Modelos Multidimensionais da Teoria de Resposta ao Item

Tiago de Miranda Fragoso

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP Data de Depósito: 18 de junho de 2010 Assinatura:

Modelos Multidimensionais da Teoria de Resposta ao Item 1

Tiago de Miranda Fragoso

Orientador: Prof^a. Dr^a. Mariana Cúri

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências de Computação e Matemática Computacional.

São Carlos - SP Junho/2010

 $^{^1\}mathrm{Mestrado}$ financiado pelo CNP
q e pela FAPESP

"Existe uma teoria que diz que, se um dia alguém descobrir exatamente para que serve o Universo e por que ele está aqui, ele desaparecerá instantaneamente e será substituido por algo ainda mais estranho e inesplicável." Douglas Adams em **O Restaurante no fim do Universo**

Agradecimentos

À minha orientadora, Mariana Cúri pela inestimável experiência, conselhos e paciência nos dois anos da construção do presente trabalho e ao professor Mário de Castro Andrade Filho pelas valiosas discussões e opiniões, de grande valia para a minha formação e a desse trabalho.

A Imyra, por ouvir mais sobre esse trabalho ao longo desses anos do que qualquer publico especializado, ao Gilberto pela companhia e providenciais ajudas e aos meus pais e familiares pelo suporte dado à sua própria maneira.

Resumo

Avaliações educacionais, de distúrbios psicológicos e da aceitação de um produto no mercado são exemplos de estudos que buscam quantificar um construto de interesse através de questionários compostos por itens de múltipla escolha. A Teoria de Resposta ao Item (TRI) é muito utilizada na análise de dados provenientes da aplicação desses questionários. Há vários modelos da TRI já muito utilizados na prática com tal finalidade, tanto para respostas dicotômicas aos itens (certo/errado, presente/ausente, sim/não), quanto para itens com mais de duas categorias de resposta (nominais ou ordinais). No entanto, a grande maioria supõe que apenas um traço latente é necessário para explicar a probabilidade de resposta ao item (modelos unidimensionais). Como as situações práticas são usualmente caracterizadas por várias aptidões (traços latentes) influenciando a probabilidade de um indivíduo apresentar certa resposta ao item, os modelos multidimensionais são de grande importância. Neste trabalho, após um levantamento bibliográfico dos principais modelos multidimensionais da TRI existentes na literatura, realizou-se um estudo detalhado de um deles: o modelo logístico multidimensional de dois parâmetros. O método de estimação dos parâmetros dos itens por máxima verossimilhança marginal e dos traços latentes por máxima verossimilhança são explicitados assim como a estimação por métodos bayesianos. Todos os métodos foram implementados em R, comparados e aplicados a um conjunto de dados reais para avaliação do Inventário de Depressão de Beck (BDI) e do Exame Nacional do Ensino Médio (ENEM).

Abstract

Educational evaluations, psychological testing and market surveys are examples of studies aiming to quantify an underlying construct of interest through multiple choice item tests. Item Response Theory (IRT) is a class of models used to analyse such data. There are several IRT models already being used in applied studies to such end, either for dichotomical answers (right/wrong, present/absent, Yes/No) or for itens with nominal or ordinal answers. However, the large majority of those models make the assumption that only one latent trait is sufficient to explain the probability of a correct answer to an item (unidimensional models). Since many situations in practice are characterized by multiple aptitudes (latent traits) influencing such probabilities, multidimensional models that take such traits into consideration gain great importance. In the present work, after a thorough review of the litterature regarding multidimensional IRT models, we studied in depth one model: the two parameter multidimensional logistic model for dichotomical items. The marginal maximum likelihood method used to estimate the item parameters and the maximum likelihood method used for the latent traits as well as bayesian methods for parameter estimation were studied, compared, implemented in the R software and then applied to a real dataset to infere depression using the *Beck Depression Inventory* (BDI) and the Exame Nacional do Ensino Médio (ENEM).

Sumário

1	Introdução		
2 Análise Fatorial		álise Fatorial	23
	2.1	Analise Fatorial Exploratória	23
	2.2	Análise Fatorial com Correlações Tetracóricas	25
	2.3	Análise Fatorial de Informação Plena (AFIP)	26
3 Modelos Multidimensionais da TRI		delos Multidimensionais da TRI	29
	3.1	Modelo de Ogiva Normal Compensatório	29
	3.2	Modelo Logístico Compensatório	30
	3.3	Modelos Não-Compensatórios	32
4 Estimação dos Parâmetros		33	
	4.1	Identificabilidade dos Modelos	34
	4.2	Estimação dos Parâmetros dos Itens	35
		4.2.1 Método da Máxima Verossimilhança Marginal	35
		4.2.2 Método Bayesiano da Moda a Posteriori	37
	4.3	Estimação dos Traços Latentes	38
	4.4	Estimação Conjunta: MCMC	39
		4.4.1 Diagnóstico de Convergência do Método MCMC	41
	4.5	Adequação do Modelo	43
5	5 Simulação		51
	5.1	Simulação 1	51
	5.2	Simulação 2	55

6	Apli	icação a Dados Reais	65
	6.1	Inventário de Depressão de Beck	65
	6.2	Exame Nacional do Ensino Médio (ENEM)	72
7	Con	clusão	79
A	Inventário de Depressão de Beck		
в	Esti	mativas dos Parâmetros dos Itens do ENEM	87
\mathbf{C}	Algorítmos		91
	C.1	Estimação dos Parâmetros dos Itens por MVM	91
	C.2	Erros-Padrão dos Estimadores de MVM	96
	C.3	Estimação Bayesiana dos Parâmetros dos Itens por MCMC	96
	C.4	Estimação dos Traços Latentes por Máxima Veros similhança e EAP $\ .$.	98
	C.5	Resíduos Padronizados	100
	C.6	Escore Esperado	101
	C.7	Cálculo da Estatística $s - \chi^2$	102
	C.8	Checagens Preditivas da Posteriori e DIC	103

Lista de Figuras

3.1	Superfície característica para um item (SCI) sob o modelo logístico (3.3) com dois traços latentes ($\theta_1 \in \theta_2$) e $a_1 = 1$, $a_2 = 0.7$, $b = 1, 0 \in c = 0$.	31
4.1	Ajuste Gráfico pela CCI: os pontos em azul representam os valores esti- mados, enquanto os valores em rosa representam os valores observados. Fonte: IRT Lab, University of Illinois-Urbana	45
5.1	Valores absolutos da diferença entre os valores reais e os ajustados pe- los métodos MVM e MCMC, os pontos representam os parâmetros de discriminação, enquanto os triângulos representam os parâmetros de di- ficuldado	54
5.2	Históricos dos valores amostrados no método MCMC e densidades de	94
0.2	probabilidade <i>a posteriori</i> estimadas para os parâmetros do Item 1	55
5.3	Média dos viéses para as amostras simuladas nas diferentes configura- ções. Os círculos representam os resultados para 10 itens, enquanto os triângulos representam os de 25 itens.	58
5.4	Média dos EQM para as amostras simuladas nas diferentes configura- ções utilizando o MPL e os parâmetros 5.2. Os círculos representam os resultados para 10 itens, enquanto os triângulos representam os de 25	
	itens.	63
6.1	Diagrama de Venn para os itens do BDI	67
6.2	Gráfico de vetores dos resultados da aplicação do ML2P aos itens do BDI	68
6.3	Histogramas das componentes de $\widehat{\theta}$ para estimativas por EAP (itens por MVM) e MCMC.	72
6.4	Normal Q-Q Plots das componentes de $\widehat{\boldsymbol{\theta}}$ para estimativas por EAP	
	obtidas de ambos os métodos	73

6.5 $$ Escores esperados pelo M2LP (linha pontilhada) e observados (pontos		
	para os dados do BDI	73
6.6	Escores observados (pontos interligados), escores esperados obtidos pelo	
	método MCMC (linha cheia) e intervalo de 95% de confiança (linha	
	pontilhada)	74
6.7	Poder discriminativo (quadrados) e dificuldade (círculos) para os itens	
	do ENEM	75
6.8	Histogramas das componentes de $\widehat{oldsymbol{ heta}}$ para estimativas por EAP \ldots .	76
6.9	Q-Q plots das componentes de $\widehat{\boldsymbol{ heta}}$ para estimativas por EAP \ldots	77
6.10	Escores esperados pelo M2LP (linha pontilhada) e observados (pontos	
	interligados) para os dados do ENEM	77

Lista de Tabelas

2.1	Probabilidades de resposta aos items $l \in m$	25
5.1	Estimativas dos parâmetros dos itens utilizando o MVM, o programa TESTFACT (TF), o Amostrador de Gibbs (MCMC) e respectivos erros	
	padrão $(\hat{\sigma})$, parâmetros de discriminação $(a_1,\ a_2)$ e dificuldade $(b).$	53
5.2	Parâmetros utilizados nas simulações	56
5.3	Média dos viéses para cada configuração de amostra simulada $\ .\ .\ .$	57
5.4	Máximo dos viéses para cada configuração de amostra simulada $\ .\ .\ .$	59
5.5	Mínimo dos viéses para cada configuração de amostra simulada $\ \ .\ .\ .$	60
5.6	Média dos EQM para cada configuração de amostra gerada utilizando	
	os parâmetros em 5.2 e o MPL	62
5.7	Máximo dos EQM para cada configuração de amostra simulada $\ .$	64
6.1	Valores das estatísticas de seleção de modelos para os dados do BDI, gl:	
	graus de liberdade da estatística qui-quadrado	66
6.2	Estimativas dos Parâmetros do ML2P aplicados aos dados do BDI utili-	
	zando o Método da Máxima Verossimilhança Marginal (MVM) e MCMC.	
	As estimativas dos erros padrão são denotados entre parênteses	70
6.3	Média($\boldsymbol{\mu}$), variância (σ^2) e correlações (σ_{12}) das estimativas obtidas por	
	EAP para o vetor de traços latentes utilizando ambos os métodos	71
6.4	Valores das estatísticas de seleção de modelos para diversos modelos	
	utilizados no ajuste dos dados do ENEM	74
6.5	Média(μ) e variância (σ^2) das estimativas obtidas para o vetor de traços	
	latentes por EAP	75
B.1	Estimativas para os parâmetros dos itens do ENEM	87

Lista de Abreviações

AF: análise fatorial
AFIP: análise fatorial de informação plena
BDI: inventário de depressão de Beck
SCI: superficie característica do item
EAP: esperança a *a posteriori*MAP: moda a *a posteriori*MCMC: monte carlo via cadeia de Markov
ML1P: modelo logístico multidimensional de um parâmetro
ML2P: modelo logístico multidimensional de dois parâmetros
MN2P : modelo logístico não-compensatório de dois parâmetros
ML3P: modelo logístico multidimensional de três parâmetros
ML3P: modelo logístico multidimensional de três parâmetros
MV: máxima verossimilhança
MVM: máxima verossimilhança marginal
UL2P: modelo logístico unidimensional de dois parâmetros

TRI: teoria de resposta ao item

1

Capítulo

Introdução

A Teoria de Resposta ao Item (TRI) é um conjunto de modelos estatísticos que relacionam traços latentes (tais como o conhecimento em uma determinada disciplina, na área de Educação, e presença ou intensidade de uma determinada patologia ou comportamento, na área de Saúde) com a probabilidade de uma certa resposta a itens de múltipla escolha compondo uma prova ou questionário.

Os modelos mais usuais da TRI apoiam-se na suposição de que existe um único traço latente que influencia a probabilidade de determinada resposta, ou pelo menos existe um único traço dominante que responde pelo desempenho do indivíduo na prova ou questionário. Estes modelos são denominados modelos unidimensionais.

Contudo, tal suposição nem sempre é válida na prática, onde comumente mais de um traço latente tem papel significativo no desempenho do indivíduo ao longo da prova. Por exemplo, em uma prova que objetiva medir a fluência de um indivíduo em um determinado idioma, vários fatores podem influenciar na resposta correta ou não a um item, tais como seu vocabulário, interpretação de textos ou habilidade de leitura no idioma em questão. Também pode ser intuito do examinador estimar o conhecimento do examinado em cada uma dessas sub-áreas para uma interpretação mais apurada da fluência do indivíduo.

Outro problema em que um único traço latente pode não ser suficiente é quando temos disparidade entre as populações avaliadas. Por exemplo, anseia-se medir o conhecimento básico de Matemática em uma população. Se a prova for aplicada a uma população instruída, a leitura e interpretação das perguntas dos itens não influenciam as respostas dos indivíduos, que serão relacionadas unicamente ao conhecimento de Matemática. Entretanto, se aplicarmos o mesmo procedimento a uma população com baixa instrução, a estimação do conhecimento em Matemática pode não condizer com a realidade, pois os indivíduos são influenciados por fatores alheios, tais como falta de compreensão do vocabulário ou tempo insuficiente para a resolução da questão.

Nesse contexto, um modelo que considere vários traços latentes deve ser utilizado para descrever os atributos da população. Na literatura, são propostas extensões dos modelos unidimensionais da TRI considerando que o traço latente θ outrora um escalar é definido como um vetor $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$ de traços latentes que influenciam as respostas aos itens.

O número de traços latentes que deve ser considerado no modelo depende da heterogenidade da população alvo com relação aos níveis dessas dimensões e de quais traços efetivamente influenciam nas respostas dos indivíduos aos itens (Linden & Hambleton 1996). Escolher um número que superestime a quantidade de traços latentes não é recomendável devido à dificuldade na interpretação de cada traço, enquanto que subestimar o número de dimensões pode prejudicar a adequação do modelo aos dados.

No propósito de determinarmos a dimensão do vetor de parâmetros $\boldsymbol{\theta}$, técnicas de Análise Fatorial (AF) são utilizadas. Contudo, a natureza dicotomica dos itens limita o uso da análise fatorial exploratória mais frequentemente encontrada na literatura (Johnson & Wichern 2003) e implementada nos programas estatísticos mais tradicionais. Nesse sentido, utilizaremos a abordagem descrita por Bock & Aitkin (1981) e Bock, Gibbons & Muraki (1988), na qual tratamentos para a análise dos itens dicotomicos e estimação das cargas dos fatores são propostas através da técnica denominada análise fatorial de informação plena.

