod_{j=1}^J \frac{f_2(\eta_j)}{|\Sigma|^{1/2} \lambda_j \phi(\Phi^{-1}(F_2(\eta_j)))} \exp(-\boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma} / 2) \\ &\times \pi(\Sigma), \end{aligned}$$

em que $\pi(\mathbf{\Sigma})$ é a distribuição *a priori* para $\mathbf{\Sigma}$ e π_{ij} é uma função de a_i , b_i , c_i , θ_j , η_j e λ_j como definido em (2.7). Assim, obtemos as seguintes condicionais completas;

$$\begin{aligned}
 \pi(\mathbf{a}|\mathbf{y}, \mathbf{b}, \mathbf{c}, \mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\eta\lambda}) &\propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \\
 &\times \prod_{i=1}^I \exp(-(a_i - \mu_a)^2 / 2\sigma_a^2) I(a_i > 0), \\
 \pi(\mathbf{b}|\mathbf{y}, \mathbf{a}, \mathbf{c}, \mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\eta\lambda}) &\propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \prod_{i=1}^I \exp(-(b_i - \mu_b)^2 / 2\sigma_b^2), \\
 \pi(\mathbf{c}|\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\eta\lambda}) &\propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \prod_{i=1}^I c_i^{e_1-1} (1 - c_i)^{e_2-1}, \\
 \pi(\mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}|\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \rho_{\theta\eta}, \rho_{\theta\lambda}, \rho_{\eta\lambda}) &\propto \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \\
 &\times \prod_{j=1}^J \frac{f_2(\eta_j)}{|\mathbf{\Sigma}|^{1/2} \lambda_j \phi(\Phi^{-1}(F_2(\eta_j)))} \exp(-\boldsymbol{\gamma}' \mathbf{\Sigma}^{-1} \boldsymbol{\gamma} / 2), \\
 \pi(\mathbf{\Sigma}|\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}) &\propto \prod_{j=1}^J \frac{\exp(-\boldsymbol{\gamma}' \mathbf{\Sigma}^{-1} \boldsymbol{\gamma} / 2)}{|\mathbf{\Sigma}|^{1/2}} \times \pi(\mathbf{\Sigma}).
 \end{aligned}$$

De posse das condicionais completas podemos então usar o algoritmo de Gibbs. Para gerar valores das quatro primeiras condicionais completas podemos usar o *slice sampling* e para amostrar da distribuição de $\mathbf{\Sigma}|\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{\theta}, \mathbf{\eta}, \mathbf{\lambda}$ usamos o procedimento apresentado na Subseção 3.2.2.

Note que temos um passo de Metrópolis em Gibbs. Conforme [Paulino et al. \(2003, pág. 332\)](#) para gerarmos um valor da distribuição alvo o procedimento seria o seguinte: deve-se gerar uma cadeia de tamanho T de modo a garantir a convergência para a distribuição alvo (distribuição de equilíbrio) e com isso um valor amostral de interesse seria o obtido na iteração T . Porém, estudos teóricos e práticos mostram que basta considerar $T = 1$. Dessa forma, durante a implementação do processo de estimação consideramos $T = 1$.

Uma observação importante a ser feita é que no caso em que $\mathbf{\Sigma}$ corresponde a matriz identidade temos o modelo com independência entre os traços latentes. Sendo assim caso desejamos estimar o modelo considerando estrutura de independência entre os traços latentes basta fixar $\mathbf{\Sigma}$ como matriz identidade e desconsiderar o passo de metropolis.

Capítulo 4

Estudo de Simulação

Nesse capítulo, no intuito de analisar a recuperação dos parâmetros dos modelos 2PLMG tanto com a suposição de independência entre as distribuições *a priori* quanto com a suposição de dependência entre as distribuições *a priori* fizemos o seguinte. A escolha do modelo 2PLMG para esse estudo de simulação deve-se ao fato de termos utilizado o mesmo para estudar os dados do EGRA. Consideramos como verdadeiras as estimativas obtidas na aplicação do EGRA, apresentadas, graficamente, no Capítulo 5 e que satisfazem $\rho_{\theta\eta} = 0.49$, $\rho_{\theta\lambda} = -0.177$ e $\rho_{\eta\lambda} = -0.723$. Assim, considerando haver dependência entre os traços latentes dos indivíduos simulamos 10 conjuntos contendo as respostas de 512 indivíduos a 50 itens. Posteriormente, ajustamos os modelos 2PLMG para cada réplica. Na Seção 4.1 consideramos distribuições *a priori* independentes para os traços latentes dos indivíduos e na Seção 4.2 distribuições *a priori* dependentes para θ , η e λ sendo essa dependência modelada por uma cópula normal.

4.1 Modelo com distribuições *a priori* independentes

No que segue, quando nos referirmos a parâmetros estaremos fazendo referência a quaisquer quantidades desconhecidas e usaremos os termos parâmetros dos itens ou traços latentes quando quisermos ser mais específicos. Para a estimação dos parâmetros dos itens e dos traços latentes correspondentes aos indivíduos fizemos uso da abordagem totalmente Bayesiana, assim como proposto em [Bazan et al. \(2010\)](#). O uso dessa abordagem é feito quando especificamos distribuições *a priori* para todos os parâmetros. Assim, ao invés de usarmos estimadores *plug-in* em um procedimento de dois passos realizamos uma estimação conjunta dos parâmetros com base na distribuição *a posteriori* completa. Nessa seção apresentamos as estimativas obtidas dos parâmetros do modelo 2PLMG considerando que as distribuições *a priori* são independentes. Para estimação dos parâmetros desses modelos consideramos em (2.6) a reparametrização $a_i\theta_j - b_i^*$ ao invés de $a_i(\theta_j - b_i)$ e implementamos o algoritmo de Gibbs (*Gibbs Sampling*) similar ao da Seção 3.3, porém sem o passo de metropolis uma vez que neste caso Σ corresponde a matriz identidade de ordem 3. O uso dessa reparametrização é usual na TRI, pois reduz o tempo computacional significativamente. As especificações das distribuições *a priori* foram feitas do seguinte modo: Para os parâmetros dos itens seguimos [Bazan et al. \(2010\)](#), ou seja, consideramos que $a_i \sim N(1, 0.5)I(0,)$ e $b_i^* \sim N(0, 2)$, onde $b_i = b_i^*/a_i$ é o parâmetro de dificuldade e bs_i é o parâmetro de intercepto de $a_i\theta_j - b_i^*$; Para os traços latentes dos indivíduos consideramos que $\theta_j \sim N(0, 1)$ (como é feito em geral na TRI); $\eta_j \sim Beta(1, 1)$, ou seja, atribuímos uma distribuição *a priori* não informativa ao mesmo; por fim consideramos que $\lambda_j \sim LN(1, 1)$. A especificação de λ_j como sendo uma distribuição log-normal com parâmetros $\mu_\lambda = 1$ e $\sigma_\lambda^2 = 1$ ocorreu porque $E(\lambda_j) \approx 4.48$ e $var(\lambda_j) \approx 34.51$. Dessa forma essa distribuição *a priori* é menos informativa que a distribuição $LN(0, 1)$, proposta por [Bazan et al. \(2010\)](#), que possui 1.64 e 4.67 como sendo sua média e variância, respectivamente.

O modelo foi ajustado utilizando o *software* Matlab versão 2013a em um computador com processador intel i7 e 8Gb de memória ram. Para o ajuste consideramos um *burn-in* de 24.000 iterações com saltos de tamanho 20, sendo o tamanho dos saltos suficiente para eliminar as autocorrelações

(os gráficos foram omitidos devido à grande quantidade de parâmetros). Assim, após 64.000 iterações foram obtidas amostras de tamanho 2.000 para cada um dos parâmetros do modelo citado e para cada uma das 10 réplicas em estudo. As convergências das cadeias foram verificadas utilizando o pacote *coda* (Plummer *et al.*, 2006) do *software* R no qual os seguintes métodos foram analisados: método de Geweke (Geweke *et al.*, 1991) e o método de Gelman e Rubin (Gelman e Rubin, 1992) (neste último, foram consideradas duas cadeias paralelas). A integração entre os *softwares* R e Matlab foi feita através do pacote R.matlab de Bengtsson (2015). Através das amostras das distribuições *a posteriori* dos parâmetros η_j e λ_j notamos que as mesmas eram assimétricas, sendo assim como estimativa dos traços latentes, bem como dos parâmetros associados aos itens, adotamos a mediana das distribuições *a posteriori*.

Nas Figuras 4.1 e 4.2 temos os *boxplots* das estimativas dos parâmetros de discriminação e dificuldade, respectivamente, para as 10 réplicas. Os *boxplots* foram dispostos lado a lado sendo o primeiro *boxplot* dos dados reais e os demais correspondentes as respectivas estimativas obtidas em cada réplica. De acordo com os mesmos podemos notar uma grande semelhança entre os valores reais e estimados, indicando um bom ajuste dos parâmetros dos itens.

Figura 4.1: *Boxplots dos valores reais dos parâmetros de discriminação, a_i , e das estimativas obtidas nas réplicas de 1 à 10 usando distribuições a priori independentes para os traços latentes. Consideramos que $a_i \sim N(1; 0, 5)I(0,)$.*

Figura 4.2: *Boxplots dos valores reais dos parâmetros de intercepto, b_i^* , e das estimativas obtidas para as 10 réplicas usando distribuições a priori independentes para os traços latentes. Consideramos que $b_i^* \sim N(0, 2)$.*

Similarmente aos *boxplots* apresentados para os parâmetros dos itens construímos os *boxplots* associados aos traços latentes dos indivíduos. Nas Figuras de 4.3 à 4.5 temos os *boxplots* correspondentes às estimativas dos θ_j , η_j e λ_j , respectivamente. De acordo com os mesmos observamos um comportamento para as estimativas dos traços latentes, θ_j e η_j , similar aos dos valores considerados reais para cada uma das réplicas, uma vez que os primeiros e terceiros quartis, bem como a mediana de ambos, ficaram bastante próximos. Com relação às estimativas dos λ_j 's, notamos que para as réplicas, o terceiro quartil em geral foi bastante menor que o real. Cabe ressaltar que isso não necessariamente significa que o modelo não estimou bem tais parâmetros. Como consideramos as estimativas obtidas do EGRA como sendo os verdadeiros valores dos parâmetros, construímos as Tabelas 4.1 e 4.2 para esclarecer essa diferença nas estimativas dos λ_j 's. Nas mesmas dispomos alguns padrões de respostas (dicotomizados) conjuntamente com o valor do fator de penalização e a probabilidade de um indivíduo responder corretamente ao respectivo item dado os seus traços latentes e os parâmetros dos itens. Conforme observamos na Tabela 4.1 isso ocorre porque em geral esse modelo possui dificuldade em replicar (escore nulo) respostas de indivíduos que sofrem o efeito *speededness* desde o primeiro item, o que corresponde a um caso extremo e inesperado. Note no primeiro quadro da Tabela 4.1 que mesmo um indivíduo que errou todos os itens o esperado era que na verdade ele venha a acertar pelo menos o primeiro, pois a probabilidade de que o mesmo viesse a acertá-lo é de 0,63. Além disso, também temos o fato do parâmetro λ ser sensível a mudanças nos padrões de respostas, para ver isso basta compararmos as diferenças dos valores dos mesmos nas Tabelas 4.1 e 4.2. Além disso, com base nessas tabelas notamos que as estimativas desses parâmetros estão coerentes com os padrões de respostas.

Item	Resposta	Penalização	Prob de acerto	Resposta	Penalização	Prob de acerto
1	0	0.65	0.63	1	1	0.99
2	0	0.35	0.33	1	1	0.98
3	0	0.18	0.18	1	1	1
4	0	0.09	0.09	1	1	0.99
5	0	0.05	0.05	1	1	0.98
6	0	0.02	0.02	1	1	0.99
7	0	0.01	0.01	1	1	0.99
8	0	0.01	0.01	1	1	0.95
9	0	0	0	1	1	1
10	0	0	0	1	1	1
11	0	0	0	1	1	0.98
12	0	0	0	1	1	0.99
13	0	0	0	1	1	0.99
14	0	0	0	1	1	0.97
15	0	0	0	1	1	0.99
16	0	0	0	1	1	0.98
17	0	0	0	1	1	0.98
18	0	0	0	1	1	0.99
19	0	0	0	1	1	0.99
20	0	0	0	1	1	1
21	0	0	0	1	1	1
22	0	0	0	1	1	0.96
23	0	0	0	1	1	0.91
24	0	0	0	1	1	1
25	0	0	0	1	1	0.96
26	0	0	0	1	1	1
27	0	0	0	1	1	0.99
28	0	0	0	1	1	0.99
29	0	0	0	1	1	0.99
30	0	0	0	1	1	1
31	0	0	0	1	1	1
32	0	0	0	1	1	1
33	0	0	0	1	1	1
34	0	0	0	1	1	1
35	0	0	0	1	1	1
36	0	0	0	1	1	1
37	0	0	0	1	1	1
38	0	0	0	1	1	1
39	0	0	0	1	1	1
40	0	0	0	1	1	1
41	0	0	0	1	1	1
42	0	0	0	1	1	1
43	0	0	0	1	1	0.99
44	0	0	0	1	1	0.99
45	0	0	0	1	1	0.98
46	0	0	0	1	1	0.97
47	0	0	0	1	0.99	0.96
48	0	0	0	1	0.96	0.92
49	0	0	0	1	0.93	0.87
50	0	0	0	1	0.90	0.80

Tabela 4.1: No quadro a esquerda temos os fatores de penalização aplicados a cada item bem como a probabilidade de resposta correta para um indivíduo que teve $\eta = 0.006$ e $\lambda = 30.65$ e escore nulo enquanto no quadro ao lado direito temos um indivíduo que obteve escore perfeito e $\eta = 0.935$ e $\lambda = 1.525$.