No Capítulo 2 apresentamos uma breve descrição com diferenças e semelhanças entre as analises fatoriais exploratória para respostas contínuas, para dados dicotômicos substituindo-se a matriz de correlação de Pearson pelas correlações tetracóricas e a de informação plena, foco principal deste trabalho.

No Capítulo 3 apresentamos alguns modelos multidimensionais da TRI: modelos compensatórios de ogiva normal e logístico de três parâmetros e um modelo logístico não-compensatório.

Descrevemos a estimação dos parâmetros para os modelos multidimensionais compensatórios utilizando o método da Máxima Verossimilhança Marginal e métodos Bayesianos no Capítulo 4 e terminamos o presente trabalho no Capítulo 5 com um estudo de simulação para um estudo comparativo dos métodos, sendo um deles aplicado a um conjunto de dados reais para o diagnóstico de depressão através do Inventário de Depressão de Beck (BDI).

Capítulo

2

Análise Fatorial

A teoria de AF surgiu das idéias de Thurstone e Spearman (Reckase 1997) com o estudo das associações entre variáveis que refletem a inteligência e identificação dos fatores latentes que justifiquem tais associações.

No contexto da AF, são medidas diversas respostas, que podem ser explicadas por um número menor de traços latentes (variáveis não observadas, chamadas de *fatores*) para as quais analisamos as respectivas estruturas de correlação. Um breve relato da metodologia da Análise Fatorial será apresentado nessa seção, tanto para respostas observadas contínuas (Seção 2.1) quanto dicotômicas (Seções 2.2 e 2.3).

2.1 Analise Fatorial Exploratória

Suponha *n* variáveis resposta contínuas padronizadas (ou seja, com média zero e variância unitária), $Y_1, Y_2, ..., Y_n$, associadas a um grupo de *p* variáveis aleatórias não observadas, os fatores $F_1, F_2, ..., F_p$, segundo o seguinte modelo:

$$Y_{1} = l_{11}F_{1} + \dots + l_{1p}F_{p} + \epsilon_{1}$$

$$Y_{2} = l_{21}F_{1} + \dots + l_{2p}F_{p} + \epsilon_{2}$$

$$\vdots$$

$$Y_{n} = l_{n1}F_{1} + \dots + l_{np}F_{p} + \epsilon_{n}$$
(2.1)

Em que os coeficientes l_{ij} i = 1, ..., n, j = 1, ..., p, são denominados cargas fatoriais e representam o peso de cada fator na resposta observada e $\epsilon_1, ..., \epsilon_n$ são erros aleatórios. Equivalentemente, em forma matricial,

$$\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon}, \tag{2.2}$$

em que L, de dimensão (nxp), é demominada matriz de cargas fatoriais, $\mathbf{F} = (F_1, \ldots, F_p)^t$ e $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)^t$. De tal formulação, segue que podemos decompor a matriz de covariâncias das variáveis observadas (ou a matriz de correlações, já que utilizamos as variáveis padronizadas), $\boldsymbol{\Phi}$, na seguinte forma:

$$\boldsymbol{\Phi} = \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^t + \boldsymbol{\Psi},\tag{2.3}$$

em que Θ é a matriz de covariâncias dos fatores e Ψ é a matriz de covariâncias dos erros. É comum a suposição de que F e ϵ são independentes, E(F) = 0, $Cov(F) = I_p$, $E(\epsilon) = 0$ e $Cov(\epsilon) = \Psi$, sendo que Ψ é uma matriz diagonal.

Tais suposições caracterizam o *modelo de fatores ortogonais*. Dessa forma, o sistema de equações anterior pode ser simplificado para

$$\boldsymbol{\Phi} = \boldsymbol{L}\boldsymbol{L}^t + \boldsymbol{\Psi}. \tag{2.4}$$

Assim temos claramente que $Cov(F_i, Y_j) = l_{ij}$; logo, as cargas fatoriais representam a covariância entre os fatores e as respostas e o objetivo da análise é estimá-las juntamente com a matriz de covariâncias dos erros, os parâmetros do modelo da Análise Fatorial. Para tal estimação, vários métodos podem ser empregados.

Um deles é o método dos fatores principais, herdado da Análise de Componentes Principais. Nele, assim como feito na análise citada, decompomos a matriz de covariâncias ou correlações em seu espectro de autovalores e autovetores, de onde estimamos a matriz L das cargas fatoriais, e determinamos o número de fatores necessários para uma boa explicação da variabilidade da amostra observando a proporção da variância explicada adicionando-se cada fator.

Outro método para a estimação é o método de Máxima Verossimilhança. Para isso, supomos que \mathbf{F} tem distribuição normal *p*-variada com vetor de médias nulo e matriz de covariâncias igual à identidade de ordem *p*, enquanto $\boldsymbol{\epsilon}$ tem distribuição normal *n*-variada, de vetor de médias nulo e matriz de covariância $\boldsymbol{\Psi}$ como definida anteriormente, o que implica que o vetor de variáveis resposta \boldsymbol{Y} tem distribuição normal *n*-variada com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\boldsymbol{L}\boldsymbol{L}^t + \boldsymbol{\Psi}$.

Contudo, temos um problema de idetificabilidade no modelo da Análise Fatorial. Observe que a matriz de covariâncias Φ permanece a mesma para transformações do tipo $L^* = LM$ e $F^* = M^t F$, sendo M uma matriz ortogonal. Neste caso,

$$\boldsymbol{L}^* \boldsymbol{F}^* + \boldsymbol{\epsilon} = \boldsymbol{L} \boldsymbol{M} \boldsymbol{M}^t \boldsymbol{F} + \boldsymbol{\epsilon} = \boldsymbol{L} \boldsymbol{F} + \boldsymbol{\epsilon}. \tag{2.5}$$

Ambas $L \in L^*$ (e, consequentemente, $F \in F^*$) representam uma solução matemática para o problema. No entanto, no sentido da interpretação dos fatores em questão, uma delas provavelmente será mais adequada do que a outra. Neste contexto, a literatura propõe alguns métodos de rotação dos fatores, que apontam determinadas transformações de L como mais interpretáveis. As rotações mais comuns são a rotação *Varimax* e as oblíquas. Tais tecnicas podem ser encontradas em Johnson & Wichern (2003) e Lawley & Maxwell (1971), respectivamente.

Quando temos variáveis contínuas, os elementos da matriz de correlações Φ considerados na Análise Fatorial são estimados pelos coeficientes da matriz de correlação de Pearson. Contudo, quando as variáveis respostas representam o acerto ou erro em um determinado item, isto é, quando as variaveis observadas são dicotômicas, a utilização das correlações de Pearson como estimadores dos elementos matriz Φ tornase inadequada. Quando tratamos itens de dificuldades muito díspares,ou seja, itens nos quais a maioria da população avaliada responde incorretamente, a correlação amostral estima valores espúrios com pequenas variações nas respostas dos itens em questão.

2.2 Análise Fatorial com Correlações Tetracóricas

Uma alternativa para a análise fatorial para respostas dicotômicas, proposta por Bock et al. (1988), é supor uma variável aleatória não observada Y_i contínua, associada à resposta observada X_i ao *i*-ésimo item de múltipla escolha, sendo 1 definido arbitrariamente como um acerto e 0 como um erro, tal que

$$X_i = \begin{cases} 0 & se \ Y_i < \gamma_i \\ 1 & se \ Y_i \ge \gamma_i \end{cases}$$
(2.6)

em que γ_i é um parâmetro desconhecido denominado ponto de corte (*threshold*) associado ao item $i, i = 1, \dots, I$.

Sejam $\pi_{10}^{l,m}, \pi_{.0}^{m}, \pi_{01}^{l,m}$ as probabilidades da tabela de contingência para os itens $l \in m$ conforme representados na Tabela 2.1.

		-	
$l\backslash m$	Acerto	Erro	Total
Acerto	$\pi_{11}^{l,m}$	π^l_{10}	$\pi_{1.}^l$
Erro	$\pi_{01}^{l,m}$	π^l_{00}	$\pi_{0.}^{l}$
Total	$\pi^m_{.1}$	$\pi^m_{.0}$	1

Tabela 2.1: Probabilidades de resposta aos items $l \in m$

O coeficiente de correlação tetracórica entre os itens $l \in m$, ρ_{lm} é o número real que satisfaz:

$$\pi_{10}^{l,m} = P\left(Y_l \ge \gamma_l, Y_m < \gamma_m\right) = \int_{\gamma_l}^{\infty} \int_{-\infty}^{\gamma_m} f(y_l, y_m, \rho_{lm}) dy_m dy_l$$

$$\pi_{01}^{l,m} = P\left(Y_m \ge \gamma_m, Y_l < \gamma_l\right) = \int_{-\infty}^{\gamma_l} \int_{\gamma_m}^{\infty} f(y_l, y_m, \rho_{lm}) dy_m dy_l \qquad (2.7)$$

$$\pi_{.0}^m = P\left(Y_m < \gamma_m\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\gamma_m} f(y_l, y_m, \rho_{lm}) dy_m dy_l,$$

em que $f(\cdot)$ é a função densidade de probabilidade da distribuição normal bivariada e $\rho_{lm} = Corr(Y_l, Y_m)$. Não existem fórmulas fechadas para o cálculo da correlação tetracórica. Divgi (1979) apresenta diversos métodos de estimação para a correlação tetracórica, que comumente estão implentados nos programas estatísticos que fazem a estimação das cargas fatoriais e da matriz de covariâncias.

Além disso, a matriz de correlações tetracóricas nem sempre é positiva definida. Bock et al. (1988) apresentam correções para o cálculo da matriz de correlações tetracóricas em situações de acertos casuais e respostas omitidas, que podem causar esse problema.

2.3 Análise Fatorial de Informação Plena (AFIP)

As dificuldades apresentadas na seção anterior foram contornadas por Bock et al. (1988) ao proporem a AFIP, uma metodologia baseada na Teoria de Resposta ao Item que evita o problema de calcular as correlações tetracóricas entre os itens. A AFIP toma o modelo

$$Y_{1} = l_{11}\theta_{1} + \dots + l_{1p}\theta_{p} + \epsilon_{1}$$

$$Y_{2} = l_{21}\theta_{1} + \dots + l_{2p}\theta_{p} + \epsilon_{2}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$Y_{i} = l_{i1}\theta_{1} + \dots + l_{ip}\theta_{p} + \epsilon_{i}$$

$$\vdots \qquad \vdots$$

$$Y_{n} = l_{n1}\theta_{1} + \dots + l_{np}\theta_{n} + \epsilon_{n}$$

$$(2.8)$$

em que os erros aleatórios ϵ_i têm distribuição normal, de média 0 e variância σ_i^2 , sendo independentes entre si e do vetor de traços latentes $\boldsymbol{\theta}$, e que $\boldsymbol{\theta}$ tem distribuição normal multivariada com vetor de médias **0** e matriz de covariâncias \boldsymbol{I}_p , implicando uma distribuição normal *n*-variada para o vetor de respostas \boldsymbol{Y} como em (2.6). Desta forma, pode-se escrever a probabilidade do indivíduo acertar o i-esimo item como

$$P(X_{i} = 1|\boldsymbol{\theta}) = P(Y_{i} \ge \gamma_{i}|\boldsymbol{\theta})$$

$$= \frac{1}{\sqrt{2\pi\sigma_{i}^{2}}} \int_{-\infty}^{-\gamma_{i}} \exp\left\{-\frac{1}{2\sigma_{i}^{2}}\left(y - \sum_{k=1}^{p} l_{ik}\theta_{k}\right)\right\} dy$$

$$= \Phi\left(-\frac{\gamma_{i} + \sum_{k=1}^{p} l_{ik}\theta_{k}}{\sigma_{i}}\right) = \Phi_{i}(\boldsymbol{\theta}), \qquad (2.9)$$

em que $\Phi(\cdot)$ é a função distribuição acumulada da normal padrão e $\Phi_i(\boldsymbol{\theta})$ é definido como a função distribuição acumulada da distribuição normal calculada no referido ponto, dependendo assim de todos os parâmetros da Análise Fatorial (cargas fatoriais, variância dos erros e ponto de corte do item) e dos traços latentes do indivíduo.

Como é usual na análise fatorial que Y_i tenha média 0 e variância 1 (variável padronizada) a suposição de que $\sigma_i^2 = 1 - \sum_{k=1}^p l_{ik}^2$ também é adotada. O claro benefício de usar a AFIP é a economia nos cálculos, já que um item a mais indica apenas mais uma probabilidade a ser calculada, ao invés da grande quantidade de tabelas de acertos entre itens necessárias para a correlação tetracórica, além de prover estimativas com interpretações no contexto da TRI, como apresentamos no capítulo 3.

Capítulo

Modelos Multidimensionais da TRI

A TRI multidimensional pode ser considerada um tipo de análise fatorial para dados observados não contínuos (dicotômicos ou politonicos). O que diferencia as duas metodologias é essencialmente a interpretação. A análise fatorial normalmente é usada para encontrar um pequeno número de variáveis latentes (fatores) capazes de explicar a variabilidade de um número maior de variáveis observadas, i.e., é basicamente uma técnica de redução de dimensão.

Por outro lado, a TRI modela a interação entre indivíduos e itens, buscando determinar caracteristicas dos indivíduos (traços latentes) e dos itens que influenciam nas respostas observadas (Reckase 1997).

No presente capítulo são apresentados modelos multidimensionais da TRI para itens dicotômicos, separados em três classes: os modelos compensatórios de ogiva normal e logísticos (herdados da AFIP) e os modelos não compensatórios. Um modelo é dito compensatório quando a mesma probabilidade de acerto do item pe mantida quando a diminuição do valor de um traço latente é compensada pelo aumento no valor de outro traço latente.

3.1 Modelo de Ogiva Normal Compensatório

Considere o modelo 2.9 com a seguinte reparametrização proposta por Bock et al. (1988)

$$b_i = \frac{-\gamma_i}{\sigma_i} \quad e \quad a_{ik} = \frac{l_{ik}}{\sigma_i},\tag{3.1}$$

em que i = 1, ..., I, sendo I o número de itens na prova, e k = 1, ..., p, sendo p a dimensão do vetor de traços latentes. O parâmetro b_i é denominado dificuldade do *i*-esimo item e cada componente do vetor $\mathbf{a}_i = (a_{1i}, \cdots, a_{pi})$ é a discriminação do *i*-ésimo item sobre o k-ésimo traço latente.