Item	Resposta	Penalização	Prob de acerto	Resposta	Penalização	Prob de acerto
1	1	1	0.96	1	1	0.94
2	0	1	0.93	1	1	0.91
3	1	0.96	0.94	1	1	0.95
4	0	0.92	0.89	1	1	0.96
5	1	0.87	0.80	1	1	0.88
6	1	0.83	0.80	1	1	0.95
7	1	0.79	0.75	0	1	0.93
8	0	0.75	0.68	1	1	0.90
9	1	0.71	0.68	1	1	0.92
10	0	0.67	0.66	1	1	0.98
11	1	0.63	0.61	1	1	0.96
12	1	0.60	0.58	1	1	0.95
13	1	0.56	0.55	1	1	0.97
14	1	0.53	0.46	1	1	0.83
15	1	0.50	0.48	1	1	0.97
16	1	0.47	0.45	1	1	0.96
17	0	0.44	0.40	1	1	0.89
18	0	0.41	0.40	1	1	0.96
19	0	0.38	0.37	1	1	0.98
20	0	0.35	0.35	1	1	0.99
21	0	0.33	0.32	1	1	0.97
22	0	0.30	0.25	1	1	0.79
23	1	0.28	0.23	1	1	0.78
24	1	0.26	0.26	1	1	0.99
25	0	0.24	0.22	1	1	0.92
26	0	0.22	0.21	1	1	0.96
27	0	0.20	0.19	1	1	0.97
28	1	0.18	0.17	1	1	0.93
29	0	0.16	0.16	1	1	0.94
30	0	0.15	0.14	1	1	0.91
31	0	0.13	0.11	1	1	0.61
32	0	0.12	0.09	0	1	0.33
33	0	0.10	0.07	0	1	0.23
34	0	0.09	0.05	0	1	0.14
35	0	0.08	0.03	0	1	0.07
36	0	0.07	0.02	0	0.97	0.03
37	0	0.06	0.01	0	0.92	0.02
38	0	0.05	0.01	0	0.87	0.01
39	0	0.04	0	0	0.82	0
40	0	0.04	0	0	0.77	0
41	0	0.03	0	0	0.73	0
42	0	0.02	0	0	0.69	0
43	0	0.02	0	0	0.64	0
44	0	0.01	0	0	0.60	0
45	0	0.01	0	0	0.57	0
46	0	0.01	0	0	0.53	0
47	0	0	0	0	0.49	0
48	0	0	0	0	0.46	0
49	0	0	0	0	0.43	0
50	0	0	0	0	0.40	0

Tabela 4.2: No quadro a esquerda temos os fatores de penalização aplicados a cada item bem como a probabilidade de resposta correta para um indivíduo que teve $\eta = 0.044$ e $\lambda = 2.35$, enquanto no quadro ao lado direito apresentamos os mesmos dados para um indivíduo que possui $\eta = 0.708$ e $\lambda = 2.665$.

Figura 4.3: Boxplots dos valores reais das habilidades, θ_j , e das estimativas obtidas para as réplica de 1 à 10 obtidas usando distribuições a priori independentes para os traços latentes. Foi considerado que $\theta_j \sim N(0, 1)$.

Figura 4.4: Boxplots dos valores reais de tolerância ao speededness, η_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições a priori independentes para os traços latentes. Consideramos que $\eta_j \sim \text{Beta}(1, 1)$.

Figura 4.5: Boxplots dos valores reais das propensões ao speededness, λ_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições a priori independentes para os traços latentes. Consideramos que $\lambda \sim LN(1,1)$.

Na Tabela 4.3 apresentamos as correlações médias das estimativas obtidas para cada uma das réplicas com os valores reais dos correspondentes parâmetros. De acordo com as mesmas notamos que há uma alta correlação linear indicando também um bom ajuste do modelo.

Parâmetros	Média	Mínimo	Máximo	Desvio Padrão
a	0.98	0.98	0.99	<0.01
b	0.99	0.99	0.99	<0.01
θ	0.92	0.91	0.93	<0.01
η	0.97	0.97	0.97	<0.01
λ	0.88	0.86	0.90	<0.01

Tabela 4.3: Resumo das correlações reais e estimadas

Por fim, na Tabela 4.4 apresentamos as médias das estimativas das correlações induzidas *a posteriori* nos traços latentes e notamos que houve uma boa recuperação, quando comparado com os valores reais, mesmo utilizando o modelo supondo a independência entre as distribuições *a priori* dos traços latentes.

Correlação	$\rho_{\theta\eta}$	$\rho_{\theta\lambda}$	$\rho_{\eta\lambda}$
real	0.49	-0.18	-0.72
estimada	0.41	-0.10	-0.70

Tabela 4.4: Tabela contendo as correlações reais entre os traços latentes e as correlações médias estimadas obtidas através das 10 réplicas

Construímos também os intervalos HPD's com 95% de confiança para cada parâmetro em cada uma das 10 réplicas e observamos o percentual de vezes que os valores reais estavam no correspondente intervalo. Na Tabela 4.5 são apresentados esses percentuais em valores aproximados através dos quais notamos que em geral, estão próximos do esperado, com exceção do parâmetro de dificuldade, o que pode estar ocorrendo devido ao fato de estarmos levando em consideração somente 10 réplicas.

Parâmetros	a	b^*	θ	η	λ
Percentual de Cobertura	0.92	0.82	0.99	0.98	0.97

Tabela 4.5: *Percentual de cobertura dos hpds com 95% de credibilidade para o modelo que considera independência entre as distribuições a priori para os traços latentes*

Como um questionamento natural a ser feito podemos nos perguntar se o uso de um modelo que leva em consideração uma dependência entre as distribuições *a priori* não nos conduziria a um ajuste melhor. Para responder a essa pergunta faremos então uma análise similar a feita nessa seção na próxima seção, porém levaremos em consideração uma estrutura dependência entre as distribuições *a priori*.

4.2 Modelo com distribuições *a priori* dependentes

Nessa seção, diferentemente da seção anterior, faremos um estudo de simulação estimando os parâmetros do modelo 2PLMG considerando que *a priori* há uma relação de dependência entre os traços latentes dos indivíduos. Conforme visto anteriormente essa estrutura de dependência é construída de modo a preservar o conhecimento prévio sobre cada parâmetro e para isso usaremos uma cópula normal. Novamente faremos uso da reparametrização $a_i\theta_j - b_i^*$ ao invés de $a_i(\theta_j - b_i)$. As especificações das distribuições *a priori* foram feitas conforme a seção anterior. Para os parâmetros dos itens seguimos Bazan *et al.* (2010), ou seja, consideramos que $a_i \sim N(1, 0.5)I(0,)$ e $b_i^* \sim N(0, 2)$, onde $b_i = b_i^*/a_i$ enquanto que para os traços latentes dos indivíduos consideramos que $\theta_j \sim N(0, 1)$, $\eta_j \sim Beta(1, 1)$, e $\lambda_j \sim LN(1, 1)$. Porém, conforme mencionado o diferencial dessa seção está no fato de que iremos supor que conjuntamente θ , η e λ são modelados *a priori* por uma cópula normal com matriz de correlação desconhecida.

Novamente ajustamos o modelo usando o *software* Matlab onde implementamos o algoritmo de Gibbs apresentado na Seção 3.3 que consiste de um algoritmo de Gibbs com um passo de Metropolis-Hastings. Consideramos então um *burn-in* de 60.000 iterações com saltos de tamanho 20, sendo o tamanho dos saltos suficiente para eliminar as autocorrelações. Assim, após 100.000 iterações foram obtidas uma amostra de tamanho 2.000 para cada um dos parâmetros do modelo em estudo. As convergências das cadeias foram verificadas utilizando o pacote *coda* (Plummer *et al.*, 2006) do *software* R no qual os seguintes métodos foram analisados: método de Geweke e o método de Gelman e Rubin (neste último, foram consideradas duas cadeias paralelas). A integração entre os *softwares* R e Matlab foi feita novamente através do pacote R.matlab de Bengtsson (2015). Como estimativas pontuais dos respectivos parâmetros consideramos a mediana de suas distribuições *a posteriori*.

As estimativas dos parâmetros, em geral, apresentaram bastante semelhança com as obtidas na seção anterior conforme podemos ver nos *boxplots* apresentados nas Figuras de 4.6 à 4.10, levando consequentemente, praticamente às mesmas correlações estimadas entre os parâmetros dos indivíduos conforme vemos na Tabela 4.6 bem como os mesmos percentuais de cobertura conforme a Tabela 4.7 e por fim às mesmas interpretações da seção anterior, a qual omitiremos para não nos tornarmos repetitivos.

Correlação	$\rho_{\theta\eta}$	$\rho_{\theta\lambda}$	$\rho_{\eta\lambda}$
real	0.49	-0.18	-0.72
estimada	0.42	-0.09	-0.70

Tabela 4.6: Tabela contendo as correlações reais entre os traços latentes e as correlações médias estimadas obtidas através das 10 réplicas considerando distribuições a priori dependentes para os traços latentes

Parâmetros	a	b^*	θ	η	λ
Percentual de Cobertura	0.92	0.82	0.99	0.98	0.96

Tabela 4.7: Percentual de cobertura dos hpds com 95% de credibilidade para o modelo que considera dependência entre as distribuições a priori para os traços latentes

Figura 4.6: Boxplots dos valores reais dos parâmetros de discriminação, a_i , e das estimativas obtidas nas réplicas de 1 à 10 usando distribuições a priori dependentes para os traços latentes. Consideramos que $a_i \sim N(1; 0, 5)I(0,)$.

Figura 4.7: *Boxplots dos valores reais dos parâmetros de intercepto, b_i^* , e das estimativas obtidas para as 10 réplicas usando distribuições a priori dependentes para os traços latentes. Consideramos que $b_i^* \sim N(0, 2)$.*

Figura 4.8: *Boxplots dos valores reais das habilidades, θ_j , e das estimativas obtidas para as réplica de 1 à 10 obtidas usando distribuições a priori dependentes para os traços latentes. Foi considerado que $\theta_j \sim N(0, 1)$.*

Figura 4.9: *Boxplots dos valores reais de tolerância ao speededness, η_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições a priori dependentes para os traços latentes. Consideramos que $\eta_j \sim \text{Beta}(1,1)$.*

Figura 4.10: *Boxplots dos valores reais das propensões ao speededness, λ_j , e das estimativas para as réplicas de 1 à 10 obtidas usando distribuições a priori dependentes para os traços latentes. Consideramos que $\lambda \sim \text{LN}(1,1)$.*

Diante disso, vamos fazer agora uma comparação das estimativas obtidas em ambos os modelos.

4.3 Comparações dos modelos sem/com dependência entre as distribuições *a priori* e conclusões do estudo de simulação

Com base nas estimativas obtidas, nas Seções (4.1) e (4.2), para cada uma das réplicas construímos duas tabelas resumo. A Tabela 4.8 apresenta as medidas resumo dos valores reais dos parâmetros e as estimativas obtidas considerando o modelo sob a hipótese de que as distribuições *a priori* são independentes para os traços latentes. Já a Tabela 4.9 apresenta as medidas resumos dos valores reais dos parâmetros e as estimativas obtidas considerando o modelo sob a hipótese de dependência entre as distribuições *a priori*. Comparando as medidas resumo dos valores reais e as obtidas considerando distribuições *a priori* independentes notamos que as mesmas são bastante próximas indicando um bom ajuste do modelo independentemente da suposição de dependência das distribuições *a priori* dos traços latentes. Para ambos os casos o desvio absoluto mediano das estimativas do vetor λ apresentou-se relativamente maior que os demais o que é aceitável já que a escala de variação dos mesmos é bem maior. Interpretação similar é feita quando observamos o modelo com distribuições *a priori* dependentes na Tabela 4.9.