Incorpora-se uma probabilidade de acerto casual do item c_i , i = 1, ..., I quando as habilidades são baixas. Assim, temos o Modelo de Ogiva Normal de 3 Parâmetros (discriminação, dificuldade e acerto ao acaso):

$$P(X_i = 1 | \boldsymbol{a}_i, b_i, c_i, \boldsymbol{\theta}) = c_i + (1 - c_i) \Phi\left(b_i + \sum_{k=1}^p a_{ik} \theta_k\right) = \Phi_i(\boldsymbol{\theta})$$
(3.2)

em que $\mathbf{a}_i = (a_1, \ldots, a_p), b_i, \Phi(.)$ e $\boldsymbol{\theta}$ são definidos como em (2.9) e (3.1). Vale ressaltar que a equação (3.2) equivale a (2.9) reformulada segundo os parâmetros usuais da TRI.

Os modelos de ogiva normal são amplamente utilizados e encontram-se implementados em diversos programas existentes, como o NOHARM (Frasier 1988), que estima os parâmetros do modelo por uma adaptação do método dos mínimos quadrados e o TESTFACT (Wilson, Wood & Gibbons 1987) que estima os parâmetros dos itens segundo o método da máxima verossimilhança marginal e os parâmetros dos indivíduos por métodos bayesianos de esperança ou moda a posteriori a serem expostos no capítulo 4.

3.2 Modelo Logístico Compensatório

McKinley & Reckase (1980) propõem o modelo logístico de três parâmetros com expressão dada por

$$P(X_i = 1 | \boldsymbol{\theta}, \boldsymbol{a}_i, b_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + \exp\left[-\sum_{k=1}^{p} a_{ki} \theta_k + b_i\right]} = P_i(\boldsymbol{\theta}), \quad (3.3)$$

em que i = 1, ..., I, a_i , b_i e c_i são definidos como em (3.2) e $P_i(\boldsymbol{\theta})$ é definido para simplificar a notação, dependendo dos parâmetros dos itens (vetor de discriminações, dificuldade e acerto casual) e dos traços latentes dos indivíduos. Note que este modelo é muito parecido com o modelo (3.2) substituindo a função de distribuição de probabilidades da distribuição normal pela função logística.

Assim como nos modelos unidimensionais temos uma curva relacionando a probabilidade de acerto de cada item com a habilidade do indivíduo. Nos modelos multidimensionais da TRI temos uma superfície para cada relacionando a probabilidade de acerto ao item e habilidades. Para fins de ilustração, tomemos o caso bidimensional com $a_1 = 1$, $a_2 = 0.7$, b = 1,0 e c = 0. A superfície de resposta do modelo (3.3) encontra-se representada na Figura 3.1.



Figura 3.1: Superfície característica para um item (SCI) sob o modelo logístico (3.3) com dois traços latentes ($\theta_1 \in \theta_2$) e $a_1 = 1$, $a_2 = 0.7$, $b = 1, 0 \in c = 0$

Note que o parâmetro de acerto casual é a assíntota inferior da superfície, enquanto os parâmetros de discriminação representados no vetor \boldsymbol{a}_i são responsáveis pela sua inclinação em cada eixo dos traços latentes, cada elemento do vetor indica a sensibilidade do item para avaliar cada traço latente. Para medir o *poder discriminativo* do *i*-ésimo item para a combinação de todos os traços latentes, usa-se a estatística

$$MDISC_i = \sqrt{\sum_{k=1}^{p} a_{ki}^2},\tag{3.4}$$

ou seja, a norma do vetor de discriminações. Essa estatística mede a intensidade da transição da região de baixa probabilidade de acerto ao item para a região de alta probabilidade na SCI.

Nos modelos unidimensionais, o parâmetro de dificuldade do item é medido na mesma escala do traço latente e significa o ponto de inflexão da curva característica do item, ou seja, o ponto onde obtemos uma probabilidade de acerto de $\frac{1+c_i}{2}$. Nos modelos multidimensionais, b_i não pode ser interpretado dessa forma. Reckase (1996) propõe uma estatística análoga à existente no caso unidimensional para medir a distancia de origem até o ponto de maior inclinação da superfície (SCI), dada por

$$MDIFF_i = \frac{-b_i}{MDISC_i}.$$
(3.5)

Sheu, Chen, Su & Wang (2005) implementam a estimação dos parâmetros do modelo logístico no SAS utilizando o proc NLMIXED, e programas comerciais como o MAXLOG (McKinley & Reckase 1983), o MULTIDIM (McKinley 1987) e o MIRTE (Carlson 1987) realizam a estimação dos parâmetros dos itens.

Assim como no caso unidimensional, os modelos a três parâmetros podem ser simplificados para dois ou um parâmetro igualando-se c_i a 0 para i = 1, ..., I ou $c_i = 0$ e $a_{ik} = 1$ para i = 1, ..., I e k = 1, ..., p respectivamente. Mais detalhes são discutidos em McKinley & Reckase (1980).

3.3 Modelos Não-Compensatórios

O modelo logístico não compensatório de três parâmetros proposto por Whitely (1980) e Sympson (1977) é dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}, \boldsymbol{a}_i, \boldsymbol{b}_i, c_i) = c_i + (1 - c_i) \prod_{k=1}^p \frac{1}{1 + \exp(-Da_{ki}\theta_{jk} + b_{ik})}, \quad (3.6)$$

em que a_i , $c_i \in \boldsymbol{\theta}$ são definidos como em (3.3) e $\boldsymbol{b}_i = (b_{i1}, ..., b_{pi})$ representa a dificuldade do item em questão associada a cada traço latente.

Diversos outros modelos compensatórios foram propostos (Reckase 1997), entretanto tais modelos não são frequentemente utilizados por serem analiticamente complicados, tornando a implementação de métodos de estimação dos parâmetros extremamente dificil. Com o aumento do poder computacional e novos métodos de estimação, os modelos não-compensatórios podem vir a obter alguma popularidade nas aplicações da TRI (Bolt & Lall 2003).

Todos os modelos aqui descritos são adequados para itens dicotômicos. Propostas para itens politômicos também são encontradas na literatura. Entre elas podemos citar Adams, Wilson & Wang (1997), Ayreshi (1993) e Kelderman & Rijkes (1994).

Capítulo

4

Estimação dos Parâmetros

Os modelos da TRI envolvem um grande número de parâmetros a serem estimados, tanto referentrees a itens (denominados parâmetros estruturais), quanto referentes aos traços latentes dos indivíduos (parâmetros incidentais). Claramente, o número de parâmetros referentes aos traços latentes aumenta com o tamanho da amostra, caracterizando o problema descrito em Neyman & Scott (1948).

Bock & Aitkin (1981) sinalizam que esse fato gera estimadores de MV que não necessariamente apresentam as propriedades assintóticas como ausência de viés, consistência e eficiência mesmo em modelos mais simples como os modelos unidimensionais da TRI (Baker & Kim 2004).

Como consequência, métodos que marginalizam a função de verossimilhança com respeito aos parâmetros incidentais ou métodos bayesianos são mais utilizados na TRI.

Neste capítulo detalharemos o processo de estimação dos parâmetros do modelo (3.3) supondo $c_i = 0, i = 1, ..., I$. Para os parâmetros dos itens, serão descritos os métodos de máxima verossimilhança marginal (MVM) proposto por Bock et al. (1988) e bayesiano por moda a posteriori (MAP).

Os traços latentes serão estimados supondo os parâmetros dos itens conhecidos através dos um métodos bayesiano de esperança a posteriori (EAP) e Máxima Verossimilhança.

Adicionalmente, o método monte carlo via cadeias de Markov (MCMC) será adotado para a estimação conjunta dos parâmetros dos itens e traços latentes, como proposto por Patz & Junker (1999) no contexto da TRI.

4.1 Identificabilidade dos Modelos

Assim como os modelos unidimensionais, os modelos multidimensionais da TRI não são identificáveis. Nos modelos unidimensionais, tomando $\alpha, \beta \in \mathbb{R}, \alpha \neq 0, i \in 1, \dots, I$

$$\theta^* = \alpha \theta + \beta \tag{4.1}$$
$$a_i^* = \frac{a_i}{\alpha}$$
$$b_i^* = b_i - \frac{a_i \beta}{\alpha}$$

criava um problema de identificabilidade. A transformação (4.1) foi obtida modificandose a transformação exposta em Andrade, Tavares & Valle (2000) para a probabilidade de resposta correta ao item obtida fixando-se p = 1 nos modelos do capítulo 3.

Para os parâmetros de habilidade, discriminação e dificuldade, respectivamente, geram a mesma probabilidade de acerto ao item *i*. Analogamente, no modelo multidimensional tomando uma matriz inversivel Λ de dimensão (pxp) e um vetor β e considerando que a combinação linear entre traços latentes e discriminações dos modelos (3.2) e (3.3) podem ser escritas como $a^t \theta$, *t* indicando a transposta do vetor, temos analogamente das transformações feitas em (4.1)

$$\boldsymbol{\theta}^* = \Lambda \boldsymbol{\theta} + \boldsymbol{\beta}$$

$$\boldsymbol{a}_i^* = \boldsymbol{a}_i^t \Lambda^{-1}$$

$$\boldsymbol{b}_i^* = \boldsymbol{b}_i - \boldsymbol{a}_i^t \Lambda^{-1} \boldsymbol{\beta}.$$
(4.2)

Podemos observar que $\boldsymbol{a}_i^* \boldsymbol{\theta}^* + b_i^* = \boldsymbol{a}_i \boldsymbol{\theta} + b_i$ fornecendo assim as mesmas probabilidades de acerto.

Para resolver esse problema, programas que realizam a estimação por Máxima Verossimilhança Marginal como o TESTFACT assumem que os traços latentes seguem uma distribuição normal p-variada, com vetor de médias nulo e matriz de variânciacovariâncias idêntica à matriz identidade de ordem p.

Não é difícil observar que tal premissa resolve o problema de identificabilidade no processo de estimação, tendo em vista que uma transformação linear Λ nos traços latentes mudariam a estrutura da matriz de variâncias-covariâncias enquanto uma translação na direção de um vetor não-nulo β alteraria a média da variável transformada. Reforçando essa premissa, portanto, garantem-se estimativas únicas dos parâmetros dos itens.

Contudo, a identificabilidade obtida por essa premissa adicional não vem sem um custo. Como destacado por Reckase em Linden & Hambleton (1996), tais premissas podem aumentar a variância e covariância das estimativas dos parâmetros de discriminação, resultando em erros maiores.

Da mesma maneira que a Análise Fatorial, entretanto, podemos tirar vantagem da invariância do modelo por rotações dos eixos de traços latentes. Após o processo de estimação, podemos da mesma maneira executar rotações como o Varimax para obter interpretações melhores dos parâmetros dos itens.

4.2 Estimação dos Parâmetros dos Itens

4.2.1 Método da Máxima Verossimilhança Marginal

Denotando por j = 1, ..., n os indivíduos que respondem aos itens $i = 1, ..., I, \boldsymbol{\theta}_j$, o vetor dos traços latentes do *j*-ésimo indivíduo, $x_{ij} = 0$ ou 1, a resposta do *j*-ésimo indivíduo ao *i*-ésimo item, $\boldsymbol{x}_j = (x_{1j}, ..., x_{Ij})$ e $P_i(\boldsymbol{\theta})$ como definido em (3.3) para $c_i = 0, i = 1, ..., I$, podemos escrever a função de verossimilhança como

$$L_j(\boldsymbol{\theta}) = P(\boldsymbol{X}_j = \boldsymbol{x}_j | \boldsymbol{\theta}_j, \boldsymbol{a}_i, b_i) = \prod_{i=1}^{I} \left[P_i(\boldsymbol{\theta}_j) \right]^{x_{ij}} \left[1 - P_i(\boldsymbol{\theta}_j) \right]^{1 - x_{ij}}.$$
 (4.3)

Considerando, $g(\cdot)$ a densidade de probabilidade da normal *p*-variada padrão suposta para $\boldsymbol{\theta}_j$ como mencionada no capítulo 2, a função de verossimilhança é

$$P(\boldsymbol{X}_{j} = \boldsymbol{x}_{j} | \boldsymbol{a}_{i}, b_{i}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^{I} [P_{i}(\boldsymbol{\theta})]^{x_{ij}} [1 - P_{i}(\boldsymbol{\theta})]^{1 - x_{ij}} g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
$$= \int_{\boldsymbol{\theta}} L_{j}(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} = \tilde{P}_{j}.$$
(4.4)

Calcularemos a integral utilizando o método de quadratura gaussiana, no qual utilizamos a seguinte aproximação

$$\tilde{P}_{l} = \sum_{q_{p}=1}^{Q} \dots \sum_{q_{1}=1}^{Q} L_{l}(\boldsymbol{K}) A(K_{q_{1}}) \dots A(K_{q_{p}})$$
(4.5)

de modo que temos Q pontos de quadratura $\mathbf{K} = (K_{1k}, \dots, K_{Qk})$ em cada dimensão de traços latentes $k = 1, \dots, p$, cada um com um peso A(·). Essa aproximação consiste de um agrupamento das amostras em torno de certos níveis de traços latentes, o que nos induz a escrever a função verossimilhança por uma distribuição multinomial, reduzindo o tempo de processamento computacional.