Parâmetros	Real		Estimado		Desvio Absoluto Mediano
	Mediana	D.P	Mediana	Desvio Padrão	
a	0.71	1.43	0.78	1.18	0.27
b	-2.94	2.62	-2.86	2.24	0.28
θ	-0.15	1.01	-0.05	0.91	0.21
η	0.67	0.27	0.65	0.26	0.03
λ	3.37	6.11	3.11	4.81	0.86

Tabela 4.8: Comparativo das medidas resumo dos valores reais dos parâmetros conjuntamente com as medidas resumo obtidas das estimativas obtidas considerando distribuições *a priori* independentes para as 10 réplicas estudadas.

Parâmetros / Estatísticas	Real		Estimado		Desvio Absoluto Mediano
	Média	D.P	Média	Desvio Padrão	
a	0.71	1.43	0.78	1.18	0.27
b	-2.94	2.62	-2.86	2.24	0.28
θ	-0.15	1.01	-0.05	0.91	0.21
η	0.67	0.27	0.66	0.26	0.03
λ	3.37	6.11	3.10	4.82	0.89

Tabela 4.9: Comparativo das medidas resumo dos valores reais dos parâmetros dos itens e dos traços latentes conjuntamente com as medidas resumo obtidas das estimativas, das 10 réplicas, considerando distribuições *a priori* dependentes.

Comparando então as estimativas dos parâmetros obtidas para cada réplica levando em consideração os dois modelos em questão acabamos observando na Tabela 4.10 que as mesmas são extremamente próximas. Além disso comparando o comprimento médio dos intervalos HPD's com 95% para os modelos sem/com dependência *a priori* para os traços latentes, nas Tabelas 4.11 e 4.12, respectivamente, observamos que as estimativas do modelo com a suposição dependência entre as distribuições *a priori* são um pouco mais precisas, uma vez que o comprimento médio é um pouco menor. Porém pode-se observar que a diferença entre os comprimentos médios também é extremamente pequena, ou seja, acabamos concluindo que sob as condições do EGRA temos que a suposição de independência entre as distribuições *a priori* é preferível uma vez que diante da mesma as cadeias convergiram mais rapidamente necessitando de menos iterações no algoritmo. Além disso, a suposição de independência entre as distribuições *a priori* possui a vantagem de ser facilmente implementada requerendo um aparato teórico bem menor.

Parâmetros	a	b	θ	η	λ
Diferença média	0.00048	0.00118	0.00026	0.00013	0.03367

Tabela 4.10: Diferenças médias entre as estimativas obtidas considerando o modelo com distribuições *a priori* dependentes com as do modelo considerando distribuições *a priori* independentes.

Parâmetros	a	b	θ	η	λ
Comprimento médio	1.154	1.372	2.545	0.390	16.855

Tabela 4.11: Comprimento médio dos intervalos para HPD's com 95% de credibilidade para os parâmetros do modelo com independência entre as distribuições *a priori* para os traços latentes quando analisado as 10 réplicas.

Parâmetros	a	b	θ	η	λ
Comprimento médio	1.152	1.372	2.543	0.389	16.404

Tabela 4.12: Comprimento médio dos intervalos para HPD's com 95% de credibilidade para os parâmetros do modelo com dependência entre as distribuições *a priori* para os traços latentes quando analisado as 10 réplicas.

Para efeito ilustrativo do que ocorre caso utilizemos o modelo 2PL com dados na presença de efeito *speededness* chegamos a ajustar o mesmo para cada uma das réplicas. Nas Figuras de 4.11 a 4.13 apresentamos os valores dos parâmetros estimados com o modelo 2PL para a réplica 1 versus os valores considerados reais. Como ocorreu um comportamento similar para as estimativas das demais réplicas apresentamos somente as estimativas dos parâmetros obtidas para réplica 1. De acordo com as Figuras 4.11 e 4.12 notamos que o uso do modelo 2PL na presença de *speededness* acaba superestimando tanto o valor da discriminação dos itens quanto os respectivos valores do parâmetro de dificuldade dos mesmos. Além disso, pela Figura 4.13 notamos que o uso do modelo 2PL também acaba subestimando os valores dos θ 's, sendo portanto inadequado seu uso em situações que ocorre *speededness*.

Figura 4.11: Comparação das estimativas dos parâmetros de discriminação do modelo 2PL com os seus correspondentes valores reais.

Figura 4.12: Comparação das estimativas dos parâmetros de intercepto do modelo 2PL com os seus respectivos valores reais.

Figura 4.13: Comparação das estimativas dos θ 's do modelo 2PL com os valores reais das habilidades.

Capítulo 5

Aplicação

Nesse Capítulo, faremos uma aplicação dos modelos de dois parâmetros logito (2PL) e do modelo de dois parâmetros logito com processo de mudança de resposta gradual (2PLMG) (ou seja, em ambos os modelos $c_i = 0, \forall i = 1, \dots, I$) com dados na presença de *speededness*. Para isso buscamos um conjunto de dados que se adeque à situação proposta pelo modelo (2.6), ou seja, uma situação em que esteja evidente a presença do efeito de *speededness* como um fator da limitação do tempo nas respostas dos indivíduos.

5.1 Descrição dos dados

Estudos do EdData II¹ (*education data for decision making*) comprovam que as crianças que não aprendem a ler nos primeiros anos são mais propensas a repetirem de ano posteriormente. Diante da importância desse fato, aparece a necessidade de se avaliar a capacidade de leitura das crianças nessa fase inicial. Uma metodologia denominada de EGRA (*Early Grade Reading Assessment*), desenvolvida pelo EdData II, vem sendo utilizada nesse intuito. O EGRA corresponde a uma avaliação dos alunos por via oral (ou seja, o aluno ler a palavra e sua resposta é dicotomizada como certa, 1, ou errada, 0) e foi desenvolvida para medir as habilidades fundamentais mais básicas para alfabetização nas séries iniciais, são elas: reconhecer as letras do alfabeto; ler palavras simples; compreender frases e parágrafos; e ouvir com compreensão.

O EdData II desenvolveu a metodologia EGRA e a aplicou em 11 países e em 19 línguas. Posteriormente, essa metodologia foi adotada e implementada por outros parceiros em mais de 30 países e em mais de 60 outras línguas. Os dados do EGRA tem sido utilizados para analisar a prática docente de forma rigorosa, mas de um modo fácil de entender. Muitos países tem mostrado interesse em utilizá-la como uma ferramenta para avaliar e melhorar o aprendizado da leitura, e passaram a redesenhar a formação de seus professores com base nos resultados. Nesse trabalho, consideramos um teste de leitura de palavras sem sentido (*Nonsense Words Fluency* - NWF), que corresponde a uma das etapas de avaliação do EGRA, realizado no Peru de setembro de 2007 a fevereiro de 2008. Uma cópia da metodologia utilizada nesse exame, bem como o relatório final, pode ser encontrado em <https://www.eddataglobal.org>.

Os dados obtidos do NWF consistem das respostas dicotomizadas de 512 alunos de 8 anos de idade que foram submetidos à leitura de 50 palavras sem sentidos, dentro de um tempo limite de 60 segundos. Esse conjunto de dados também foi trabalhado em Bazan *et al.* (2010). É importante ressaltar que se um indivíduo não conseguisse no decorrer do teste chegar numa determinada palavra, a mesma era considerada incorreta e que tal consideração pode ocasionar um viés nas estimativas dos parâmetros caso seja utilizado um modelo que não leve em consideração o efeito ocasionado pelo *speededness*. Um fato interessante desse banco de dados é que 12 indivíduos tiveram escore nulo e 39 escores perfeitos, totalizando praticamente 10% dos respondentes.

¹corresponde a um mecanismo de contrato cujo papel principal, de acordo com o documento de concessão, é melhorar “a exatidão, pontualidade e acessibilidade, assim como o uso de dados para a política de educação básica e planejamento do programa”.

Na Figura 5.1 temos a proporção de respostas corretas para cada item do subteste NWF, nela podemos observar que quanto mais próximo do fim do teste, menor é a proporção de indivíduos que respondem o item corretamente, o que é um indício da presença de *speededness* no teste.

Figura 5.1: *Proporção de respostas corretas - EGRA*

Apesar de termos um indício de *speededness*, não podemos afirmar com certeza a ocorrência do mesmo apenas observando a Figura 5.1, pois pode ocorrer dos itens no final do teste serem realmente mais difíceis fazendo com que os indivíduos não possuam habilidade suficiente para respondê-los. Uma alternativa para quantificar o efeito de *speededness* foi dada por Stafford (1971) e discutida posteriormente em Lu e Sireci (2007). Sob a hipótese de que os respondentes possuem a noção do tempo durante o teste, o que parece ocorrer no EGRA, Stafford (1971) sugere usar uma medida chamada de *speededness quotient* definida como

$$SQ = \frac{\sum_{j=1}^n U_j}{\sum_{j=1}^n W_j + \sum_{j=1}^n U_j}, \quad (5.1)$$

em que n corresponde ao número de respondentes, U_j é o número de itens não atingidos pelo indivíduo j e W_j é o número de itens respondidos incorretamente ou não atingidos pelo respondente j . Essa medida tem a vantagem de ser facilmente calculada e interpretada, além disso, a mesma é invariante sob a quantidade de itens e de respondentes. A interpretação de SQ é feita de modo que se dentre todas as questões consideradas como respondidas incorretamente houver uma grande proporção de questões que não chegaram a ser respondidas, então o teste apresentou *speededness*. Em outras palavras, quanto mais próximo de 1 o SQ estiver, maior é o efeito *speededness* no teste.

Para o conjunto de dados do EGRA, foi obtido $SQ = 0.9081$, o que nos embasa a usar o modelo (2.6). Como o NWF consistia de uma avaliação oral levamos em consideração a utilização do modelo 2PLMG ao invés do modelo 3PLMG descartando a possibilidade de acerto casual, um vez que a prova era oral e não objetiva. É importante salientar que também chegamos a ajustar os modelos 3PL e 3PLMG para esse conjunto de dados e obtivemos valores de c_i extremamente próximo de zero o que condiz com nossa proposta de uso dos modelos 2PL e 2PLMG, ou seja, consideramos $c_i = 0$, $\forall i$ nos modelos 3PL e 3PLMG, respectivamente. Como no processo de estimação usando a abordagem totalmente Bayesiana devemos especificar as distribuições *a priori*, realizamos uma análise de sensibilidade para os mesmos, a qual será apresentada a seguir.

5.2 Análise de Sensibilidade

A verificação ou avaliação do ajuste de um modelo é uma parte crucial de qualquer análise estatística. Stern e Sinharay (2005) citam que antes de tirarmos conclusões definitivas a partir da aplicação de um modelo estatístico para um conjunto de dados, devemos avaliar o ajuste do modelo para nos certificarmos de que características importantes do conjunto de dados sejam reportadas adequadamente pelo mesmo. Cabe ressaltar que mesmo um modelo sendo apontado como modelo final para uma aplicação, deve-se avaliar sua adequabilidade, a fim de estar ciente das suas limitações antes de se fazer quaisquer inferência. O objetivo não é determinar se um modelo é perfeito, porque um modelo nunca é perfeito; em vez disso, o interesse está em determinar se as limitações do modelo possuem um efeito perceptível sobre as análises estatísticas.

De acordo com Gelman *et al.* (2014), na estatística bayesiana, um pesquisador pode verificar o ajuste do modelo usando uma variedade de estratégias: (1) verificando se as inferências *a posteriori* são razoáveis, dado o contexto do modelo; (2) examinando a sensibilidade das inferências a mudanças razoáveis na distribuição *a priori* e na verossimilhança; e (3) verificando se o modelo pode explicar os dados ou, em outras palavras, se o modelo é capaz de gerar dados como os dados observados. Para conferir o ajuste do modelo 2PLMG aos dados do EGRA faremos uso dessas abordagens.

Em relação à análise de robustez Bayesiana que corresponde à estratégia (2) é importante salientar que há três possíveis abordagens para analisar a robustez de um modelo. A primeira é uma abordagem informal, em que algumas distribuições *a priori* são consideradas e os ajustes dos correspondentes modelos obtidos são comparados. Essa abordagem é bastante utilizada devido à sua simplicidade. A segunda abordagem corresponde a conhecida por análise de robustez global (ver Moreno (2000) para uma ilustração completa). Nessa abordagem o pesquisador considera a classe de todas as distribuições *a priori* compatíveis com o que o conhecimento prévio diz e em seguida calcula os possíveis valores da média *a posteriori* (obtendo um intervalo) quando as distribuições *a priori* variam ao longo da classe. Este intervalo é normalmente encontrado utilizando-se as distribuições *a priori* “extremas” na classe que deram as médias *a posteriori* mínima e máxima. Tais cálculos podem ser difíceis em problemas multidimensionais. A terceira abordagem, chamada robustez local, é descrita por Gustafson (2000) e Sivaganesan (2000). Essa última abordagem está interessada na taxa de variação em inferências, no que diz respeito às mudanças na distribuição *a priori*, e usa técnicas diferenciais para avaliar a taxa. Medidas de sensibilidade locais são normalmente mais fáceis de serem obtidas em situações complicadas do que as medidas globais, mas a sua interpretação nem sempre é clara. Para maiores detalhes e exemplos dos três métodos de análise de robustez Bayesiana bem como suas diversas variações ver Insua e Ruggeri (2012).