Seja r_l a frequência do padrão de resposta \boldsymbol{x}_l , para cada um dos $s \leq \min \{2^I, n\}$ padrões de resposta possíveis. Temos que a função verossimilhança pode ser escrita como:

$$L = \frac{n!}{r_1! \dots r_s!} \tilde{P}_1^{r_1} \dots \tilde{P}_s^{r_s}.$$
 (4.6)

A derivada da função log-verossimilhança com respeito a um parâmetro de item (discriminação com relação a um traço latente ou dificuldade do item), v_i , é dada por

$$\frac{\partial \log(L)}{\partial v_{i}} = \sum_{l=1}^{s} \frac{r_{l}}{\tilde{P}_{l}} \left(\frac{\partial \tilde{P}_{l}}{\partial v_{i}} \right)$$

$$= \sum_{l=1}^{s} \frac{r_{l}}{\tilde{P}_{l}} \int_{\boldsymbol{\theta}} \left(\frac{L_{l}(\boldsymbol{\theta})}{\left[P_{i}(\boldsymbol{\theta})\right]^{x_{li}} \left[1 - P_{i}(\boldsymbol{\theta})\right]^{1 - x_{li}}} \cdot \frac{\partial \left[P_{i}(\boldsymbol{\theta})\right]^{x_{li}} \left[1 - P_{i}(\boldsymbol{\theta})\right]^{1 - x_{li}}}{\partial v_{i}} g(\boldsymbol{\theta}) \right) d\boldsymbol{\theta}$$

$$= \sum_{l=1}^{s} \frac{r_{l}}{\tilde{P}_{l}} \int_{\boldsymbol{\theta}} \left(\frac{x_{li} - P_{i}(\boldsymbol{\theta})}{P_{i}(\boldsymbol{\theta}) \left[1 - P_{i}(\boldsymbol{\theta})\right]} \right) L_{l}(\boldsymbol{\theta}) \frac{\partial P_{i}(\boldsymbol{\theta})}{\partial v_{i}} g(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{4.7}$$

Definindo

$$\bar{R}_l = \sum_{l=1}^s \frac{r_l x_{li} L_l(\boldsymbol{\theta})}{\tilde{P}_l} \tag{4.8}$$

е

$$\bar{n} = \sum_{l=1}^{s} \frac{r_l L_l(\boldsymbol{\theta})}{\tilde{P}_l},\tag{4.9}$$

temos que a derivada da log-verossimilhança pode ser re-escrita como

$$\frac{\partial \log(L)}{\partial v_i} = \int_{\boldsymbol{\theta}} \frac{\bar{R}_l - \bar{n} P_i(\boldsymbol{\theta})}{P_i(\boldsymbol{\theta}) [1 - P_i(\boldsymbol{\theta})]} \cdot \frac{\partial P_i(\boldsymbol{\theta})}{\partial v_i} g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \qquad (4.10)$$

ou utilizando-se aproximação para quadratura gaussiana,

$$\frac{\partial \log(L)}{\partial v_i} \approx \sum_{q_p=1}^Q \dots \sum_{q_1=1}^Q \frac{\bar{R}_{i,q_1\dots q_p} - \bar{n}_{q_1\dots q_p} P_i(\boldsymbol{K})}{P_i(\boldsymbol{K})[1 - P_i(\boldsymbol{K})]} \left[\frac{\partial P_i(\boldsymbol{K})}{\partial v_i} A(K_{q_1}) \dots A(K_{q_p}) \right].$$
(4.11)

As equações definidas em (4.8) e (4.9) representam a frequência esperada de acertos do item *i* e o número esperado de indivíduos com niveis dos traços latentes iguais a Kcaracterizando o passo E do algorítmo EM usado na obtenção das estimativas. O passo M é apresentado na equação (4.10) ou na sua aproximação (4.11), que é maximizada utilizando o método de aceleração do algoritmo EM de Varadhan & Roland (2008) implementado em R no pacote BB (Varadhan & Gilbert 2009). O programa em R com tal metodologia implementada encontra-se no Apêndice C.
4.2.2 Método Bayesiano da Moda a Posteriori

No método da Máxima Verossimilhança Marginal supomos uma distribuição de probabilidade para o vetor de traços latentes $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$, como uma premissa razoável e necessária para garantir boas propriedades dos estimadores dos parâmetros associados aos itens e para tornar a metodologia de estimação aplicável na prática. Em uma abordagem bayesiana, assumimos uma distribuição de probabilidade a priori para todos os parâmetros do modelo, assim obtendo do Teorema de Bayes a chamada distribuição de probabilidade a posteriori.

$$\pi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\theta} | \boldsymbol{X}, \boldsymbol{\eta}, \boldsymbol{\tau}) \propto L(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\theta} | \boldsymbol{X}) \pi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\theta} | \boldsymbol{\eta}, \boldsymbol{\tau}), \qquad (4.12)$$

de maneira que $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_I)$ são os vetores de discriminação $\boldsymbol{a}_i = (a_1, \cdots, a_p)$ de cada traço latente para cada item $i = 1, \cdots, I$, $\boldsymbol{b} = (b_1, \cdots, b_I)$ é o vetor de dificuldades para cada item da prova, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n)$ são os vetores de traços latentes $\boldsymbol{\theta}_j = (\theta_1, \cdots, \theta_p)$ associados a cada $j = 1, \cdots, n$ indivíduo da população avaliada e \boldsymbol{X} é a matriz de respostas dos N indivíduos aos I itens com função de verossimilhança L(.) como definida em 4.3, $\pi(.)$ a distribuição de probabilidade a priori assumida para os parâmetros e dependente dos hiperparâmetros $\boldsymbol{\tau}$ para os traços latentes e $\boldsymbol{\eta}$ para os parâmetros dos itens.

Assim como no MVM da subseção anterior, marginalizamos a distribuição (4.12) sobre os valores do vetor de traços latentes $\boldsymbol{\theta}$ e sobre os hiperparâmetros das distribuições dos parâmetros dos itens, assumidos com distribuição $g(\boldsymbol{\eta}_a)$ e $g(\boldsymbol{\eta}_b)$ para os hiperparâmetros da discriminação e dificuldade, respectivamente, obtendo a distribuição de probabilidade

$$\pi(\boldsymbol{a}, \boldsymbol{b} | \boldsymbol{X}) \propto \int \left(\left[\int_{\mathbb{R}^p} L(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\theta} | \boldsymbol{X}) \pi(\boldsymbol{\theta} | \boldsymbol{\tau}) d\boldsymbol{\theta} \right] \pi(\boldsymbol{a} | \boldsymbol{\eta}_{\boldsymbol{a}}) \pi(\boldsymbol{b} | \boldsymbol{\eta}_{\boldsymbol{b}}) \right) g(\boldsymbol{\eta}_{\boldsymbol{a}}) g(\boldsymbol{\eta}_{\boldsymbol{b}}) d\boldsymbol{\eta}_{\boldsymbol{a}} d\boldsymbol{\eta}_{\boldsymbol{b}},$$
(4.13)

onde $\pi(.)$ no lado direito da equação são as distribuições a priori dos parâmetros, assumidas independentes entre si (Baker & Kim 2004).

Tomando a função de verossimilhança de um padrão de respostas $\boldsymbol{x}_s, s = 1, \dots, S$ como definida em (4.4), podemos escrever o logaritmo da distribuição a posteriori condicionada ao vetor de respostas \boldsymbol{x}_s como

$$\log\left(\pi(\boldsymbol{a}, \boldsymbol{b} | \boldsymbol{x}_s)\right) = \widetilde{P}_s + \log(\pi(\boldsymbol{a})) + \log\left(\pi(\boldsymbol{b})\right).$$
(4.14)

Repetindo a modelagem multinomial da função verossimilhança feita em (4.6), analogamente observa-se que adotando as premissas usuais de independência entre parâmetros dos itens, itens e indivíduos, obtemos a expressão para a derivada do logaritmo da distribuição a posteriori com relação a um parâmetro de item v_i , $i = 1, \dots, I$ pela soma da Equação (4.14) para todos os padrões de resposta observados, obtendo

$$\frac{\partial \log \left(\pi(\boldsymbol{a}, \boldsymbol{b} | \boldsymbol{X})\right)}{\partial v_i} = \frac{\partial \log(L)}{\partial v_i} + \frac{\partial \log(\pi(v_i))}{\partial v_i}, \qquad (4.15)$$

 $\pi(v_i)$ é a distribuição a priori do parâmetro do item em questão, L é como definido em (4.6) e a derivada da log-verossimilhança é como definida em (4.10), de maneira que podemos realizar a otimização utilizando o algorítmo EM de maneira completamente análoga, apenas levando em consideração as componentes extras devidas às distribuições a priori.

4.3 Estimação dos Traços Latentes

Supondo conhecidos os parâmetros dos itens, a estimação do vetor de traços latentes pode ser feita resolvendo as equações de verossililhança para $k = 1, \dots, p$

$$\frac{\partial l\left(\boldsymbol{\theta}_{s} | \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{X}\right)}{\partial \theta_{k}} = 0$$

 $l(\boldsymbol{\theta}_s | \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{X})$ é a função log-verossimilhança do s-ésimo padrão de resposta observado, $s = 1, \dots, S$ condicionada à matriz de respostas observadas \boldsymbol{X} e aos parâmetros dos itens $\boldsymbol{a} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_I)$ e $\boldsymbol{b} = (b_1, \dots, b_I)$ obtida tomando-se o logarítmo da função de verossimilhança (4.3).

Porém, a estimação dos traços latentes por máxima verossimilhança envolve a resolução de pS equações não lineares, ou a maximização de S funções de verossimilhança, o que pode ser extremamente custoso.

Uma alternativa é a estimação bayesiana pela por esperança a posteriori (EAP) utilizada por exemplo no programa de *Lee*& *Terry* (2005). Aproveitando os nós de quadratura da estimação por máxima verossimilhança marginal e a premissa de normalidade multivariada do vetor de traços latentes, estima-se a k-ésima componente do vetor de traços latentes, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k, \dots, \theta_p)$ pelo valor esperado da distribuição a posteriori do vetor de traços latentes

$$\hat{\theta}_{ks} = \frac{\int_{\mathbb{R}^{1}} \theta_{k} L_{s}(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\widetilde{P}_{s}}, \qquad (4.16)$$

dado que \widetilde{P}_s é a probabilidade marginal do s-ésimo padrão de resposta definida em (4.4) ou em sua aproximação por quadratura de Gauss-Hermite

$$\theta_{ks} \approx \frac{\sum_{m \in \mathbf{K}} K_{mk} L_s(\mathbf{K}_m) A(\mathbf{K}_m)}{\widetilde{P}_s},\tag{4.17}$$

para a qual usamos as aproximações definidas em (4.5) para o conjunto dos pontos de quadratura p- dimensionais K.

A estimação por EAP tem a vantagem de envolver apenas o cálculo de somas para aproximar as integrais ao invés dos sistemas não-lineares e otimizações do método de máxima verossimilhança e a garantia de atender sempre ao principio da verossimilhança por se tratar de um procedimento bayesiano.

Contudo, em ambas as estimações, escores perfeitos (i.e. padrões de resposta consistindo apenas de acertos) ou nulos (apenas erros) devem ser retirados do processo de estimação, tendo em vista que as estimativas dos traços latentes associados a tais padrões de resposta tendem a $+\infty$ ou $-\infty$ respectivamente.

4.4 Estimação Conjunta: MCMC

O método de simulação MCMC consiste na definição de uma cadeia de Markov estacionária com estados representados por $M_t = \left(\boldsymbol{\zeta}^{(t)}, \boldsymbol{\theta}^{(t)}\right)$, para $t = 1, 2, \cdots$ e $\boldsymbol{\zeta}^{(.)} = (\boldsymbol{a}, \boldsymbol{b}), \boldsymbol{\theta}^{(.)}$ definidos como em (3.3).

Definido um espaço de estados, especifica-se uma probabilidade de transição entre estados, uma função distribuição de probabilidade definindo a probabilidade do t-ésimo passo no processo pertencer a uma região A_t do espaço paramétrico:

$$P\left((\boldsymbol{\zeta}^{(t)}, \boldsymbol{\theta}^{(t)}) \in A_t | (\boldsymbol{\zeta}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}) = M_{t-1}\right).$$
(4.18)

A função densidade de probabilidade $p(M_t|M_{t-1})$ obtida diferenciando (4.18) é denominada núcleo de transição da cadeia.

Em posse dessa função, é possivel obter funções de densidade de probabilidade para os valores do espaço paramétrico em um determinado passo $t(M_t)$ em função da distribuição do estado anterior por uma aplicação das equações de Chapman-Kolmogorov

$$\pi^{(t)}(.) = \int_{\boldsymbol{\zeta},\boldsymbol{\theta}} p(.|s) \pi^{(t-1)}(s) ds.$$
(4.19)

Se a cadeia definida pelo espaço de estados e probabilidades de transição atendem a algumas condições de regularidade (Robert & Casella 1999), a distribuição de M_t irá convergir para a distribuição estacionária da cadeia.

Para obterem-se estimativas dos parâmetros de interesse, construimos núcleos de transição que convenientemente produzam cadeias tendo a distribuição a *posteriori* de interesse como sua distribuição estacionária. Assim, a partir de um certo momento t^* , os elementos M_t podem ser considerados amostras aleatórias da distribuição a *posteriori* em questão.

O valor de t^* é escolhido a partir da convergência da cadeia. As observações geradas antes da iteração t^* , ou seja, antes da convergência são descartadas e esse período é denominado "burn in". As demais observações são usadas para realizar inferências e obter estimativas dos parâmetros de interesse. Diferentes métodos para determinar o número de iterações a serem descartadas e a convergência da cadeia são propostos na literatura e serão discutidos posteriormente nesse capítulo.

Para obter amostras dos parâmetros, utilizamos o método do Amostrador de Gibbs (Geman & Geman 1984) que obtém amostras da distribuição estacionária da cadeia utilizando o procedimento iterativo:

- 1. Simulação de um valor inicial para a cadeia $(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)})$
- 2. Simulação da observação a amostra $\theta_{jk}^{(1)}$ da distribuição condicional completa $\pi(\theta|\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)})$ obtida a partir da distribuição estacionária $\pi(.)$ para todo j, k em $j = 1, \dots, N$ e $k = 1, \dots, p$.
- 3. Simulação de $b_i^{(1)}$ da distribuição $\pi(b|\boldsymbol{a}^0, \boldsymbol{b}^0, \boldsymbol{\theta}^{(1)})$ para todo $i = 1, \cdots, I$
- 4. Simulação de $a_{ik}^{(1)}$ da distribuição $\pi(a|\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(1)}, \boldsymbol{\theta}^{(1)})$ para todo $i = 1, \cdots, I, k = 1, \cdots, p$
- 5. Utilizando o ponto $(\boldsymbol{a}^{(1)}, \boldsymbol{b}^{(1)}, \boldsymbol{\theta}^{(1)})$ no passo 1, repita o procedimento.