Nesse trabalho fazemos uso da primeira, ou seja, a abordagem informal. Como para o modelo 3PL já há estudos de análise de sensibilidade consideramos então as distribuições *a priori* para os parâmetros dos itens obtidas nesses estudos que são: $a_i \sim N(1, 0.5)I(0,)$ e $b_i^* \sim N(0, 2)$ conforme Sahu (2002) e Patz e Junker (1999a). Sendo assim focaremos na análise de sensibilidade dos traços latentes adicionais do modelo 2PLMG, que correspondem, respectivamente, a η_j e λ_j .

Considerando inicialmente o modelo com distribuições *a priori* independentes especificamos na Tabela 5.1 alguns pares de distribuições *a priori* que analisamos para η_j e λ_j . No caso denominado M_1 , primeiro que analisamos, especificamos que $\eta_j \sim \text{Beta}(1, 1)$ e $\lambda_j \sim \text{LN}(0, 1)$ assim como em Bazan *et al.* (2010). Em M_2 propomos que $\eta_j \sim \text{Beta}(1, 1)$, porém $\lambda_j \sim \text{LN}(1, 1)$ que conforme já comentado anteriormente corresponde a uma distribuição *a priori* menos informativa para os λ_j que a anterior. Diante dos valores estimados em ambos os casos para os λ_j que pertencem, respectivamente, aos intervalos (0, 18; 23, 5) e (0, 3; 31, 3), resolvemos flexibilizar um pouco mais a distribuição *a priori* (ampliando o conjunto de valores *a priori* para λ_j) e propomos então o caso M_3 no qual $\eta_j \sim \text{Beta}(1, 1)$ e $\lambda_j \sim U(0; 50)$. No caso M_3 acabamos não constatando a convergência das cadeias. Observamos também que a inserção de uma distribuição *a priori* uniforme acabou nos levando a algumas distribuições *a posteriori* platicúrtica para alguns λ_j . No intuito de analisar a sensibilidade do modelo com relação aos η_j propomos os casos M_4 , M_5 e M_6 . Isso foi feito de maneira similar aos três primeiros casos, porém desta vez consideramos que η_j possui uma distribuição *a priori* $\text{Beta}(0, 5; 0, 5)$ a qual, também, corresponde a uma distribuição *a priori* não informativa.

Para esses três últimos casos acabamos não obtendo a convergência de todas as cadeias (para um número relativamente alto de cadeias) mesmo considerando 100000 iterações. Como o tempo computacional é bastante custoso e não obtivemos a convergência de todas as cadeias, optamos por não nos aprofundar nos casos M_4 , M_5 e M_6 . Como conseguimos obter a convergência apenas para os casos M_1 e M_2 , considerando 64000 iterações, seguimos com os mesmos. Cabe ressaltar que chegamos a implementar o caso com dependência entre as distribuições *a priori* e, mesmo deixando o algoritmo rodar até atingirmos um número de iterações correspondente ao dobro comparado ao caso sob independência. Obtivemos a convergência de todas as cadeias somente para os casos M_1 e M_2 . Como as estimativas dos parâmetros foram similares ao caso com a suposição de independência similar ao que ocorreu no estudo de simulação, decidimos então trabalhar apenas com o modelo considerando independência.

Modelo	Priori para η_j	Priori para λ_j
M_1	$Beta(1, 1)$	$LN(0, 1)$
M_2	$Beta(1, 1)$	$LN(1, 1)$
M_3	$Beta(1, 1)$	$U(0, 50)$
M_4	$Beta(0.5, 0.5)$	$LN(0, 1)$
M_5	$Beta(0.5, 0.5)$	$LN(1, 1)$
M_6	$Beta(0.5, 0.5)$	$U(0, 50)$

Tabela 5.1: Tabela contendo propostas de distribuições *a priori* para os traços latentes η_j e λ_j .

Na Figura 5.2 temos os gráficos das tolerâncias ao *speededness*, η_j 's, contra as propensões ao *speededness*, λ_j 's, correspondentes aos casos M_1 e M_2 . Através das mesmas podemos notar que as estimativas dos λ_j 's são sensíveis à escolha da distribuição *a priori*. Observamos que o caso M_1 praticamente não consegue distinguir a taxa do *speededness* para indivíduos que possuem propensão ao *speededness* a partir do item 20, enquanto o caso M_2 por permitir uma maior variabilidade nas estimativas dos λ_j consegue distinguir melhor essas taxas de aceleração para diferentes padrões de respostas.

Para especificarmos qual conjunto de distribuições é preferível calculamos então o DIC, $\overline{D}$ e o PsBF, medidas apresentadas em B.2, e exibimos os seus respectivos valores na Tabela 5.2. De acordo com todas essas medidas as distribuições *a priori* especificadas em M_2 são preferíveis, ou seja, $\eta_j \sim Beta(1, 1)$ e $\lambda_j \sim LN(1, 1)$, uma vez que apresentaram menor DIC e $\overline{D}$ e o maior PsBF.

Figura 5.2: Gráfico das estimativas das tolerâncias ao speededness, η , vs a propensão ao speededness, λ , para as distribuições *a priori* em M_1 e M_2 .

Tabela 5.2: Medidas de seleção de modelos

Casos	$\bar{D}$	DIC	PsBF (Soma dos logaritmos)
M_1	7913	8465	-4246
M_2	7681	8217	-4173

5.3 Ajuste do Modelo

Inicialmente fizemos uma análise dos dados sob três modelos para saber qual proporcionava o melhor ajuste, foram eles: (I) modelo 2PL; (II) modelo 2PLMG com independência entre as distribuições *a priori* e; (III) modelo 2PLMG com dependência entre as distribuições *a priori*.

Todos os modelos foram analisados sob as perspectivas clássica e Bayesiana. O modelo 2PLMG foi estudado pelo ponto de vista da abordagem clássica, conforme Goegebeur *et al.* (2008). A abordagem totalmente Bayesiana para o caso com independência entre as distribuições *a priori* foi feita conforme Bazan *et al.* (2010), enquanto que sob a condição de dependência *a priori* utilizamos a teoria estudada no Capítulo 3, isto é, fizemos uso de uma cópula normal para a distribuição conjunta dos traços latentes, para isso consideramos que marginalmente as mesmas possuíam as distribuições *a priori* especificadas no caso M_2 apresentado na análise de sensibilidade.

Estimação Usando a Abordagem Clássica

A seguir, fazemos um breve resumo dos programas/pacotes utilizados para ajustar os respectivos modelos ao conjunto de dados do NWF.

Inicialmente temos o *software* R. Project em que no mesmo utilizamos dois pacotes distintos para ajustar o modelo 2PL os quais apresentamos à seguir.

1. Pacote *ltm* (*Latent Trait Model*):

Desenvolvido por Rizopoulos (2013) o pacote *ltm* é utilizado na análise multivariada de dados dicotômicos e politômicos usando modelos com traço latente sob a abordagem da TRI. Ele

inclui o modelo de Rasch, os modelos logísticos de Birnbaum com dois e três parâmetros, o de resposta gradual, e o modelo de crédito parcial generalizado. Nesse pacote a estimação dos parâmetros dos itens é obtida através do método da máxima verossimilhança marginal por meio da função *tpm*. Nessa função a verossimilhança é integrada numericamente sob o traço latente dos respondentes por meio do método de Gauss-Hermite (o *default* considera 21 pontos de quadratura) e posteriormente a maximização nos parâmetros dos itens é realizada pelo método de Quasi-Newton por meio da subrotina BFGS com um *default* de 150 iterações. Por fim, no referido pacote as estimativas das habilidades são obtidas através do uso da função *factor.score* através da média *a posteriori*.

2. Pacote *irtoys* (*Simple interface to the estimation and plotting of IRT models*).

O pacote *irtoys* de Partchev (2014) utiliza a função *est* para estimar os parâmetros dos itens, essa função possui características similares as do software BILOG com relação às opções de estimação. Inclusive, é possível utilizar as ferramentas de estimação do BILOG caso o tenha instalado em seu computador. A seguir usamos a função *est* para estimar os parâmetros dos itens fazendo uso da *engine ltm*. Nessa função consideramos 20 pontos de quadratura para a integração uma vez que estamos considerando o modelo mais simples. As estimativas das habilidades utilizando esse pacote são obtidas via média *a posteriori* utilizando a função *eap*.

Um segundo *software* levado em consideração foi o SAS. Utilizamos o mesmo para estimar os parâmetros dos modelos 2PL e 2PLMG. Para o último seguimos o mesmo procedimento adotado em Goegebeur *et al.* (2008). Nesse *software* utilizamos uma *procedure* chamada de NLMIXED. É importante salientar inicialmente que para utilizar essa *procedure* o usuário deve estar atento na forma como o conjunto de dados está disposto. Para mais detalhes de como estruturar o conjunto de dados para a utilização dessa *procedure* veja Sheu *et al.* (2005). Para o modelo 2PLMG a estimação do vetor de parâmetros ξ , em (2.8), nessa *procedure* é feita via MVM sendo as integrais calculadas através do método da quadratura gaussiana e a maximização feita posteriormente através do algoritmo de Newton-Raphson. Para utilização do método da quadratura gaussiana deve-se especificar inicialmente a quantidade de pontos quadratura a ser utilizada e seguindo Goegebeur *et al.* (2008) especificamos somente 5 pontos de quadratura. Sob essas condições o procedimento de estimação levou aproximadamente 2 dias para ser concluído. Porém, de acordo com Sheu *et al.* (2005) o uso de 5 pontos de quadratura é considerado insuficiente, uma vez que os mesmos recomendam o uso de pelo menos 20 pontos de quadratura. Dessa forma buscamos aumentar a quantidade de pontos de quadratura para 15 e após 15 dias a convergência não havia sido alcançada inviabilizando o uso desse método com o SAS. Por outro lado, considerando 20 pontos de quadratura para o modelo 2PL a convergência foi alcançada em menos de 2 horas e os resultados são apresentados adiante.

Um fato curioso que ocorreu quando analisamos o modelo 2PL foi que como todos os métodos descritos acima apontaram para convergência no processo de estimação dos parâmetros dos itens, o esperado é que tais estimativas sejam extremamente próximas, uma vez que o modelo em estudo é o mesmo, o que está sendo alterado é apenas o método de estimação. Portanto, ao fazermos um gráfico das estimativas obtidas em um método contra as obtidas com outro método as mesmas devem estar sobre a reta da função identidade. Porém, conforme podemos ver na Figura 5.3 isso não ocorre, indicando que o uso da abordagem clássica pode não ser uma boa opção para estimar os parâmetros do modelo 2PL quando utilizado o conjunto de dados do EGRA. Isso pode estar ocorrendo, possivelmente, por dois motivos. O primeiro seria o efeito do *speededness* enquanto o segundo consiste do fato de termos 12 indivíduos com escore nulo e 39 indivíduos com escore perfeito totalizando praticamente 10% dos respondentes do teste.

Figura 5.3: Comparação das estimativas dos parâmetros de discriminação, a_i , do Modelo 2PL e dos parâmetros de intercepto, b_i^* , estimados com diferentes métodos/programas.

Estimação Usando a Abordagem Totalmente Bayesiana

Realizamos então a estimação dos parâmetros com a abordagem totalmente Bayesiana para cada um dos três modelos: modelo 2PL; modelo 2PLMG sob independência entre as distribuições *a priori*; e o modelo 2PLMG com a suposição de dependência *a priori*. Cabe ressaltar que para

os dois primeiros modelos é possível utilizar o pacote *rjags* disponível no R. Project ou o *software* WinBUGS. Neste último *software* os programas utilizados para estimação dos parâmetros pode ser encontrado em [Bazan et al. \(2010\)](#). Por questões de comparação de desempenho do tempo computacional realizamos a estimação dos parâmetros nos três casos utilizando o *software* Matlab já que o mesmo oferecia comandos que facilitavam a implementação do caso com dependência entre as distribuições *a priori* otimizando o tempo computacional. Para a estimação dos parâmetros desses modelos consideramos em (2.6) a reparametrização $a_i\theta_j - b_i^*$ já mencionada anteriormente. As especificações das distribuições *a priori* foram feitas com base no estudo de análise de sensibilidade apresentado na seção anterior em que optamos por seguir: $\theta_j \sim N(0, 1)$, $\eta_j \sim \text{Beta}(1, 1)$, $\lambda_j \sim \text{LN}(1, 1)$, $a_i \sim N(1, 0.5)I(0,)$ e $b_i^* \sim N(0, 2)$, onde $b_i = b_i^*/a_i$ é o parâmetro de dificuldade e b_i^* é o parâmetro de intercepto de $a_i\theta_j - b_i^*$.