Entretanto, obter as distribuições condicionais de cada parâmetro não é uma tarefa trivial, e frequentemente tais distribuições não tem forma analítica fechada (como no caso do modelo do modelo logístico). Para isso, utilizamos o algorítmo de Metropolis-Hastings nos passos 2 a 4 do amostrador de Gibbs para obter amostras das distribuições desejadas.

O algoritmo de Metrópolis-Hastings consiste em tomar um núcleo de transição mais conveniente $q(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{a}^{(1)}, \boldsymbol{b}^{(1)}, \boldsymbol{\theta}^{(1)})$, gerar uma observação dessa distribuição $(\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{\theta}^*)$ e calcular a probabilidade de aceitação do valor gerado como o próximo passo da cadeia, dada por

$$\alpha(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{a}^{(*)}, \boldsymbol{b}^{(*)}, \boldsymbol{\theta}^{(*)}) = \min\left\{\frac{\pi(\boldsymbol{a}^{(*)}, \boldsymbol{b}^{(*)}, \boldsymbol{\theta}^{(*)})q(\boldsymbol{a}^{*}, \boldsymbol{b}^{*}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)})}{\pi(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)})q(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{a}^{(*)}, \boldsymbol{b}^{(*)}, \boldsymbol{\theta}^{(*)})}, 1\right\}.$$
(4.20)

Um valor aleatório U de uma distribuição uniforme no intervalo (0, 1) é gerado e, se $U \leq \alpha$, o valor da observação é atualizado como $(\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{\theta}^*)$.

Existem modificações do algorítmo de Metrópolis-Hastings focando na escolha da distribuição q(.) e na obtenção de boas propriedades da cadeia como *mixing*, onde os valores amostrados preenchem o conjunto dos valores que podem assumir e menor

tempo computacional por obter maiores probabilidades de aceitação. Tais métodos estão expostos e detalhadamente discutidos em Robert & Casella (1999).

Um problema da estimação por MCMC é a alta correlação entre as observações, obrigando a adoção de modificações nos algoritmos de amostragem como a amostragem em blocos de certos conjuntos de parâmetros, como o vetor de discriminações e de traços latentes. Ambos os problemas são tratados em Patz & Junker (1999).

Existem ainda algumas questões envolvendo a qualidade das estimativas obtidas pelos métodos MCMC e propriedades desejaveis da cadeia de Markov, as quais discutimos na seção seguinte.

4.4.1 Diagnóstico de Convergência do Método MCMC

As amostras desejáveis para as estimativas são obtidas da distribuição estacionária (??), que construimos de maneira que seja a distribuição *a posteriori* (4.12). Contudo, a distribuição estacionária é obtida após um número suficientemente grande de iterações da cadeia (esse período de iterações, nos quais a cadeia converge à distribuição estacionária é denominado *burn-in*), as quais devemos descartar no cálculo dos estimadores pontuais dos parâmetros.

Existem diversas maneiras de verificar a convergência. Abordaremos nessa seção os critérios de Geweke (1992) e Gelman-Rubin (Gelman & Rubin 1992) que permitem inferir convergência baseados apenas nas amostras. A determinação do período de *burn-in* e do intervalo entre iterações que minimize a autocorrelação é feita pelo critério de Raftery-Lewis (Raftery & Lewis 1995).

Simplificaremos a notação ao longo dos seguintes parágrafos utilizando ζ_t para denotar uma amostra de um parâmetro do modelo obtida de algum dos métodos do tipo MCMC discutidos anteriormente.

O critério de Raftery-Lewis se baseia no processo de estimação de um quantil fixado $q = P(\pi(\zeta_t) \leq u)$ da distribuição a posteriori com um erro r e probabilidade de cobertura s. Constroi-se a sequência de variáveis aleatórias

$$Z_t = \begin{cases} 1 & se \ \zeta_t \le u \\ 0 & c.c. \end{cases}$$
(4.21)

sobre as quais obtém sub-sequências, para k > 1

$$Z_t^{(k)} = Z_{1+(t-1)k}, (4.22)$$

que nada mais são do que valores da sequência (4.21) espaçados de k valores.

Raftery e Lewis então assumem que a autocorrelação entre valores de (4.22) decresce na medida que k aumenta e que para valores suficientemente grandes de k, a sequencia se comporta como uma cadeia de Markov. Assim, essas variáveis aleatórias são ajustadas como sendo advindas de uma cadeia de Markov de primeira ordem (o próximo valor da sequência depende apenas do valor anterior), de uma cadeia de Markov de segunda ordem (o valor seguinte é determinado pelos últimos dois valores), e o modelo é escolhido utilizando algum método de seleção de modelos. O salto k entre variáveis é então determinado como sendo o menor k em (4.22) para o qual o modelo de primeira ordem é escolhido.

O período de *burn-in* é então determinado pelo número de iterações necessárias para que se obtenham valores suficientemente próximos da distribuição estacionária da cadeia de Markov construida, mais simples de obter. Tal número, contudo, costuma ser excessivamente pequeno, sendo utilizada frequentemente na literatura uma regra prática de descartar o primeiro 1% das iterações como *burn-in*. O método de Raftery-Lewis está implementado no pacote **coda** do software estatístico **R** o qual assume como padrão o quantil q = 0.025, um erro de r = 0.125 e s = 0.95 de probabilidade de cobertura.

Um outro critério para verificar a convergência da cadeia baseado nas amostras é o de Geweke (Geweke 1992). Tomam-se partes da cadeia, $n_a \in n_b$, frequentemente os primeiros 10% das iterações após o *burn-in* m e os últimos 50% das iterações e calculam-se as médias ergódicas para cada parâmetro amostrado ζ

$$\overline{\zeta}_{a} = \sum_{i=m}^{m+n_{a}} \zeta_{i}$$

$$\overline{\zeta}_{b} = \sum_{i=N^{*}-n_{b}}^{N^{*}+n_{b}} \zeta_{i},$$
(4.23)

para as quais calculamos o valor

$$z_G = \frac{\overline{\zeta}_a - \overline{\zeta}_b}{\sqrt{\widehat{\sigma}_a - \widehat{\sigma}_b}},\tag{4.24}$$

 $\hat{\sigma}_a$, $\hat{\sigma}_b$ são as variâncias amostrais calculadas para os n_a e n_b elementos da amostra. O valor z_G tende em distribuição para uma distribuição normal padrão, cujos valores utilizamos para avaliar a significância de z_G . Frequentemente, supôe-se convergência da cadeia para valores de z_G entre -1.96 e 1.96.

Tanto o critério de Geweke quanto o de Raftery-Lewis tentam observar a convergência para a distribuição estacionária (e portanto, a procedência das amostras utilizadas na estimação dos parâmetros) baseados em amostras advindas de uma única cadeia. O critério de Gelman-Rubin (Gelman & Rubin 1992) supôe que após a convergência para a distribuição de intresse, as amostras obtidas em cadeias diferentes não devem ter diferença significativa, e propôe um índice que permite observar a diferença baseado em uma análise da variância.

Suponha que obtemos amostras para um parâmetro ζ do modelo simulando C cadeias de Markov, resultando em C sequências de amostras $\{\zeta^{(c)}\}_{t=1,\dots,N^*}, c = 1,\dots,C$.

A variância entre cadeias é calculada por

$$B = \frac{C}{N^* - 1} \sum_{c=1}^{C} \left(\overline{\zeta}^{(c)} - \overline{\zeta}\right)^2, \qquad (4.25)$$

 $\overline{\zeta}^{(c)}$ é a média amostral dos elementos da cadeia c, e $\overline{\zeta}$ é a média amostral de todos os valores amostrados.

Calcula-se também a variância entre as amostras de cada cadeia por

$$W = \frac{1}{C(N^* - 1)} \sum_{c=1}^{C} \sum_{N^*}^{t=1} \left(\zeta_t^{(c)} - \overline{\zeta}^{(c)} \right)^2.$$
(4.26)

Então, estima-se a variância do parâmetro ζ por

$$\widehat{\sigma}_{\zeta} = \left(1 - \frac{1}{N^*}\right)W + \frac{1}{N^*}B.$$
(4.27)

Se as cadeias tiverem convergido, as equações (4.25), (4.26) e (4.27) serão bons estimadores para a variância de ζ . Entretanto, se isso não ocorre, (4.27) tenderá a superestimar o valor da variância, enquanto (4.26) o subestimará. Gelman e Rubin utilizam o valor

$$\widehat{R} = \sqrt{\frac{\widehat{\sigma}_{\zeta}}{W}},\tag{4.28}$$

que é sempre maior do que 1, mas que tende a 1 na medida que $N^* \to \infty$. Avalia-se então a convergência das cadeias pela proximidade de (4.28) a 1, sendo sugerido pelos autores valores abaixo de 1.2 como sugestivos de convergência.

4.5 Adequação do Modelo

Diversas modificações da estatística qui-quadrado são utilizadas para verificar o ajuste do modelo aos dados e sua dimensionalidade. Bock et al. (1988) utilizam a estatística

$$G^{2} = 2\sum_{l=1}^{S} r_{l} \log\left(\frac{r_{l}}{N\widetilde{P}_{l}}\right), \qquad (4.29)$$

soma essa efetuada sobre os S padrões de resposta observados utilizando o número de cada padrão de resposta r_l e a probabilidade marginal \tilde{P}_l de ocorrência do l-ésimo padrão de resposta.

Tal estatística tem distribuição qui-quadrado com $2^Q - Q(p+1) + \frac{p(p-1)}{2}$ graus de liberdade e testa o ajuste do modelo contra a hipótese nula de que um modelo multinomial com probabilidades iguais para todos os itens ajusta melhor os dados.

Contudo, como destacado em Bock et al. (1988) o cálculo dessa estatística requer um número de individuos maior que os 2^Q padrões de resposta possíveis para valores confiáveis.

Bock & Schilling (2005) utilizam a estatística

$$-2\hat{l} = 2\sum_{s=1}^{S} r_s \log\left(\widetilde{P}(\boldsymbol{x}_s|\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})\right)$$
(4.30)

testar a dimensionalidade do modelo utilizando os valores obtidos por MVM $\hat{a} \in \boldsymbol{b}$. Fazendo a diferença entre as para dois modelos com $p \in p + 1$ dimensões do vetor de traços latentes obtém-se uma estatistica para testar a hipótese de que um modelo com p traços latentes ajusta bem os dados, contra a alternativa de que p + 1 traços latentes são necessários. Bock & Schilling (2005) sugere que ajustes sucessivos sejam realizados aumentando-se a dimensão até que a hipótese nula seja aceita.

Kang & Cohen (2007) discutem outras estatísticas para a determinação da dimensionalidade do modelo, como o Akaike Information Criterion(AIC)

$$AIC_{Modelo} = -2\hat{l} + 2n, \tag{4.31}$$

-2l é como definida em (4.30) e n é o número de parâmetros do modelo, e o *Bayesian* Information Criterion(BIC) definido por

$$BIC_{Modelo} = -2\hat{l} + n\log(N), \tag{4.32}$$

sendo N o número de individuos na amostra. Em ambos os critérios, o modelo com o menor índice é selecionado como mais adequado.

Contudo, todas as estatisticas até agora definidas dependem de valores para a função de verossimilhança da amostra obtidos na convergência do algorítimo EM, o que nem sempre ocorre tendo em vista que frequentemente interrompe-se o processo iterativo quando observam-se mudanças pequenas o suficiente nos valores dos parâmetros, o que não necessariamente implica em mudanças insignificantes no valor da função verossimilhança. Kang & Cohen (2007) também ressaltam que os valores de AIC e BIC nem sempre selecionam o mesmo modelo, sendo o BIC mais inclinado a escolher modelos com menos parâmetros, mesmo erroneamente.

A premissa de independência local é verificada verificando-se a matriz de covariâncias entre os itens (Linden & Hambleton 1996)

$$cov_{i,k} = \frac{\sum_{j=1}^{N} (x_{ij} - P_i(\boldsymbol{\theta}_j))(x_{kj} - P_k(\boldsymbol{\theta}_j))}{N}, \ i, k = 1, \cdots, Q$$
(4.33)

de maneira que resíduos grandes indicam dependência entre os itens.

Uma vantagem da TRI é a de que itens podem ser estimados e avaliados separadamente, dessa maneira, podemos avaliar o ajuste dos diversos itens ao modelo proposto.

A maneira tradicionalmente feita para modelos unidimensionais é a estimação dos parâmetros dos itens e dos traços latentes, obtendo a curva característica do item (CCI) e compara-la com as proporções de acerto ao item para cada agrupamento de valores próximos do traço latente.



Figura 4.1: Ajuste Gráfico pela CCI: os pontos em azul representam os valores estimados, enquanto os valores em rosa representam os valores observados. Fonte: IRT Lab, University of Illinois-Urbana

Contudo, os modelos multidimensionais formam superfícies que não podem ser avaliadas graficamente, sendo necessária uma verificação analítica do ajuste.

O programa BILOG, após estimar os parâmetros dos itens, estimativas dos traços latentes são computadas e classificadas em $c = 1, \dots, C$ categorias e a estatística

$$2\sum_{c=1}^{C} \left[p_c \log\left(\frac{p_c}{N_c P(\boldsymbol{\theta}_c)}\right) + (N_c - p_c) \log\left(\frac{p_c}{N_c}(1 - P(\boldsymbol{\theta}_c))\right) \right]$$
(4.34)

é calculada para cada item tomando-se p_c a proporção de acertos na categoria c, o número de respondentes na categoria, N_c e a probabilidade de resposta ao item considerando-se o nível dos traços latentes $\boldsymbol{\theta}_c$ em torno do qual a categoria foi feita. A estatística (4.34) tem uma distribuição qui-quadrado com C graus de liberdade (Embretson & Reise 2000).