Tanto para o modelo 2PL quanto para o modelo 2PLMG com distribuições *a priori* independentes consideramos um *burn-in* de 24.000 com saltos de tamanho 20, sendo o tamanho dos saltos suficiente para eliminar as autocorrelações (os gráficos foram omitidos devido a sua grande quantidade). Assim, após 64.000 iterações foram obtidas amostras de tamanho 2.000 para cada um dos parâmetros dos respectivos modelos citados. Para analisar a convergência das cadeias foi utilizado o pacote *coda* ([Plummer et al., 2006](#)) no qual os seguintes métodos foram analisados: método de Geweke ([Geweke et al., 1991](#)) e o método de Gelman e Rubin ([Gelman e Rubin, 1992](#)). Para o método de Gelman e Rubin obtivemos a convergência das cadeias para todos os parâmetros sendo que nesse método foram consideradas duas cadeias paralelas.

Na Figura 5.4 temos um comparativo das estimativas obtidas utilizando a abordagem totalmente Bayesiana e as estimativas obtidas utilizando as abordagens clássicas mencionadas anteriormente para o modelo 2PL. Podemos notar claramente que as estimativas de alguns itens usando abordagem clássica diferem das estimativas usando a abordagem totalmente Bayesiana. Devido a inconsistência entre as estimativas utilizando métodos clássicos fizemos uso somente da abordagem totalmente Bayesiana.

Figura 5.4: Comparação das estimativas dos parâmetros de discriminação dos itens, a_i , e dos parâmetros de intercepto, b_i^* , do Modelo 2PL obtidas utilizando a abordagem totalmente Bayesiana e a abordagem Clássica com diferentes programas/pacotes.

Para o caso com dependência entre as distribuições *a priori* utilizamos a metodologia apresentada na Subseção (3.3). Nesse caso, consideramos um *burn-in* de tamanho 60000 e saltos de tamanho 20. Assim, após 100000 iterações obtivemos uma amostra de tamanho 2000. A análise da convergência foi feita de maneira similar ao caso do modelo 2PLMG com distribuições *a priori* independentes sendo que o método de Gelman e Rubin novamente indicou a convergência das cadeias para todos os parâmetros. Conforme ocorreu no estudo de simulação as estimativas obtidas no modelo considerando dependência entre as distribuições *a priori* foram praticamente idênticas as obtidas considerando que as distribuições *a priori* eram independentes, dessa forma, seguiremos apenas comentando os resultados com distribuições que são independentes *a priori*.

Seguindo então usando apenas a abordagem Bayesiana, estimamos os parâmetros dos modelos 2PL e 2PLMG sem dependência entre as distribuições *a priori* (que, como vimos, oferece resultados idênticos ao com dependência) conforme mencionamos anteriormente. De acordo com a Figura 5.5 observamos que em conformidade com o estudo de simulação as estimativas dos parâmetros dos itens ao utilizar o modelo 2PL na presença de *speededness* acabam sendo superestimadas. Além disso, conforme a Figura 5.10 observamos que as habilidades de alguns indivíduos acabam sendo subestimadas. Nas Figuras de 5.6 à 5.9 apresentamos as estimativas dos parâmetros dos itens (mediana da distribuição *a posteriori*) dos modelos 2PL e 2PLMG conjuntamente com seus respectivos intervalos de máxima densidade *a posteriori*, HPD, com 95% de confiança e podemos notar que, as distribuições *a priori* utilizadas foram flexíveis permitindo às estimativas assumirem um conjunto de valores bastante diversificado.

Figura 5.5: Comparação entre as estimativas dos parâmetros de discriminação, a_i , e de posição, b_i^* , dos modelo 2PL com o modelo 2PLMG com distribuições *a priori* independentes.

Figura 5.6: Intervalo com 95% de credibilidade (HPD) do parâmetro de discriminação, a_i , e gráfico da distribuição a priori dos respectivos parâmetros para o modelo 2PL com independência a priori.

Figura 5.7: Intervalo com 95% credibilidade (HPD) do parâmetro de discriminação, a_i , e gráfico da distribuição a priori dos respectivos parâmetros para o modelo 2PLMG com independência a priori.

Figura 5.8: Intervalo com 95% de credibilidade (HPD) do parâmetro de intercepto, b_i^* , e gráfico da distribuição a priori dos respectivos parâmetros para o modelo 2PL.

Figura 5.9: Intervalo com 95% de credibilidade (HPD) do parâmetro de intercepto, b_i^* , e gráfico da distribuição *a priori* dos respectivos parâmetro para o modelo 2PLMG com independência *a priori*.

A Tabela 5.3 resume os resultados das estimativas das variáveis latentes dos modelos 2PL e 2PLMG com independência entre as distribuições *a priori*. A partir da Tabela 5.3 e das Figuras de 5.10 e 5.11 notamos que as estimativas das habilidades, θ_j e das tolerâncias ao *speededness*, η_j , são positivamente correlacionadas com os escores. Em resumo, indivíduos que possuem uma habilidade alta ou uma tolerância maior à aceleração tendem a possuírem um escore maior. Observando a correlação entre os escores e a taxa de aceleração notamos que a mesma é negativa indicando que indivíduos que possuem uma taxa de aceleração baixa possuem maior escore.

Tabela 5.3: Escores observados e medidas resumo para as estimativas Bayesianas das variáveis latentes dos indivíduos sob os modelos 2PL e 2PLMG com independência *a priori*

parâmetro	Mediana	Mínimo	Máximo	SD	θ 2PL	θ 2PLMG	η	λ	score
θ 2PL	-0.010	-3.843	2.356	1.458	1				
θ 2PLMG	-0.150	-2.148	1.956	1.007	0.615	1			
η	0.667	0.006	0.937	0.265	0.950	0.490	1		
λ	3.374	0.331	31.23	6.107	-0.777	-0.177	-0.723	1	
score	30.701	0.000	50.000	12.976	0.996	0.622	0.955	-0.757	1

Figura 5.10: Matriz de correlação das habilidades, θ_j , estimadas com os modelos 2PL e 2PLMG com independência entre as distribuições a priori e o escore.

Figura 5.11: Correlações entre as estimativas das variáveis latentes (habilidades, θ_j , tolerância ao speededness, η_j e propensão ao speededness, λ_j) do modelo 2PLMG com independência a priori.

Tabela 5.4: Tabela comparativa das estimativas dos traços latentes dos modelos 2PL e 2PLMG com o escore

Respostas dicotomizadas	escore	θ 2PL	θ 2PLMG	η	λ
11111111111101111111101110111111101111111111110000	42	1.158	1.023	0.489	0.663
11111111111111111110000000000000000000000000000000	18	-1.492	0.245	0.325	15.827
101000	2	-3.352	-0.121	0.017	18.542
00	0	-3.818	-0.100	0.006	30.646

De acordo com a Tabela 5.4 podemos observar que quanto mais cedo ocorre o efeito do *speededness* mais significativa é a diferença entre as estimativas dos θ 's, ou seja, de uma forma geral estimativas mais discrepantes dos valores de θ quando comparado ambos os modelos estão associados a um baixo escore e a uma alta propensão ao *speededness*.

Na Tabela 5.5 apresentamos algumas medidas para comparação dos modelos, são elas, $\overline{D}$, DIC e preditiva a posteriori média. Segundo todas essas medidas o modelo preferível deve ser o modelo 2PLMG com independência entre as distribuições *a priori*, como era esperado.

Dados simulados	Nº de Parâmetros	$\overline{D}$	DIC	PsBF (soma dos log)	Qtd de iterações por hora
2PL	612	9050	10048	-4504	≈ 1600
2PLMG	1636	7681	8217	-4173	≈ 1440

Tabela 5.5: Comparação dos modelos 2PL e 2PLMG para os dados do EGRA

É importante salientar que a quantidade de iterações por hora apresentada pode variar conforme o tempo de simulação (necessidade de cadeias mais longas) uma vez que há maior exigência de memória ram do computador. Os valores apresentados corresponde apenas a quantidade de iterações obtidas durante a primeira hora em cada programa.

5.4 Considerações Finais

Nessa seção apresentamos as principais contribuições e as conclusões do estudo realizado nesse trabalho bem como algumas possibilidades de trabalhos futuros.

Principais Contribuições

Sobre o presente trabalho ressaltamos então como principais contribuições às seguintes:

- i) Enriquecimento da literatura de modelos, da TRI, que levam em consideração o efeito de *speededness*.
- ii) Códigos para estimação dos parâmetros do modelo de resposta ao item com processo de mudança gradual implementados, em código aberto, no MATLAB.
- iii) A realização de um estudo de análise de sensibilidade do modelo 2PLMG para os dados do EGRA.
- iv) Estudo de simulação comparando o modelo de resposta ao item com processo de mudança de resposta gradual que considera independência entre as distribuições *a priori* para os traços latentes com o que considera dependência entre as distribuições *a priori*.

Conclusões

De acordo com o exposto neste trabalho ficou constatado que:

O uso de modelos que não levam em consideração o efeito de *speededness*, quando o mesmo é constatado, podem nos levar a interpretações errôneas, ocasionando a superestimação dos parâmetros dos itens e a subestimação das habilidades dos indivíduos.

O uso do *software* SAS utilizando a *procedure* NLMIXED, para estimar os parâmetros dos modelos 2PLMG, não é viável devido a enorme exigência computacional para estimar tais parâmetros. Isso ocorre, porque uma quantidade de pontos de quadratura inferior a 20 não é suficiente para estimar os parâmetros dos itens com uma precisão considerada razoável.

Por fim, a conclusão mais importante do referido trabalho é que o uso do modelo 2PLMG com a suposição de independência entre as distribuições *a priori* leva as mesmas interpretações obtidas com o modelo que supõe dependência entre as distribuições *a priori* quando analisado os dados do EGRA. Portanto, o uso do modelo 2PLMG considerando distribuições *a priori* independentes para os traços latentes deve ser preferível, pois além de ser mais fácil de implementar o mesmo apresenta outras vantagens dentre as quais podemos destacar uma convergência mais rápida para as cadeias dentro do algoritmo de Gibbs e o menor aparato teórico. É importante ressaltar a possibilidade de

que para algum outro banco de dados haja diferenças significativas nas estimativas dos parâmetros dos dois modelos em questão.

Trabalhos Futuros

Concluimos então esse trabalho deixando algumas propostas de trabalhos futuros.

- Construir um estudo de análise de sensibilidade formal para o modelo 2PLMG. Essa análise de sensibilidade poderia ser global ou local. Conforme mencionado, anteriormente, para uma análise global recomenda-se a leitura de [Moreno \(2000\)](#), enquanto, para uma análise local recomenda-se a leitura de [Gustafson \(2000\)](#) e [Sivaganesan \(2000\)](#).
- Realizar uma segunda aplicação, em outro conjunto de dados, a qual apresente diferenças significativas nas estimativas dos traços latentes quando utilizado os modelos sem/com dependência entre os traços latentes.
- Estudar outras funções de ligações no modelo de resposta ao item com processo de mudança de resposta gradual.
- Fazer um estudo inserindo assimetria nos traços latentes do modelo de resposta ao item com processo de mudança de resposta gradual.
- Atualmente, devido aos avanços computacionais os tempos de respostas são facilmente coletados através de testes adaptativos ou questionários respondidos no computador. Como os tempos de resposta correspondem a uma informação valiosa sobre os itens do teste e os respondentes, pode-se realizar um estudo de modelagem conjunta das respostas e dos tempos respostas dos respondentes. Assim, outra proposta seria estudar os modelos usuais da TRI e os tempos de respostas (TRs) dos indivíduos, conjuntamente, para estimar o(s) traço(s) latente(s) e assim pensar em propor um novo modelo para os tempos de resposta e por fim fazer um comparativo entre os mesmos para ver qual apresenta o melhor ajuste. Um ponto de partida inicial seria focar no modelo log-normal para os tempos de respostas proposto em [Van der Linden \(2006\)](#), para que posteriormente se possa fazer uma modelagem conjunta seguindo [Fox \(2010\)](#). Assim poderia se

Apêndice A

Algumas Distribuições

Nesse Apêndice apresentamos três distribuições de probabilidade utilizadas no texto e que geralmente são poucos abordadas.

A.1 Distribuição Categórica

A distribuição Categórica consiste de uma generalização da distribuição de Bernoulli em que, ao invés de dois possíveis resultados, podemos ter k possíveis resultados associados a uma variável aleatória, X , num experimento. Dessa forma considere o conjunto $\{1, 2, \dots, k\}$ composto pelos k possíveis resultados de um experimento e denote $p_k = \mathbb{P}(X = k)$. Dizemos então que X tem função de probabilidade categórica se,

$$f(x) = \prod_{i=1}^k p_i^{[x=i]} \quad \text{e} \quad \sum_{i=1}^k p_i = 1,$$

em que $[x = i]$ corresponde a 1, se $x = i$ e 0 caso contrário.