Como descrito em Masters & Wright (1996), podemos avaliar o ajuste dos indivíduos e dos itens utilizando resíduos padronizados. Para cada resposta X_{si} , $s = 1, \dots, S$, $i = 1, \dots, Q$ no *s*-ésimo padrão de resposta ao *i*-ésimo item, o valor esperado da resposta é

$$E[X_{si}] = P_i(\boldsymbol{\theta}_s), \qquad (4.35)$$

simplesmente a probabilidade de acerto ao item, pois a variável aleatória segue uma distribuição Bernoulli. Analogamente, temos a variância da variável aleatória X_{si}

$$V[X_{si}] = \sum_{k=0}^{1} (k - X_{si})^2 P_i(\boldsymbol{\theta}_s), \qquad (4.36)$$

com os quais calculamos o resíduo padronizado

$$z_{si} = \frac{X_{si} - E\left[X_{si}\right]}{\sqrt{V\left[X_{si}\right]}}.$$
(4.37)

Definem-se então estatísticas baseadas na média desses resíduos para avaliar o ajuste de determinado item fazendo-se a média sobre o número de padrões de resposta observados ou sobre o número de itens. Ou seja, denotando-se por ϵ_s o ajuste do *s*-ésimo indivíduo ao modelo e ϵ_i o ajuste do *i*-ésimo item, fazemos

$$\epsilon_{i} = \sum_{s=1}^{S} \frac{z_{si}^{2}}{S}$$

$$\epsilon_{s} = \sum_{i=1}^{Q} \frac{z_{si}^{2}}{Q}$$

$$(4.38)$$

que utilizamos para verificar o ajuste de cada item ou de cada indivíduo ao modelo ajustado. Estudos recentes, contudo, apontam uma correlação positiva entre o valor das estatísticas (4.38) e o valor dos parâmetros de discriminação dos itens (Sinharay & Lu 2008), tornando quaisquer conclusões advindas desses resíduos menos confiável.

Béguin & Glas (2001) avaliam a performance dos modelos ajustados pelo MVM utilizando a frequência esperada de um escore $r = 0, \dots, I$ dada pela equação

$$f(r) = N \sum_{\boldsymbol{x}_s|r} \int_{\mathbb{R}^p} L_s(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \qquad (4.39)$$

 $\boldsymbol{x}_s | r$ é o conjunto dos padrôes de resposta que resultam em um escore r e L_s é definida como em (4.3).

Orlando & Thissen (2000) estudam uma estatística qui-quadrado para avaliar o impacto no ajuste do modelo de uma eventual falta de monotonicidade (i.e. um aumento no valor de um traço latente implica em um aumento na probabilidade de resposta correta ao item) e da falta de um parâmetro de acerto casual no modelo, estatística essa generalizada para modelos multidimensionais por Zhang & Stone (2008).

Para cada escore $K = 0, \cdots, Q$, define-se

$$S_k = \sum_{s|K=k} L_s(\boldsymbol{\theta}), \tag{4.40}$$

a verossimilhança do k-ésimo escore como a soma de todos os padrões de resposta que resultam no escore $k = 0, \dots, Q$. Para testes com um número razoável de itens (o BDI por exemplo apresenta mais de 2 milhões de padrões de resposta possiveis) tal calculo é inviável, sendo utilizado um algorítmo iterativo proposto em Orlando & Thissen (2000) e implementado em R disponível no Apêndice.

Denotando então por S_k^{-i} a verossimilhança do escore k excluindo-se o item i do teste, calcula-se a proporção esperada de respostas corretas ao item i entre os individuos de escore k como

$$E_{ik} = \frac{\int_{\mathbb{R}^p} P_i S_{k-1}^{-i}(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^p} S_k(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$
(4.41)

Integral essa que, como as outras integrais utilizadas nesse trabalho, são calculadas utilizando quadratura de Gauss-Hermite com 7 pontos em cada dimensão do vetor de traços latentes. Sendo O_{ik} a frequência observada de acertos ao item *i* entre os individuos de escore *k*, calcula-se a estatística

$$s - \chi^2 = \sum_{k=1}^{Q-1} \frac{N_k (O_{ik} - E_{ik})^2}{E_{ik} (1 - E_{ik})},$$
(4.42)

que sob a hipótese nula de que a proporção observada é próxima da predita pelo modelo tem distribuição qui-quadrado com Q - m + 1 graus de liberdade, m = p + 1o número de parâmetros do item do modelo para o modelo multidimensional de dois parâmetros.

Como observado por Zhang & Stone (2008), a estatística (4.42) pode ser usada para verificar violações nas premissas de monotonicidade e de ausência de acerto casual, sendo mais eficiente na discriminação da primeira, com a maior complexidade computacional em seu cálculo compensada pela ausência de correlação entre os valores da estatística e parâmetros do modelo (Sinharay & Lu 2008). Todas as estatísticas discutidas até o presente ponto verificam o ajuste do modelo segundo um ponto de vista frequentista, desenvolvendo uma estatística de teste $T(\mathbf{X})$ que permita observar se os dados apresentam o comportamento previsto pelo modelo.

Ajustando o modelo por métodos bayesianos como os discutidos anteriormente, temos também maneira de verificar o ajuste dos modelos empregados na sua capacidade de prever características do conjunto de dados e obter índices que permitam escolher modelos.

Checagens preditivas da posteriori (Gelman, Carlin & Rubin 2004) permitem generalizar o contexto de p-valor para dados ajustados segundo métodos bayesianos. Escolhendo uma estatística de teste $T(\mathbf{X})$ simulam-se R replicações do conjunto de dados \mathbf{X}_r utilizando-se as R amostras obtidas pelo método MCMC e calcula-se um sumário da estatística de teste, quantis e o p-valor bayesiano pela aproximação

$$p - valor \approx \frac{1}{R} \sum_{r=1}^{R} I(T(\boldsymbol{X}_r \ge T(\boldsymbol{X}))).$$
(4.43)

I(.) é a função indicadora, tendo como valor 1 se a condição explicitada em seu argumento ocorre ou 0 caso contrário. O p-valor indica irregularidades com a propriedade avaliada do modelo quando apresenta valores extremos, próximos de 0 ou 1.

Béguin & Glas (2001) por exemplo utiliza checagens preditivas da distribuição de escores preditos pelo modelo para verificar o ajuste geral do modelo aos dados, e calcula o p-valor utilizando como estatística de teste

$$T(\boldsymbol{X}_{r}) = \sum_{k=0}^{Q} \frac{(N_{k}^{(r)} - f^{(r)}(k))^{2}}{f^{(r)}(k)},$$
(4.44)

de maneira que $N_k^{(r)}$ é a frequencia observada do escore $k = 0, \dots, Q$ na repetição $r = 1, \dots, R$ e $f^{(r)}(k)$ é a frequencia esperada calculada em (4.39) utilizando-se os parâmetros amostrados na r- ésima iteração do amostrador utilizado. Tal método foi utilizado no conjunto de dados reais para o qual ajustamos os modelos no Capítulo 5.

Um índice que pode ser utilizado na escolha dos modelos utilizando as amostras obtidas por MCMC é o DIC (Gelman et al. 2004), calculado por

$$DIC = \overline{D} + p(\boldsymbol{\theta}, \boldsymbol{\zeta}), \qquad (4.45)$$

de maneira que \overline{D} é a esperança da deviance, calculada para cada amostra por $D \propto -2\log(L(\boldsymbol{X}_r)|\boldsymbol{\theta}^{(r)},\boldsymbol{\zeta}^{(r)}), r = 1, \cdots, R$ e $p(\boldsymbol{\theta},\boldsymbol{\zeta})$ é uma penalidade pelo número de parâmetros, definida por $p(\boldsymbol{\theta},\boldsymbol{\zeta}) = \overline{D} - \widehat{D}, \widehat{D}$ sendo a deviance calculada utilizando a esperança dos parâmetros do modelo.

Analogamente ao BIC e ao AIC expostos anteriormente, o modelo com o menor DIC é escolhido como o modelo que melhor se ajusta ao conjunto de dados. A vantagem do DIC com relação aos outros critérios expostos é que o DIC não depende dos valores dos estimadores de máxima verossimilhança e as esperanças são bem aproximadas pelas médias amostrais dos valores obtidos no método MCMC.

Capítulo

Simulação

Após a implementação dos métodos expostos no Capítulo 4, conduzimos simulações com o intuito de estudar e comparar as propriedades dos estimadores propostos.

5.1 Simulação 1

Simulamos uma amostra de n = 1000 valores da distribuição normal bivariada, com vetor de médias nulo e matriz de variâncias igual à identidade. Valores para os parâmetros de I = 10 itens foram fixados conforme na Tabela 5.1 (coluna 'Real'), de maneira a obter 10 pontos igualmente espaçados para valores do parâmetro de dificuldade entre -3 e 3 e parâmetros de discriminação entre 0.5 e 2.5, contemplando itens muito ou pouco discriminativos em ambas as dimensões e em dificuldades tanto altas quanto baixas.

Utilizamos o modelo logístico mulditimensional de dois parâmetros ((3.3) com $c_i = 0, i = 1, \dots, 10$) e p = 2 para gerar as respostas dicotômicas (X_{ij}) .

Os parâmetros do modelo foram estimados utilizando o método de MVM implementado em R assumindo o modelo logístico (3.3), o programa TESTFACT assumindo o modelo de ogiva normal (3.2) utilizando o MVM exposto em Bock et al. (1988) e o método MCMC implementado no WinBUGS com o modelo logístico, assumindo as distribuições *a priori* para os parâmetros de acordo com o sugerido por Patz & Junker (1999)

$$\begin{aligned} a_{ik} &\sim log - normal(1, 0.5) \\ b_i &\sim N(0, 1) \\ \boldsymbol{\theta}_j &\sim N_p(\mathbf{0}, I_p), \end{aligned}$$
(5.1)

para $i = 1, \dots, 10, k = 1, \dots, 2$ e $j = 1, \dots, 1000$. A notação N(.) denota a distribuição normal p-variada, I_p denota a matriz identidade de ordem $p \in \mathbf{0}$ denota o vetor com p componentes nulas.

As estimativas obtidas para os parâmetros dos itens encontram-se na tabela 5.1 nas respectivas colunas e representados graficamente na figura 5.1.

Os valores σ_{MCMC} correspondem aos erros padrão do estimador, obtido pela raiz quadrada da variância das observações amostradas por Monte Carlo. O parâmetro $\hat{\sigma}_{MVM}$ por sua vez é obtido tomando a raiz quadrada dos elementos da diagonal principal do inverso da matriz de informação de Fisher de coeficientes (Li & Lissitz 2000)

$$a_{kj} = \int_{\mathbb{R}^p} \left(\frac{\partial P_i(\boldsymbol{\theta})}{\partial \zeta_k} \right) \left(\frac{\partial P_i(\boldsymbol{\theta})}{\partial \zeta_j} \right) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
(5.2)

em que ζ_k , $k = 1, \dots, p+1$ é um parâmetro do item $i, i = 1, \dots, I$ e $k, j = 1, \dots, p+1$. O TESTFACT não tem em sua saída os erros padrão dos estimadores.

Para a estimação por MCMC, simulamos duas cadeias de tamanho 300.000 e o método de diagnóstico de Raftery-Lewis (Raftery & Lewis 1995) para definir os períodos de *burn-in* e intervalo entre iterações da cadeia.

Escolhemos descartar as primeiras 3.000 iterações, seguindo a tendência na literatura de descartar 1% das iterações, bem acima dos 500 indicados por Raftery-Lewis. O salto sugerindo pelo critério foi de 40 a 50 iterações para a maioria dos parâmetros. Escolhemos 50 como o salto visando minimizar a autocorrelação e obtendo amostras de tamanho efetivo entre 4.000 e 6.000, para os parâmetros de discriminação e entre 4 e 10 mil, para os parâmetros de dificuldade.

Verificamos a convergência da cadeia utilizando os métodos de diagnóstico de Geweke (Geweke 1992), que retornou convergência a 95% de confiança para todos os parâmetros menos um, e o critério de Gelman-Rubin (Gelman & Rubin 1992) com um valor próximo de 1 para todos os parâmetros, indicando convergência da cadeia.

Pode-se também verificar graficamente que as duas cadeias apresentaram-se próximas e houve convergência para o mesmo valor pelos históricos e densidades, ilustrados na figura 5.2 para o Item 1.

Observamos dessa maneira que ambos os métodos apresentam uma boa recuperação dos parâmetros do modelo, tendo precisões comparáveis em termos dos erros padrão e

Item			a_1			
	Real	MVM	$\hat{\sigma}_{MVM}$	TF	MCMC	$\hat{\sigma}_{MCMC}$
1	0,500	0.363	0,115	0.692	0,661	0,347
2	0,722	0.229	0,109	1.031	0,553	0,387
3	0,944	0.978	0,110	1.032	1,184	0,266
4	1,167	1.086	0,102	1.235	1,253	0,205
5	1,389	1.246	0,102	1.316	1,405	0,176
6	1,611	1.568	0,118	1.487	1,742	0,215
7	1,833	2.082	0,146	1.833	2,256	0,217
8	2,056	2.273	0,161	1.939	2,419	0,219
9	2,278	2.015	0,148	2.215	2,126	0,190
10	2,500	3.639	0,313	2.076	2,995	0,185
Item			a_2			
	Real	MVM	$\hat{\sigma}_{MVM}$	TF	MCMC	$\hat{\sigma}_{MCMC}$
1	2,500	2.794	0,218	1.663	2,662	0,202
2	2,278	2.870	0,206	2.558	2,796	0,191
3	2,056	2.276	0,158	1.931	2,293	0,208
4	1,833	1.814	0,126	1.707	1,790	0,179
5	1,611	1.546	0,112	1.122	1,512	$0,\!175$
6	1,389	1.762	0,125	0.846	1,710	0,205
7	1,167	1.497	0,122	0.640	1,420	0,254
8	0,944	1.272	0,121	0.253	$1,\!173$	$0,\!285$
9	0,722	1.007	0,111	0.391	0,919	0,246
10	0,500	0.490	0,128	0.128	0,482	0,438
Item			b			
	Real	MVM	$\hat{\sigma}_{MVM}$	TF	MCMC	$\hat{\sigma}_{MCMC}$
1	-3,000	-3.095	0,211	-2.643	-2,964	0,28
2	-2,333	-2.701	$0,\!185$	-3.055	-2,606	$0,\!28$
3	-1,667	-1.803	0,127	-2.203	-1,825	$0,\!172$
4	-1,000	-0.903	0,093	-1.209	-0,916	0,118
5	-0,333	-0.297	0,082	-0.530	-0,312	0,105
6	0,333	0.346	0,087	0.190	0,323	0,116
7	1,000	1.237	0,108	0.975	1,210	0,146
8	1,667	1.909	0,133	1.616	1,872	0,184
9	2,333	2.028	0,133	2.618	1,996	0,172
10	3,000	3.877	0,307	2.927	3,188	0,352

Tabela 5.1: Estimativas dos parâmetros dos itens utilizando o MVM, o programa TESTFACT (TF), o Amostrador de Gibbs (MCMC) e respectivos erros padrão ($\hat{\sigma}$), parâmetros de discriminação (a_1, a_2) e dificuldade (b).



Resíduos Absolutos

Figura 5.1: Valores absolutos da diferença entre os valores reais e os ajustados pelos métodos MVM e MCMC, os pontos representam os parâmetros de discriminação, enquanto os triângulos representam os parâmetros de dificuldade

das diferenças absolutas com relação aos parâmetros reais, sendo então recomendável a estimação por MVM para o modelo em questão, tendo em vista a diferença em tempos computacionais (menos de 2 minutos para o MVM e algumas horas para o método MCMC).

Avaliamos então o MVM com relação a existência de viés nas estimativas em aplitude do erro quadrático médio em diversas configurações de amostra na seção seguinte.



Figura 5.2: Históricos dos valores amostrados no método MCMC e densidades de probabilidade *a posteriori* estimadas para os parâmetros do Item 1

5.2 Simulação 2

Para avaliar a precisão do MVM simulamos R = 1000 amostras utilizando o modelo (3.3) e os parâmetros da Tabela 5.2 para diversos tamanhos de teste (I = 10, 25), de amostra (N = 500, 1000, 5000, 10000) e dimensões do vetor de traços latentes (p = 1, 2, 3). Calculamos o Erro Quadrático Médio (EQM) dos estimadores de cada parâmetro ζ ,

$$EQM(\zeta) = \frac{1}{R} \sum_{r=1}^{R} \left(\widehat{\zeta} - \zeta^*\right)^2, \qquad (5.3)$$

onde ζ é um parâmetro do item (componente do vetor de discriminações ou dificuldade), $\hat{\zeta}$ é o estimador de MVM e ζ^* é o valor real do parâmetro.

Calculamos também o viés dos estimadores,

$$B(\zeta) = \frac{1}{R} \sum_{r=1}^{R} \left(\widehat{\zeta} - \zeta^*\right).$$
(5.4)

Tabela 5.2: Parâmetros utilizados nas simulações

Item	a_1	a_2	a_3	b
1	$1,\!094$	0,998	$1,\!254$	-0,141
2	$1,\!316$	0,901	$2,\!497$	0,160
3	$0,\!994$	$0,\!810$	$2,\!140$	-0,659
4	$0,\!877$	$0,\!834$	$0,\!627$	-0,145
5	0,926	0,722	$0,\!618$	-0,127
6	$1,\!458$	0,989	$0,\!475$	-0,063
7	$1,\!676$	$0,\!936$	1,968	-0,664
8	1,202	$1,\!228$	0,728	$0,\!664$
9	0,911	1,033	1,749	0,142
10	$1,\!237$	0,747	1,072	0,007
11	0,953	$1,\!394$	0,799	-0,001
12	0,827	$1,\!203$	1,852	-0,159
13	1,076	1,096	$0,\!994$	0,223
14	$1,\!007$	0,904	$0,\!557$	$0,\!259$
15	$1,\!334$	1,034	$0,\!667$	-0,709
16	$1,\!103$	1,269	$1,\!644$	0,346
17	0,753	1,016	1,076	-0,082
18	0,977	1,098	$0,\!377$	$1,\!178$
19	$1,\!302$	$1,\!627$	0,916	0,009
20	$1,\!241$	$1,\!113$	$1,\!477$	$0,\!150$
21	$1,\!199$	$1,\!122$	$0,\!978$	$0,\!452$
22	0,786	1,072	$1,\!298$	0,162
23	0,993	0,726	$2,\!442$	-0,391
24	0,961	1,565	0,563	$0,\!617$
25	0,909	0,869	1,313	0,297

Devido a quantidade de parâmetros estimados (I(p+1) parâmetros estimados para cada configuração), utilizamos a média dos viéses e dos EQM para cada tipo parâmetro estimado sobre o número de itens. Os resultados para o viés encontram-se na Tabela 5.3 e na Figura 5.3.

				+ A				
	ø	<i>l</i> ₁					2	
tra	10 itens	25 itens					10 itens	25 itens
500	-0,037	-0,141					0,006	-0,012
000	-0,049	-0,147					0,005	-0,010
000	-0.054	-0,153					0,006	-0,010
000	-0,054	-0,154					0,007	-0,009
				p = 2				
	a	l1	a	2			1	
stra	10 itens	25 itens	10 itens	25 itens			10 itens	25 itens
500	-0,168	0,033	-0,185	-0,138			-0,003	-0,029
000	0,029	0,038	0,005	-0,174			0,009	-0,027
000	0,015	0,071	-0,018	-0,220			0,004	-0,021
000	0,010	0,079	-0,018	-0,227			0,003	-0,019
				p = 3				
	o	<i>l</i> ₁	a	2	a	ç	1	
stra	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens
500	0,163	0,265	0,162	0,217	-0,157	-0.573	0,015	-0,048
000	0,148	0,241	0,136	0,214	-0,304	-0,630	0,009	-0,048
000	0,184	$0,\!227$	0,120	0,223	-0,691	-0.678	0,012	-0,048
000	0.103	0.990	0 116	0.221	-0.801	-0.681	0.019	-0.048



Figura 5.3: Média dos viéses para as amostras simuladas nas diferentes configurações. Os círculos representam os resultados para 10 itens, enquanto os triângulos representam os de 25 itens.

Podemos observar que o parâmetro de dificuldade b é sempre estimado com um viés insignificante, enquanto o vetor de discriminações a é estimado com pouco viés apenas no caso unidimensional. Quando adotamos mais dimensões para o vetor de traços latentes, a primeira componente de a tende a ser positiva em média, enquanto as demais componentes são negativas, situação essa que piora na medida que aumentamos o número de itens.

O viés do estimador para um parâmetro de item pode ser negativo para um item do teste e positivo para outro, tornando a média dos viéses pequena. Calculamos o máximo e o mínimo dos viéses representados nas tabelas 5.4 e 5.5, nas quais podemos observar que os viéses são pequenos com p = 1, mas crescem na medida que o número de dimensões aumenta.

-	Tabela 5.4	1: Máximo o	dos viéses p	oara cada cc	nfiguração	de amostra	simulada	
				p = 1				
		g					p	
Amostra	10 itens	25 itens					10 itens	25 itens
500	-0,018	-0,090					0,012	-0,006
1000	-0,026	-0,092					0,010	-0,003
5000	-0,030	-0,096					0,016	-0,001
10000	-0,030	-0,096					0,021	0,001
				p=2				
	a	<i>l</i> 1	ø	¹			9	
Amostra	10 itens	25 itens	10 itens	25 itens			10 itens	25 itens
500	-0,137	0,487	0,159	0,394			0,050	0,006
1000	0,145	0,563	0,253	0,414			0,043	-0,013
5000	0,167	0,664	0,132	0,407			0,006	-0,011
10000	0,167	0,673	0,127	0,399			0,004	-0,008
				p = 3				
	Ø	l_1	o	l_2	a	3	q	
Amostra	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens
500	0,678	1,038	0,528	0,687	0,416	0,311	0,053	0,028
1000	0,684	1,004	0,534	0,595	0,305	0,374	0,025	-0,004
5000	0,727	1,037	0,499	0,578	0,030	0,359	0,038	-0,028
10000	0,727	1,136	0,497	0,541	-0,050	0,356	0,042	-0,032

	Tabela 5.	5: Mínimo e	dos viéses p	ara cada cc	nfiguração e	de amostra	simulada	
				p = 1				
	~~	g					<i>b</i>	
Amostra	10 itens	25 itens					10 itens	25 itens
500	-0,100	-0,251					-0,002	-0,017
1000	-0,113	-0,266					0,002	-0,017
5000	-0,124	-0,279					0,003	-0,019
10000	-0,128	-0,279					0,004	-0,015
				p = 2				
	0	<i>k</i> ₁	a	5			7	
Amostra	10 itens	25 itens	10 itens	25 itens			10 itens	25 itens
500	-0,219	-0,503	-0,465	-0,654			-0,055	-0,055
1000	-0,156	-0.550	-0,162	-0,793			-0,014	-0,039
5000	-0,139	-0.562	-0,173	-0.921			0,003	-0,030
10000	-0,144	-0,555	-0,188	-0.931			0,002	-0,028
				p = 3				
	0	<i>k</i> ₁	a	2	a	3	7	
Amostra	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens
500	-0,532	-0.558	-0,141	-0,165	-0,750	-1,792	-0,044	-0,102
1000	-0,559	-0.584	-0,162	-0,111	-0,962	-1,806	-0,027	-0,076
5000	-0,544	-0,630	-0,178	-0,081	-1,597	-1,811	0,000	-0,070
10000	-0,529	-0,681	-0,181	-0,031	-1,803	-1,810	-0,007	-0,070

As médias para o EQM encontram-se na Tabela 5.6 e a representação gráfica na Figura 5.4, nos quais observamos uma tendência similar à observada no viés. O parâmetro b é estimado com precisão crescente na medida que o número de individuos aumenta, sendo pouco afetado pelo aumento no número de itens. O vetor de discriminações é estimado com alta precisão quando simulamos um modelo unidimensional, mas a precisão diminui na medida que aumentamos o número de dimensões e de itens, conclusões essas que são repetidas utilizando-se o máximo dos EQM ao invés da média como na tabela 5.7.

Além da sua consistência frente ao aumento da dimensão do modelo, os métodos MCMC também possibilitam fácil interpretação de modelos mais complexos. Por exemplo, a estimação dos parâmetros por MVM do modelo logístico não-compensatório (3.6) é extremamente custosa, enquanto a implementação de tal modelo é tarefa relativamente simples em programas que utilizam métodos MCMC como WinBUGS (Bolt & Lall 2003).

				= u				
				r -				
	a	1					1	-
ostra	10 itens	25 itens					10 itens	25 itens
500	0,020	0,032					0,010	0,012
1000	0,012	0,028					0,005	0,006
5000	0,005	0,026					0,001	0,001
00001	0,005	0,026					0,001	0,001
				p = 2				
	a	1	a	5			1	-
ostra	10 itens	25 itens	10 itens	25 itens			10 itens	25 itens
500	0,080	0,115	0,179	0,138			0,013	0,015
1000	0,052	0,104	0,083	0,142			0,007	0,008
5000	0,021	0,094	0,024	0,155			0,001	0,002
10000	0,020	0,091	0,021	0,157			0,000	0,001
				p = 3				
	a	1	a	5	a,		7	2
ostra	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens
500	0,329	0,416	0,176	0,272	0,352	0,760	0,020	0,019
1000	0,271	0,363	0,110	0,210	0,385	0,780	0,009	0,009
5000	0,256	0,342	0,072	0,172	0,785	0,802	0,002	0,004
0000	0,254	0.359	0.066	0.147	0.933	0.801	0.001	0.003

o MPL c c) +:::+ q E O N Tabela 5.6: Média





	Tabela 5.	7: Máximo	dos EQM p	ara cada cc	nfiguração	de amostra	simulada	
				p = 1				
		a					1	
Amostra	10 itens	25 itens					10 itens	25 itens
500	0,040	0,085					0,020	0,027
1000	0,028	0,081					0,010	0,012
5000	0,019	0,080					0,002	0,002
10000	0,018	0,079					0,001	0,001
				p = 2				
	0	11	a	2			1	
Amostra	10 itens	25 itens	10 itens	25 itens			10 itens	25 itens
500	0,144	0,407	0,329	0,573			0,032	0,026
1000	0,093	0,409	0,203	0,702			0,014	0,014
5000	0,042	0,450	0,055	0,852			0,001	0,003
10000	0,042	0,457	0,048	0,868			0,001	0,002
				p = 3				
	0	l_1	a	2	a	3	1	
Amostra	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens	10 itens	25 itens
500	0,646	1,826	0,467	1,210	0,966	3,293	0,042	0,053
1000	0,526	1,623	0,359	0,919	1,227	3,298	0,019	0,020
5000	0,535	1,574	0,260	0,703	2,726	3,283	0,005	0,007
10000	0,531	1,702	0,251	0,550	3,322	3,279	0,003	0,006

Capítulo

6

Aplicação a Dados Reais

6.1 Inventário de Depressão de Beck

Aplicamos os métodos descritos durante o trabalho a um conjunto de dados advindo de uma aplicação do Inventório de Depressão de Beck (BDI) aplicado em 1.111 universitários gentilmente cedido pelo Dr. Teng Chei-Tung do Hospital das Clínicas da FM-USP.

O BDI consiste de 21 itens abrangendo diversos aspectos da vida do indivíduo possivelmente afetados pela presença de depressão, os quais são respondidos em função de uma escala de intensidade de 0 (Baixa intensidade) a 3 (Alta intensidade) referente ao quanto a depressão afeta o invididuo na atividade em questão. O BDI está disponível no Apêndice A. Os dados foram dicotomizados adotando-se um valor de 0 (fracasso) para respostas iguais a 0 ou faltantes e 1 (sucesso) para respostas de valor 1,2 ou 3.

O padrão de respostas faltantes no teste pode revelar aspectos importantes, como a falta de tempo dos indivíduos para responder aos itens, incoerências na redação dos itens e respostas diferentes para diferentes indivíduos. A presente abordagem foi adotada por não afetar grandemente as estimativas dos parâmetros quando comparadas com valores obtidos de uma amostra completa (Baker & Kim 2004). Existem outras alternativas para modelar os dados incompletos e levar tais aspectos (itens não apresentados, grupos de indivíduos, tempo de realização do teste entre outros) em consideração, tais modelagens e seu impacto na estimação dos parâmetros podem ser encontrados em Mislevy & Wu (1996). Para selecionarmos o modelo que melhor se adequa aos dados observados, ajustamos os dados ao modelo logístico unidimensional de dois parâmetros (UL2P) obtido de (3.3) tomando-se $a_{ik} = 0$ para $k \ge 2$ e $c_i = 0, i = 1, \dots, I$, aos modelos logísticos multidimensionas compensatório de dois parâmetros (ML2P) (3.3) e ao não-compensatório de dois parâmetros (MN2P) (3.6) supondo duas dimensões para o vetor de traços latentes $(a_{ik} = 0 \text{ para } k \ge 3 \text{ e } c_i = 0, i = 1, \dots, I).$

Os modelos compensatórios foram ajustados tanto utilizando o método da Máxima Verossilimilhança Marginal e o método MCMC descritos no Capítulo 4, enquanto o modelo não compensatório, devido a sua complexidade foi ajustado apenas utilizandose o método MCMC. Foram utilizadas para todos os modelos as distribuições *a priori* especificadas em 5.2.