Notação: $X \sim \text{Cat}(p_1, \dots, p_k)$.

A.2 Distribuição de Dirichlet

A distribuição de Dirichlet corresponde a uma generalização multidimensional da distribuição Beta e é usada na estatística Bayesiana por ser a distribuição conjugada da distribuição multinomial.

Definição A.1. Dizemos que um vetor aleatório, n -dimensional X , possui distribuição de Dirichlet se sua função densidade de probabilidade é proporcional a

$$\prod_{i=1}^n x_i^{\alpha_i - 1}, \tag{A.1}$$

em que, $x_i \geq 0$, $\sum_{i=1}^n x_i = 1$ e $\alpha_i > 0$.

A.3 Distribuições Wishart e Wishart Inversa

Proposta em [Wishart \(1928\)](#) a distribuição Wishart consiste de uma generalização multidimensional da distribuição gama. Essa distribuição de probabilidade é bastante utilizada em estatística Bayesiana, pois corresponde a uma família de distribuições que é conjugada da distribuição normal multivariada quando a matriz de precisão é desconhecida, de acordo com [Kadane \(2011\)](#).

Definição A.2. A distribuição Wishart, corresponde a uma distribuição sobre o espaço de $p(p+1)/2$ elementos livres de uma matriz simétrica e positiva definida X . Dizemos então que uma matriz $\mathbf{X}$, de ordem p , simétrica e positiva definida, de variáveis aleatórias possui uma distribuição Wishart

com n graus de liberdade e matriz de escala $\mathbf{S}_{p \times p}$ simétrica e positiva definida, se sua densidade é proporcional a

$$|\mathbf{X}|^{(n-p-1)/2} e^{-\frac{1}{2} \text{tr}(\mathbf{S}\mathbf{X})} \quad (\text{A.2})$$

em que $n > p - 1$ e tr denota o que chamamos de traço de uma matriz, ou seja, a soma dos elementos da diagonal principal.

Notação: $\mathbf{X} \sim W(n, \mathbf{S})$

Caso o interesse seja utilizar uma distribuição conjugada a matriz de variância do modelo no caso em que a mesma desconhecida, pode-se utilizar a distribuição Wishart Inversa, cuja definição é dada à seguir

Definição A.3. Dizemos então que uma matriz $\mathbf{Y}$, de ordem p , simétrica e positiva definida, de variáveis aleatórias possui uma distribuição Wishart Inversa com ν graus de liberdade e matriz de escala $\mathbf{S}^*_{p \times p}$ simétrica e positiva definida, se sua densidade é proporcional a

$$|\mathbf{Y}|^{-(\nu+p+1)/2} e^{-\frac{1}{2} \text{tr}(\mathbf{S}^*\mathbf{Y})} \quad (\text{A.3})$$

em que $n > p + 1$

Notação: $\mathbf{X} \sim WI(\nu, \mathbf{S}^*)$

Essas duas distribuições estão relacionadas pela seguinte proposição.

Proposição 6. $\mathbf{X} \sim W(n, \mathbf{S}) \Leftrightarrow \mathbf{Y} = \mathbf{X}^{-1} \sim WI(n, \mathbf{S}^{-1})$.

Apêndice B

Comparação de Modelos

Nesse Apêndice apresentamos algumas medidas de comparação de modelos as quais fizemos usos no decorrer do texto, são elas: Pseudo fator de Bayes $\overline{D}$ e DIC.

B.1 Pseudo Fator de Bayes

No decorrer dessa seção usaremos as seguintes notações:

- $y_{ij.obs}$: Resposta observada do indivíduo j ao item i
- $y_{i.obs}$: Resposta observada do item i atribuída por algum indivíduo
- $\mathbf{y}_{(-i)j}$: Vetor de respostas do indivíduo j sem ao item i
- $\mathbf{y}_{(-i)}$: Vetor de respostas de um indivíduo ao item i
- y_i : possível resposta atribuída ao item i por um indivíduo.

Conforme [Paulino *et al.* \(2003\)](#) ao substituírmos $p(\mathbf{y})$ pelo produto $\prod_{i=1}^n p(y_i|\mathbf{y}_{(-i)})$ no cálculo do fator de Bayes para comparar modelos, damos origem ao que é chamado de pseudo fator de Bayes denotado por PsBF (veja [Geisser e Eddy \(1979\)](#) e [Gelfand *et al.* \(1992\)](#))

$$PsBF = \prod_{j=1}^J \prod_{i=1}^I \frac{p(y_{ij.obs}|\mathbf{y}_{(-i)j}, M_1)}{p(y_{ij.obs}|\mathbf{y}_{(-i)j}, M_2)} \equiv \prod_{j=1}^J \prod_{i=1}^I C_{ij}.$$

Definido desta forma temos que se este fator for maior do que 1, prefere-se M_1 a M_2 , uma vez que o mesmo está dizendo que a posteriori há mais evidências em favor do modelo M_1 , caso contrário prefere-se M_2 . Consequentemente, as observações para as quais o termo $\prod_{i=1}^I C_{ij}$ é maior que 1 correspondem a observações que dão preferência ao modelo M_1 e observações para as quais o $\prod_{i=1}^I C_{ij}$ correspondente é menor que 1 dão preferência ao modelo M_2 .

No caso de haver vários modelos em competição pode ser calculado, para cada modelo,

$$\prod_{j=1}^J \prod_{i=1}^I p(y_{ij.obs}|\mathbf{y}_{(-i)j})$$

ou, equivalentemente a soma dos logaritmos das ordenadas preditivas condicionais, e escolher o modelo que apresente o maior valor. Esta última opção corresponde a que faremos uso.

Note que nosso problema reduz-se então a como calcular $p(y_{i.obs}|\mathbf{y}_{(-i)})$ para cada respondente j dado um modelo.

Considerando $\boldsymbol{\xi} = (\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda})$ note que

$$\begin{aligned}
p(y_{i.obs}|\mathbf{y}_{(-i)}) &= \int f(y_i, \boldsymbol{\xi}|\mathbf{y}_{(-i)})d\boldsymbol{\xi} \\
&= \int f(y_i|\boldsymbol{\xi}, \mathbf{y}_{(-i)})h(\boldsymbol{\xi}|\mathbf{y}_{(-i)})d\boldsymbol{\xi}
\end{aligned}$$

dessa forma uma maneira de obtermos uma estimativa para $p(y_{i.obs}|\mathbf{y}_{(-i)})$ seria simularmos uma amostra $\{\boldsymbol{\xi}_{(l)}^\bullet; l = 1, \dots, m\}$ para cada distribuição *a posteriori* $h(\boldsymbol{\xi}|\mathbf{y}_{(-i)})$ e aproximarmos $p(y_{i.obs}|\mathbf{y}_{(-i)})$ por

$$\sum_{l=1}^m \frac{f(y_i|\boldsymbol{\xi}_{(l)}^\bullet, \mathbf{y}_{(-i)})h(\boldsymbol{\xi}_{(l)}^\bullet|\mathbf{y}_{(-i)})}{m}.$$

Porém, isso não é prático em nosso caso pois $\mathbf{y}$ tem dimensão elevada. Uma maneira de se contornar esse problema foi proposta em [Gelfand \(1996\)](#) e consiste em resolver a integral acima as custas simplesmente da amostra $\{\boldsymbol{\xi}_{(l)}^*; l = 1, \dots, m\}$ da distribuição *a posteriori* de $\boldsymbol{\xi}$ cuja densidade é denotada por $h(\boldsymbol{\xi}|\mathbf{y})$.

A sugestão de [Gelfand \(1996\)](#) consiste em estimar $p(y_i|\mathbf{y}_{(-i)})$ utilizando a média harmônica de $\{f(y_i|\mathbf{y}_{(-i)}, \boldsymbol{\xi}_{(l)}^*), l = 1, \dots, m\}$. Para isso, o mesmo notou que

$$p(\mathbf{y})h(\boldsymbol{\xi}|\mathbf{y}) = h(\boldsymbol{\xi})f(\mathbf{y}|\boldsymbol{\xi}) = h(\boldsymbol{\xi})f(y_i|\mathbf{y}_{(-i)}, \boldsymbol{\xi})f(\mathbf{y}_{(-i)}|\boldsymbol{\xi}).$$

Assim,

$$\begin{aligned}
p(y_i|\mathbf{y}_{(-i)}) &= \frac{p(\mathbf{y})}{p(\mathbf{y}_{(-i)})} \\
&= \frac{1}{\int \frac{f(\mathbf{y}_{(-i)}|\boldsymbol{\xi})h(\boldsymbol{\xi})h(\boldsymbol{\xi}|\mathbf{y})}{p(\mathbf{y})h(\boldsymbol{\xi}|\mathbf{y})}d\boldsymbol{\xi}} \\
&= \frac{1}{\int \frac{1}{f(y_i|\mathbf{y}_{(-i)}, \boldsymbol{\xi})}h(\boldsymbol{\xi}|\mathbf{y})d\boldsymbol{\xi}}
\end{aligned}$$

e portanto, se $\{\boldsymbol{\xi}_{(l)}^*; l = 1, \dots, m\}$ é uma amostra de $h(\boldsymbol{\xi}|\mathbf{y})$, tem-se

$$\hat{p}(y_i|\mathbf{y}_{(-i)}) \approx \frac{1}{\frac{1}{m} \sum_{l=1}^m \frac{1}{f(y_i|\mathbf{y}_{(-i)}, \boldsymbol{\xi}_{(l)}^*)}}$$

expressão esta que pode ser simplificada no caso em que $(y_1, \dots, y_n)$ são condicionalmente independentes, já que nesse caso, $f(y_i|\mathbf{y}_{(-i)}, \boldsymbol{\xi}) = f(y_i|\boldsymbol{\xi})$.

B.2 $\overline{D}$ e DIC

No decorrer dessa seção apresentamos duas medidas utilizadas para fazer comparações de modelos, são elas, o $\overline{D}$ e o DIC. Essas medidas tem como base uma medida chamada de *Desvio Bayesiano*, denotada por $D(\mathbf{a}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{y})$ e definida a partir da log-verossimilhança do modelo proposto conforme [Dempster \(1973\)](#). Assim, para o modelo 2PLMG definimos os *Desvio Bayesiano* como

$$\begin{aligned}
D(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda}) &= -2 \ln(p(\mathbf{y}|\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda})) \\
&= -2 \sum_{j=1}^J \sum_{i=1}^I \ln \mathbb{P}(Y_{ij} = y_{ij} | a_i, b_i, \lambda_j, \eta_j, \theta_j),
\end{aligned}$$

em que y_{ij} denota a resposta observada 0 (incorreta) ou 1 (correta) de Y_{ij} , ou seja, da variável resposta do respondente j ao item i no modelo 2PLMG.

O Desvio bayesiano é considerado uma medida de bondade de ajuste e é utilizado para comparar modelos. Essa medida possui a vantagem de ser facilmente calculada e interpretada. De acordo com essa a mesma ao compararmos uma gama de modelos propostos para um conjunto de dados, o modelo que melhor ajusta os dados é o que apresenta o menor valor do Desvio Bayesiano. Comumente, como medida de qualidade de ajuste é utilizado $\overline{D}$, uma medida que corresponde ao desvio médio das distribuições *a posteriori* dos parâmetros cuja definição é

$$\overline{D} \equiv E_{\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda} | \mathbf{y}}(D(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda})).$$

Com o uso de técnicas MCMC como, por exemplo, o algoritmo de Gibbs podemos obter a seguinte aproximação para $\overline{D}$:

$$\overline{D} \approx \sum_{i=1}^n D(\mathbf{a}^{(i)}, \mathbf{b}^{(i)}, \boldsymbol{\theta}^{(i)}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\lambda}^{(i)}),$$

em que i , corresponde ao i -ésimo ponto amostral obtido através do algoritmo de Gibbs de uma amostra de tamanho n . Similarmente a D a interpretação de $\overline{D}$ é feita de modo que quanto menor o valor de $\overline{D}$ melhor o modelo ajusta os dados.

Outra medida bastante utilizada para comparação de modelos é o DIC (deviance information criterion) que também consiste de uma modificação no Desvio Bayesiano. O DIC que utilizaremos é o definido em Spiegelhalter *et al.* (2002) através da seguinte expressão,

$$DIC = \overline{D} + p_D$$

em que,

$$\begin{aligned} p_D &= \overline{D} - D(E_{\mathbf{a}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\lambda} | \mathbf{y}}(\mathbf{a}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\theta})) \\ &= \overline{D} - D(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}) \end{aligned}$$

corresponde ao que é chamado de fator de penalização pela complexidade do modelo e é a diferença entre o valor esperado de D e o valor de D aplicado no valor *a posteriori* dos parâmetros. Cabe ressaltar conforme Bazan (2005) que na modelagem hierárquica a verossimilhança (ou “complexidade do modelo”) não é única no sentido que possui várias versões da verossimilhança que podem ser usadas como acontece nos modelos que possuem variáveis latentes auxiliares. Portanto, o desvio Bayesiano do modelo não é único (e também o DIC e p_D que são baseados nela) e pode ser calculado de vários modos (ver Gelman *et al.* (2014)). Assim como ocorre com o $\overline{D}$ o modelo preferível é aquele que apresenta o menor valor da medida de ajuste, ou seja, o menor DIC.

Apêndice C

Prova do Teorema 2

Neste Apêndice apresentamos um breve esboço da demonstração do Teorema 2. Inicialmente, vamos enunciar um lema que será de grande importância. Posteriormente, demonstraremos um caso particular do teorema que estamos interessado, o qual enunciaremos como uma lema. Por fim, explicamos como estender o caso mais simples para o caso mais geral. No que segue para simplificarmos as contas usaremos o produto exterior também conhecido como produto cunha cuja operação será denotada por $\wedge$. Para uma introdução ao produto exterior ver [Muirhead \(1982\)](#).

Lema 3. *Seja $\mathbf{W}$ um matriz $p \times p$ positiva definida e $\mathbf{W} = \mathbf{T}'\mathbf{A}\mathbf{T}$, onde $\mathbf{T}$ é uma matriz triangular inferior cujos elementos da diagonal são todos iguais a 1 e $\mathbf{A}$ uma matriz diagonal com elementos positivos. Denote os elementos abaixo da diagonal de $\mathbf{T}$ por t_{ij} , $1 \leq j < i \leq p$, e os elementos da diagonal de $\mathbf{A}$ por t_{ii} , $i = 1, \dots, p$. Considere que $(d\mathbf{W})$ denota o diferencial da forma $(d\mathbf{W}) \equiv \wedge_{i \geq j} d\omega_{ij}$ e similarmente $(d\mathbf{T}) \equiv \wedge_{i \geq j} dt_{ij}$. Então temos*

$$(d\mathbf{W}) = \prod_{i=1}^p t_{ii}^{-i+1} (d\mathbf{T}).$$

Em outras palavras, o determinante do Jacobiano da transformação de $\mathbf{T}'\mathbf{A}\mathbf{T}$ para $\mathbf{W}$ é $\prod_{i=1}^p t_{ii}^{-i+1}$.

Demonstração. Por definição temos

$$\begin{pmatrix} \omega_{11} & \omega_{21} & \dots & \omega_{p1} \\ \omega_{21} & \omega_{22} & \dots & \omega_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \dots & \omega_{pp} \end{pmatrix} = \begin{pmatrix} 1 & t_{21} & \dots & t_{p1} \\ 0 & 1 & \dots & t_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} t_{11} & 0 & \dots & 0 \\ 0 & t_{22} & \dots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{pp} \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ t_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{p1} & t_{p2} & \dots & 1 \end{pmatrix}.$$

Agora expressamos cada ω_{ij} em termos dos t'_{ii} s e então tomamos os diferenciais. Já que vamos tomar o produto exterior dos diferenciais e o produto exterior de diferenciais iguais é zero, não é necessário exibir diferenciais em t_{ij} que já ocorreram previamente. Em geral temos

$$\begin{aligned} \omega_{ii} &= t_{ii} + \sum_{j=i+1}^p t_{ji}^2 t_{jj}, \quad i = 1, \dots, p, \\ \omega_{ij} &= t_{ij} t_{ii} + \sum_{k=i+1}^p t_{ki} t_{kj} t_{kk}, \quad 1 \leq j < i \leq p. \end{aligned}$$

Assim, tomando os diferenciais, e ignorando os que ocorreram previamente, temos

$$\begin{aligned}
d\omega_{pp} &= dt_{pp} \\
d\omega_{p,p-1} &= t_{pp}dt_{p,p-1} + \dots \\
&\vdots \\
d\omega_{p1} &= t_{pp}dt_{p1} + \dots \\
d\omega_{p-1,p-1} &= dt_{p-1,p-1} + \dots \\
&\vdots \\
d\omega_{11} &= dt_{11} + \dots
\end{aligned}$$

dessa forma tomando o produto exterior obtemos que

$$(d\mathbf{W}) \equiv \wedge_{i \geq j} d\omega_{ij} = t_{pp}^{p-1} t_{p-1,p-1}^{p-2} \dots t_{22} \wedge_{i \geq j} dt_{ij}$$

como queríamos demonstrar. ■

De posse do lema, considere as seguintes distribuições *a priori* para t_{ii} e $t_{ij}|t_{ii}$

$$t_{ii} \stackrel{ind}{\sim} IG((\nu + i - p)/2, 1/2), \quad \nu > p, \quad i = 1, 2, \dots, p \quad (\text{C.1})$$

e

$$t_{ij}|t_{ii} \stackrel{ind}{\sim} N(0, t_{ii}\mathbf{I}_{i-1}), \quad i = 1, \dots, p \quad (\text{C.2})$$

onde IG denota a gama inversa. Diante das mesmas temos o seguinte resultado que corresponde a um caso particular do teorema que estamos interessados.

Lema 4. *As distribuições a priori em (C.1) e (C.2) induzem uma distribuição Wishart para a matriz $\Sigma^{-1} \equiv \mathbf{T}'\mathbf{A}^{-1}\mathbf{T}$ tem uma distribuição Wishart, $\Sigma^{-1} \sim \mathcal{W}(\nu, \mathbf{I}_p)$. Além disso, o inverso também é verdadeiro, se $\Sigma^{-1} \sim \mathcal{W}(\nu, \mathbf{I}_p)$, então as distribuições a priori induzidas em t_{ii} , $i = 1, \dots, p$ e t_{ij} , $1 \leq j < i \leq p$ são as apresentadas em (C.1) e (C.2), respectivamente.*

Demonstração. Assumindo a mesma notação do lema anterior, para provar o teorema, é suficiente considerar o caso onde $t_{ii} \stackrel{ind}{\sim} g(\frac{\nu+i-p}{2}, 2)$, $\nu > p$, $i = 1, \dots, p$, e para cada i , temos $t_{ij}|t_{ii} \stackrel{ind}{\sim} N(0, t_{ii}^{-1})$, $1 \leq j < i \leq p$. Primeiro note que como $\det(\mathbf{T}) = 1$, temos

$$\det(\mathbf{W}) = \det(\mathbf{A}) = \prod_{i=1}^p t_{ii}.$$

Além disso, por (1) temos

$$\begin{aligned}
tr(\mathbf{W}) &= \sum_{i=1}^p \omega_{ii} \\
&= \sum_{i=1}^p t_{ii} + \sum_{i=1}^p \sum_{j=1+1}^p t_{ji}^2 t_{jj} \\
&= \sum_{i=1}^p t_{ii} + \sum_{j=2}^p \sum_{i=1}^{j-1} t_{ji}^2 t_{jj} \\
&= \sum_{i=1}^p t_{ii} + \sum_{j=2}^p \sum_{i=1}^{j-1} t_{ij}^2 t_{ii}
\end{aligned}$$

onde mudamos a ordem dos somatórios duplos na terceira igualdade e mudamos os índices $i \leftrightarrow j$ na última igualdade. Agora, o núcleo da densidade conjunta de $\mathbf{T}$ e $\mathbf{A}$ é

$$\begin{aligned}
 f(\mathbf{T}, \mathbf{A}) &= f(\mathbf{A})f(\mathbf{T}|\mathbf{A}) \\
 &\propto \left(\prod_{i=1}^p t_{ii}^{\frac{\nu+i-p}{2}-1} \exp\left(-\frac{1}{2}t_{ii}\right) \right) \left(\prod_{i=2}^p \prod_{j=1}^{i-1} t_{ii}^{\frac{1}{2}} \exp\left(-\frac{t_{ij}^2 t_{ii}}{2}\right) \right) \\
 &= \left(\prod_{i=1}^p t_{ii}^{\frac{\nu+i-p}{2}-1} \exp\left(-\frac{1}{2}t_{ii}\right) \right) \left(\prod_{i=2}^p t_{ii}^{\frac{i-1}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{i-1} t_{ij}^2 t_{ii}\right) \right) \\
 &= \left(\prod_{i=1}^p t_{ii}^{\frac{\nu-p-3}{2}+i} \right) \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^p t_{ii} + \sum_{i=2}^p \sum_{j=1}^{i-1} t_{ij}^2 t_{ii} \right) \right\}.
 \end{aligned}$$

Pelo Lema 1, o determinante do jacobiano é $\prod_{i=1}^p t_{ii}^{-i+1}$. Substituindo $tr(\mathbf{W})$ e $det(\mathbf{W})$ na expressão acima e multiplicando o jacobiano, o núcleo da densidade de $\mathbf{W}$ é

$$det(\mathbf{W})^{\frac{\nu-p-1}{2}} \exp \left\{ -\frac{1}{2}tr(\mathbf{W}) \right\},$$

que corresponde ao núcleo da densidade Wishart $\mathcal{W}_p(\nu, \mathbf{I}_p)$. ■

Lembremos então o que diz o caso geral do teorema. As distribuições *a priori* em (3.9) e (3.10) induzem uma distribuição Wishart para a matriz $\Sigma^{-1} \equiv \mathbf{L}'\mathbf{D}^{-1}\mathbf{L}$ de modo que $\Sigma^{-1} \sim \mathcal{W}(\nu, \mathbf{R})$, em que $\mathbf{R}$ é uma matriz simétrica $p \times p$ positiva definida. Além disso, o inverso também é verdadeiro, se $\Sigma^{-1} \sim \mathcal{W}(\nu, \mathbf{R})$, então as distribuições *a priori* induzidas em d_k , $k = 1, \dots, p$ e $l_{kj}|d_k$, $1 \leq j < k \leq p$ são as apresentadas em (3.9) e (3.10), respectivamente.

Demonstração. A demonstração desse teorema será feita por construção com base na propriedade da distribuição Wishart que diz, se $\mathbf{C}'\mathbf{C} = \mathbf{R}$ e $\mathbf{W} \sim \mathcal{W}(\nu, \mathbf{I}_p)$, então $\mathbf{C}'\mathbf{W}\mathbf{C} \sim \mathcal{W}(\nu, \mathbf{R})$. Pelo Teorema (4) temos que $\mathbf{C}'\mathbf{W}\mathbf{C}$ pode ser escrita como $\mathbf{C}'\mathbf{L}'\mathbf{D}^{-1}\mathbf{L}\mathbf{C}$, onde os elementos de $\mathbf{L}$ e $\mathbf{D}$ são distribuídos em (C.1) e (C.2). Neste ponto tomamos $\mathbf{C}$ como sendo uma matriz triangular inferior, por exemplo, $\mathbf{C} = (\mathbf{P}^{-1})'$, onde $\mathbf{P}$ é o fator de Cholesky de $\mathbf{R}^{-1}$ tal que $\mathbf{P}'\mathbf{P} = \mathbf{R}^{-1}$. Como $\mathbf{L}$ e $\mathbf{C}$ são triangular inferior e a diagonal principal de $\mathbf{L}$ contém uns, o produto $\mathbf{L}\mathbf{C}$ é uma matriz triangular inferior com diagonal principal igual a diagonal principal de $\mathbf{C}$ e as entradas abaixo da diagonal principal são funções lineares das entradas de $\mathbf{L}$. Isso implica que pelo simples reescalonamento das linhas de $\mathbf{L}\mathbf{C}$ pelos valores da diagonal principal de $\mathbf{C}$, podemos escrever $\mathbf{C}'\mathbf{L}'\mathbf{D}^{-1}\mathbf{L}\mathbf{C} = \tilde{\mathbf{L}}'\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{L}}$, onde $\tilde{\mathbf{L}}$ é unitriangular inferior e $\tilde{\mathbf{D}}$ é diagonal. Além disso, como $\tilde{\mathbf{L}}$ contém combinações lineares de variáveis aleatórias normais e $\tilde{\mathbf{D}}$ contém variáveis aleatórias gama inversas reescaloadas, esses elementos tem distribuições normal e gama inversa, respectivamente. Começando com as distribuições *a priori* em (C.1) e (C.2), transformando-a por $\mathbf{C}$ chegamos aos hiper-parâmetros em (3.9) e (3.10) em que necessariamente $\Sigma^{-1} \sim \mathcal{W}(\nu, \mathbf{R})$. ■

Referências Bibliográficas

- Albert e Ghosh (2000)** J Albert e M Ghosh. Item response modeling. Em *Generalized Linear Model: A Bayesian Perspective*, páginas 173–193. New York: Marcel Dekker Inc. Citado na pág. [23](#)
- Andersen e Madsen (1977)** Erling Andersen e Mette Madsen. Estimating the parameters of the latent population distribution. *Psychometrika*, 42(3):357–374. Citado na pág. [10](#)
- Andrade et al. (2000)** Dalton Francisco Andrade, Heliton Ribeiro Tavares e Raquel da Cunha Valle. Teoria da resposta ao item: conceitos e aplicações. *ABE, São Paulo*. Citado na pág. [1](#), [5](#), [8](#), [10](#), [11](#)
- Anjos et al. (2004)** U. U Anjos, F. H. Ferreira, N. V. Kolev e B. V. M. Mendes. Modelando dependências via cópulas. *Associação Brasileira de Estatística*. Citado na pág. [24](#), [25](#), [26](#)
- Bazan (2005)** J Bazan. *Uma Família de Modelos de Resposta ao Item Normal Assimétrica*. 2005. 133 f. Tese de Doutorado, Tese (Doutorado em Estatística)–Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Citado na pág. [10](#), [71](#)
- Bazan et al. (2010)** J Bazan, L Valdivieso e M. D. Branco. Bayesian analysis in a speeded item response model. Relatório técnico, Instituto de Matemática e Estatística - IME/USP. Citado na pág. [vii](#), [13](#), [23](#), [37](#), [44](#), [51](#), [53](#), [55](#), [58](#)
- Bazan et al. (2006)** Jorge L Bazan, Heleno Bolfarine e Roseli Aparecida Leandro. Sensitivity analysis of prior specification for the probit-normal irt model: an empirical study. *Estatística*, 58:170–171. Citado na pág. [23](#)
- Bengtsson (2015)** Henrik Bengtsson. R. matlab: read and write mat files and call matlab from within r. *R package version*, 3. Citado na pág. [38](#), [44](#)
- Birnbaum (1968)** Allan Birnbaum. Some latent trait models and their use in inferring an examinee’s ability. *Statistical theories of mental test scores*. Citado na pág. [3](#)
- Bock e Aitkin (1981)** R Darrell Bock e Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, 46(4):443–459. Citado na pág. [10](#), [14](#)
- Bock e Lieberman (1970)** R Darrell Bock e Marcus Lieberman. Fitting a response model for dichotomously scored items. *Psychometrika*, 35(2):179–197. Citado na pág. [10](#)
- Bolt et al. (2002)** Daniel M Bolt, Allan S Cohen e James A Wollack. Item parameter estimation under conditions of test speededness: Application of a mixture rasch model with ordinal constraints. *Journal of Educational Measurement*, 39(4):331–348. Citado na pág. [13](#), [15](#), [16](#)
- Chan e Jeliazkov (2009)** Joshua Chi-Chun Chan e Ivan Jeliazkov. Mcmc estimation of restricted covariance matrices. *Journal of Computational and Graphical Statistics*, 18(2):457–480. Citado na pág. [28](#), [29](#), [32](#)

- Dempster (1973)** A. P. Dempster. The direct use of likelihood for significance testing. In *Proceedings of Conference on Foundational Questions in Statistical Inference*, páginas 335–352. Citado na pág. 70
- Durante et al. (2013)** Fabrizio Durante, Juan Fernandez-Sanchez e Carlo Sempi. A topological proof of sklar’s theorem. *Applied Mathematics Letters*, 26(9):945–948. Citado na pág. 25
- Embrechts et al. (2002)** Paul Embrechts, Alexander McNeil e Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, páginas 176–223. Citado na pág. 25
- Fox (2010)** Jean-Paul Fox. *Bayesian item response modeling: Theory and applications*. Springer. Citado na pág. 1, 9, 10, 66
- Fu et al. (2009)** Zhi-Hui Fu, Jian Tao e Ning-Zhong Shi. Bayesian estimation in the multidimensional three-parameter logistic model. *Journal of Statistical Computation and Simulation*, 79(6): 819–835. Citado na pág. 10, 23
- Geisser e Eddy (1979)** Seymour Geisser e William F Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160. Citado na pág. 69
- Gelfand (1996)** Alan E Gelfand. Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, páginas 145–161. Citado na pág. 70
- Gelfand et al. (1992)** Alan E Gelfand, Dipak K Dey e Hong Chang. Model determination using predictive distributions with implementation via sampling-based methods. Relatório técnico, DTIC Document. Citado na pág. 69
- Gelman e Rubin (1992)** Andrew Gelman e Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, páginas 457–472. Citado na pág. 38, 58
- Gelman et al. (2014)** Andrew Gelman, John B Carlin, Hal S Stern e Donald B Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis. Citado na pág. 53, 71
- Geweke et al. (1991)** John Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA. Citado na pág. 38, 58
- Goegebeur et al. (2008)** Yuri Goegebeur, Paul De Boeck, James A Wollack e Allan S Cohen. A speeded item response model with gradual process change. *Psychometrika*, 73(1):65–87. Citado na pág. 2, 11, 13, 15, 18, 20, 23, 55, 56
- Golub e Van Loan (2012)** Gene H Golub e Charles F Van Loan. *Matrix computations*, volume 3. JHU Press. Citado na pág. 29
- Guewehr (2007)** Katrine Guewehr. Teoria da resposta ao item na avaliação da qualidade de vida de idosos. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul. Citado na pág. 1
- Gustafson (2000)** Paul Gustafson. Local robustness in bayesian analysis. Em *Robust Bayesian Analysis*, páginas 71–88. Springer. Citado na pág. 53, 66
- Hambleton et al. (1991)** Ronald K Hambleton, H Swaminathan e H Jane Rogers. Fundamentals of item response theory (measurement methods for the social sciences series, vol. 2), 1991. Citado na pág. 8
- Insua e Ruggeri (2012)** David Ríos Insua e Fabrizio Ruggeri. *Robust Bayesian Analysis*, volume 152. Springer Science & Business Media. Citado na pág. 53

- James (2006)** Barry R James. Probabilidade: Um curso em nível intermediário, coleção euclides. *Rio de Janeiro. IMPA, 3ª Edição*. Citado na pág. 24
- Kadane (2011)** Joseph B Kadane. *Principles of uncertainty*. CRC Press. Citado na pág. 29, 67
- Keller e Isaacson (1966)** Herbert Bishop Keller e Eugene Isaacson. Analysis of numerical methods. *J. Wiley & Sons, New York*. Citado na pág. 8
- Kim (2001)** Seock-Ho Kim. An evaluation of a markov chain monte carlo method for the rasch model. *Applied Psychological Measurement*, 25(2):163–176. Citado na pág. 23
- Kim et al. (1994)** Seock-Ho Kim, Allan S Cohen, Frank B Baker, Michael J Subkoviak e Tom Leonard. An investigation of hierarchical bayes procedures in item response theory. *Psychometrika*, 59(3):405–421. Citado na pág. 23
- Lord (1952)** Frederic Lord. A theory of test scores. *Psychometric monographs*. Citado na pág. 3
- Lord (1980)** Frederic M Lord. *Applications of item response theory to practical testing problems*. Routledge. Citado na pág. 1
- Lu e Sireci (2007)** Ying Lu e Stephen G Sireci. Validity issues in test speededness. *Educational Measurement: Issues and Practice*, 26(4):29–37. Citado na pág. 2, 52
- Magalhães (2006)** Marcos Nascimento Magalhães. *Probabilidade e variáveis aleatórias*. Edusp. Citado na pág. 24
- Matos (2008)** Gilberto Matos. *Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes*. Tese de Doutorado, Universidade de Sao Paulo. Citado na pág. 8, 9
- Mislevy (1986)** Robert J Mislevy. Bayes modal estimation in item response models. *Psychometrika*, 51(2):177–195. Citado na pág. 8, 10
- Moreira Junior et al. (2013)** Fernando de Jesus Moreira Junior, Rafael Tezza, Dalton Francisco de Andrade e Antonio Cezar Bornia. Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimação da usabilidade de sites de e-commerce. *Production Journal*, 23(3):525–536. Citado na pág. 1
- Moreno (2000)** Elias Moreno. Global bayesian robustness for some classes of prior distributions. Em *Robust Bayesian Analysis*, páginas 45–70. Springer. Citado na pág. 53, 66
- Mroch et al. (2005)** Andrew A Mroch, Daniel M Bolt e James A Wollack. A new multi-class mixture rasch model for test speededness. Em *annual meeting of the National Council on Measurement in Education, Montreal, Quebec*. Citado na pág. 11, 15, 16, 17
- Muirhead (1982)** Robb J Muirhead. Aspects of multivariate statistical analysis. *John Wiley & Sons, INC., 605 Third Ave., New York, NY 10158, USA, 1982, 656*. Citado na pág. 73
- Nelsen (2013)** Roger B Nelsen. *An introduction to copulas*, volume 139. Springer Science & Business Media. Citado na pág. 25
- Partchev (2014)** Ivailo Partchev. irtoys: Simple interface to the estimation and plotting of irt models. *R package version 0.1.7*, URL <http://cran.r-project.org/packages=irtoys>. Citado na pág. 56
- Patz e Junker (1999b)** Richard J Patz e Brian W Junker. Applications and extensions of mcmc in irt: Multiple item types, missing data, and rated responses. *Journal of educational and behavioral statistics*, 24(4):342–366. Citado na pág. 10, 23
- Patz e Junker (1999a)** Richard J Patz e Brian W Junker. A straightforward approach to markov chain monte carlo methods for item response models. *Journal of educational and behavioral Statistics*, 24(2):146–178. Citado na pág. 10, 53

- Paulino et al. (2003)** Carlos Daniel Mimoso Paulino, Maria Antônia Amaral Turkman e Bento Murteira. *Estatística Bayesiana*. Fundação Calouste Gulbenkian. Citado na pág. 35, 69
- Plummer et al. (2006)** Martyn Plummer, Nicky Best, Kate Cowles e Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R news*, 6(1):7–11. Citado na pág. 38, 44, 58
- Rao (2009)** C Radhakrishna Rao. *Linear statistical inference and its applications*, volume 22. John Wiley & Sons. Citado na pág. 8
- Rizopoulos (2013)** D Rizopoulos. ltm: latent trait models under irt. *R package version 0.9-9*, URL <http://cran.r-project.org/package=ltm>. Citado na pág. 55
- Rost (1990)** Jürgen Rost. Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3):271–282. Citado na pág. 13, 15, 16
- Sahu (2002)** Sujit K Sahu. Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, 72(3):217–232. Citado na pág. 23, 53
- Schmitt et al. (2010)** TA Schmitt, DA Sass, JR Sullivan e CM Walker. A monte carlo simulation investigating the validity and reliability of ability estimation in item response theory with speeded computer adaptive tests. *International Journal of Testing*, 10(3):230–261. Citado na pág. 13
- Sheu et al. (2005)** Ching-Fan Sheu, Cheng-Te Chen, Ya-Hui Su e Wen-Chung Wang. Using sas proc nlmixed to fit item response theory models. *Behavior Research Methods*, 37(2):202–218. Citado na pág. 21, 56
- Sivaganesan (2000)** Siva Sivaganesan. Global and local robustness approaches: uses and limitations. Em *Robust Bayesian Analysis*, páginas 89–108. Springer. Citado na pág. 53, 66
- Sklar (1959)** M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8. Citado na pág. 25
- Spiegelhalter et al. (2002)** David J Spiegelhalter, Nicola G Best, Bradley P Carlin e Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639. Citado na pág. 71
- Stafford (1971)** Richard E Stafford. The speededness quotient: A new descriptive statistic for tests. *Journal of Educational Measurement*, 8(4):275–277. Citado na pág. 52
- Stern e Sinharay (2005)** Hal S Stern e Sandip Sinharay. Bayesian model checking and model diagnostics. *Bayesian thinking: modeling and computation, Handbook of Statistics*, 25:171–192. Citado na pág. 53
- Van Belle e Arnold (2000)** Gerald Van Belle e Alice Arnold. Reliability of cognitive tests used in alzheimer's disease. *Statistics in medicine*, 19(11-12):1411–1420. Citado na pág. 1
- Van der Linden (2006)** Wim J Van der Linden. A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, 31(2):181–204. Citado na pág. 66
- Vianna (1978)** Heraldo Marelim Vianna. *Testes em Educação*. IBRASA. Citado na pág. 1
- Wishart (1928)** John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, páginas 32–52. Citado na pág. 67
- Wollack e Cohen (2005)** JA Wollack e AS Cohen. A model for simulating speeded test data. Em *annual meeting of the American Educational Research Association, San Diego, CA*. Citado na pág. vii, 2, 11, 18

- Yamamoto e Everson (1997)** Kentaro Yamamoto e Howard Everson. Modeling the effects of test length and test time on parameter estimation using the hybrid model. *Applications of latent trait and latent class models in the social sciences*, páginas 89–98. Citado na pág. [11](#), [13](#), [14](#), [15](#), [18](#), [20](#)