Cada um dos métodos nos oferece um critério para a seleção de modelos, todos detalhadamente expostos na Seção 4.5. Tais índices foram calculados e encontram-se na Tabela 6.1.

te inper	uaue ua	estatistic	a qui-qua
	UL2P	ML2P	MN2P
DIC	24.645	24.023	41.610
AIC	25.851	25.599	
BIC	26.167	26.125	
χ^2	10.690	10.354	
gl	960	940	

Tabela 6.1:	Valores das estatísticas de seleção de modelos para os dados do BDI, g	gl:
	graus de liberdade da estatística qui-quadrado.	

O p-valor obtido da diferença entre as estatísticas qui-quadrado e utilizado por Bock & Schilling (2005) indicou significância na adição de uma dimensão no vetor de traços latentes ($p \leq 0.01$, $\chi^2 = 396$, gl = 20). Os índices baseados nos valores dos estimadores de máxima verossimilhança (BIC, AIC) não escolhem nenhum dos modelos, podendo-se creditar qualquer diferença observada a erros numéricos.

O DIC, apesar de próximo aponta uma considerável diferença entre os modelos UL2P e ML2P, e uma diferença óbvia entre os modelos compensatórios e o nãocompensatório, concordando com os outros índices na escolha do modelo logístico bidimensional e com outras metodologias que também analisam a dimensionalidade do BDI (Steer, Ball & Ranieri (1999),Cohen (2008)).

Os parâmetros dos itens foram obtidos pelo método da Máxima Verossimilhança Marginal tomando como valores iniciais os pesos associados aos fatores obtidos por Análise Fatorial para os parâmetros de discriminação e a proporção de acertos ao item dividida pela comunalidade para os parâmetros de dificuldade como em Bock et al. (1988). Foi utilizada uma aplicação do algoritmo EM com quadratura gaussiana com convergência acelerada pelo método exposto em Varadhan & Roland (2008) e implementado no pacote BB do R (Varadhan & Gilbert 2009) com 150 iterações do algoritmo EM, ao fim das quais foi observada mudança de menos de 10^{-2} nos parâmetros.

Estimativas também foram calculadas pela média das amostras obtidas pelo amostrador de Gibbs implementado no WinBUGS. Foram utilizadas 105.000 iterações, tomando um *burn-in* de 5.000 e intervalos de 50 iterações entre os valores utilizados para minimizar a autocorrelação, obtendo uma amostra de tamanho efetivo próximo de 2.000.

A convergência da cadeia após o período de *burn-in* foi verificada pelo critério de Geweke exposto na Seção 4.2.2, onde observamos valores para todos os parâmetros dentro do intervalo entre -1.96 e 1.96, indicando convergência da cadeia a 95% de confiança. Os valores obtidos em ambos os métodos e respectivos erros padrão encontram-se na Tabela 6.2

As estimativas das componentes do vetor de discriminações podem ser interpretadas em termos da capacidade do item de avaliar especificamente um traço latente. Na Figura 6.1 separamos os itens com relação aos valores reajustados do vetor de discriminações (i.e. discriminações divididas pelo poder discriminativo $MDISC_i$ (3.4)).

Podemos observar que os itens 6 e 8 avaliam primariamente um dos traços latentes, enquanto os itens 11 e itens 16 a 21 avaliam preferencialmente a segunda componente do vetor $\boldsymbol{\theta}$. Os demais itens avaliam de forma balanceada ambas as componentes.



Figura 6.1: Diagrama de Venn para os itens do BDI

Uma outra maneira de observar graficamente as estimativas dos parâmetros dos itens é utilizando um gráfico proposto por Reckase (1996). Os itens são representados por vetores no \mathbb{R}^p , com sua orientação dada pelo ângulo com relação a cada eixo representando um traço latente. O ângulo do vetor referente a
oi-ésimoitem , $i=1,\cdots,I$ relativo a
ok-ésimotraço latente, $k=1,\cdots,p$, é dado pela equação

$$\alpha_{ik} = \arccos\left(\frac{a_{ik}}{MDISC_i}\right),\tag{6.1}$$

com $MDISC_i$ definido como em (3.4), que é mostrada no gráfico como a norma do vetor. A posição da base do vetor é o ponto de distância igual à dificuldade $MDIFF_i$ (3.5) com relação à origem do plano.

Dessa maneira, podemos observar a capacidade de cada item em discriminar cada traço latente pela direção do vetor (quanto mais paralelo ao eixo, maior a capacidade do item discriminar indíviduos com respeito ao respectivo traço latente), sua capacidade de discriminação no geral (norma do vetor) e a dificuldade pela disposição espacial dos vetores no gráfico (ordenada do 3º ao 1º quadrante). O gráfico referente aos 21 itens do BDI encontra-se na Figura 6.2.



Figura 6.2: Gráfico de vetores dos resultados da aplicação do ML2P aos itens do BDI

Observa-se no gráfico que os itens discriminam ambas as dimensões, e tem via de regra um razoável poder de discriminação. As dificuldades concentram-se em valores positivos, consistente com o intuito do questionário de avaliar a intensidade da depressão em pacientes já diagnosticados com a morbidade e não diagnosticá-la em si. Os traços latentes foram estimados por esperança *a posteriori* utilizando a aproximação (4.17) supondo os parâmetros dos itens obtidos por MVM como conhecidos e a média amostral dos valores obtidos por MCMC. Calculamos a média, a variância e a correlação entre as estimativas obtidas por ambos os métodos. Os resultados encontram-se na Tabela 6.3.

Cabe observar a presença de correlação entre os traços latentes em ambos os métodos, contrariando a premissa de ausência de correlação entre os traços latentes. As estimativas entre os dois métodos foram extremamente similares em todos os indivíduos, apresentando correlações de 0.99 entre as estimativas obtidas EAP e MCMC em ambas as componentes.

2P aplicados aos dados do BDI utilizando o Método da Máxima	s estimativas dos erros padrão são denotados entre parênteses.
abela 6.2: Estimativas dos Parâmetros do ML2	Verossimilhança Marginal (MVM) e MCMC. As

	MC	(0,09)	(0,09)	(0, 16)	(0,09)	(0, 14)	(0,09)	(0, 14)	(0,09)	(0, 15)	(0,09)	(0,07)	(0,09)	(0,09)	(0,07)	(0,09)	(0,08)	(0,09)	(0,08)	(0, 12)	(0,07)	(0,09)
9	MC	0,76	-0,69	-1,58	0,11	-1,40	-0,89	0,47	0,79	-2,49	-1,17	0,05	-0,73	-0,61	-0,73	-0,18	-0,02	0,58	-1,08	-2,26	-0,82	-1,33
	VM	(0,08)	(0,08)	(0, 12)	(0,07)	(0, 11)	(0,08)	(0,09)	(0,08)	(0, 14)	(0,08)	(0,07)	(0,08)	(0,08)	(0,07)	(0,07)	(0,07)	(0,07)	(0,08)	(0,11)	(0,07)	(0,08)
	M	0,76	-0,70	-1,60	0,09	-1,43	-0,90	$0,\!45$	0,78	-2,54	-1,18	0,04	-0,74	-0,63	-0,73	-0,19	-0,03	0,57	-1,09	-2,29	-0,83	-1,34
	MC	(0, 14)	(0,15)	(0,23)	(0,19)	(0,20)	(0,13)	(0,23)	(0,13)	(0,18)	(0,13)	(0,13)	(0,16)	(0,15)	(0,11)	(0,16)	(0,17)	(0, 19)	(0,11)	(0,13)	(0,11)	(0,12)
12	MC	0,70	0,97	0,71	1,46	0,46	0,34	0,43	0,30	0,96	0,86	1,03	1,17	1,16	0,64	1,30	1,21	1,40	0,73	0,58	0,59	0,74
0	VM	(0,08)	(0,09)	(0,10)	(0,10)	(0,09)	(0,08)	(0,10)	(0,08)	(0, 12)	(0,09)	(0,08)	(0,09)	(0,09)	(0,07)	(0,10)	(0,09)	(0,10)	(0,08)	(0,10)	(0,07)	(0,08)
	Ň	0,79	1,06	0,94	1,52	0,65	0,45	0,66	0,40	1,08	0,92	1,07	1,21	1,24	0,69	1,38	1,25	1,42	0,78	0,63	0,63	0,76
	MC	(0, 13)	(0, 13)	(0, 27)	(0, 16)	(0, 25)	(0, 13)	(0,40)	(0, 14)	(0, 16)	(0, 12)	(0, 12)	(0, 13)	(0, 14)	(0, 10)	(0, 15)	(0,11)	(0, 14)	(0,11)	(0,08)	(0,09)	(0, 10)
1	MC	1,14	1,06	2,64	1,09	$2,\!43$	1,31	3,09	$1,\!45$	$1,\!22$	$0,\!80$	0,59	0,70	0,99	0,63	$1,\!12$	0,32	0,53	0,47	0,15	0,50	0,36
a	/M	(0,09)	(0,08)	(0, 16)	(0,09)	(0,15)	(0,09)	(0,21)	(0,10)	(0, 12)	(0,08)	(0,07)	(0,08)	(0,08)	(0,07)	(0,09)	(0,07)	(0,07)	(0,07)	(0,10)	(0,07)	(0,08)
	M	1,07	0.97	2,57	0.95	2,39	1,27	$3,\!22$	$1,\!43$	1,15	0,72	0,50	0,59	0,88	0,58	0,99	0,20	0,41	0,39	-0,04	0,45	0,30
	Item		2	လ	4	ъ	9	4	∞	9	10	11	12	13	14	15	16	17	18	19	20	21

 a_k : Discriminação do item com relação ao traço latent
ek=1,2 b: Dificuldade do item

	MV	Μ	MC	MC
	θ_1	θ_2	θ_1	θ_2
μ	-0,08	0,00	-0,01	-0,01
σ^2	1,18	1,11	0,62	0,71
σ_{12}	0,49		0,25	

Tabela 6.3: Média(μ), variância (σ^2) e correlações (σ_{12}) das estimativas obtidas por EAP para o vetor de traços latentes utilizando ambos os métodos

As estimativas dos traços latentes podem ser interpretadas em termos da distribuição das respostas dos individuos dentro dos grupos de itens associados primariamente a cada traço latente. Por exemplo, o vetor de traços latentes do indivíduo de número 322 foi estimado por EAP em $\hat{\theta}_{322} = (2.02, 0.33)$, indicando um valor da primeira componente do vetor de traços latentes bem acima da média e um valor mediano da segunda. Correspondentemente, tal indivíduo apresentou ausência dos sintomas associados aos itens de número 16, 18, 19 e 21, itens esses associados primariamente à segunda componente do vetor de traços latentes.

A fim de verificarmos a premissa de normalidade em cada componente do vetor de traços latentes, observamos um histograma das estimativas para ambos os métodos na Figura 6.3.

Pode-se observar que a frequência das estimativas lembra fortemente o formato da curva normal, exceto na distribuição das estimativas de θ_2 por MCMC. Também verificamos a suposição a normalidade fazendo um Q-Q plot e comparando com os quantis teóricos da distribuição normal, como podemos observar na Figura 6.4. Todas as estimativas parecem estar próximas da distribuição normal em seu centro, mas se afastam na direção das caudas.

A adequação do modelo aos dados foi verificado como em Béguin & Glas (2001) pelos escores esperados pelo modelo, como visto na Figura 6.5.

Quando utilizamos os métodos MCMC, podemos realizar checagens preditivias da posteriori, simulando um conjunto de dados para cada amostra obtida e calculando a frequencia de cada escore. Com tais frequências, podemos verificar a frequência esperada de cada escore para o modelo e calcular intervalos de confiança para a distribuição dos escores. Tais checagens foram realizadas e o resultado pode ser visto na Figura 6.6.

Pode-se observar que o modelo adotado prevê razoavelmente os escores observados. Utilizando ainda os escores esperados (4.39), utilizamos a estatistica de teste (4.44) para calcular o p-valor bayesiano descrito na Seção 4.5, obtendo p-valor de 0,13 para o ML2P, indicando que o modelo descreve adequadamente a distribuição dos escores.



Figura 6.3: Histogramas das componentes de $\hat{\theta}$ para estimativas por EAP (itens por MVM) e MCMC.

6.2 Exame Nacional do Ensino Médio (ENEM)

O Exame Nacional do Ensino Médio (ENEM) é aplicado desde 1998 a estudantes em todo o território nacional. Consistindo de 63 questões de múltipla escolha e uma redação, o ENEM almeja avaliar as competências desenvolvidas ao longo do ensino fundamental e médio.

De forma sucinta, as cinco competências que o ENEM se propõe a avaliar são (1) Dominar linguagens,(2) Compreender fenômenos, (3) Enfrentar situações-problema, (4) Construir argumentação e (5) Elaborar propostas.

Tal estrutura sugere um modelo multidimensional para a análise dos dados. Aplicamos então o ML2P a um conjunto de 5.000 respostas obtidas no ENEM de 1999 e utilizadas por Nojosa (2002) para uma análise similar.

Estimamos os parâmetros dos itens pelo método MVM nas mesmas condições da seção anterior e supondo 3, 4 e 5 dimensões para o vetor de traços latentes para calcular os AIC, BIC e estatística qui-quadrado associados a cada um dos modelos. A complexidade dos modelos tornou a estimação dos parâmetros por MCMC inviável, impedindo assim que conjecturássemos a aplicação de um modelo não compensatório.


Figura 6.4: Normal Q-Q Plots das componentes de $\hat{\theta}$ para estimativas por EAP obtidas de ambos os métodos



Figura 6.5: Escores esperados pelo M2LP (linha pontilhada) e observados (pontos) para os dados do BDI

As estatísticas encontram-se na Tabela 6.4. O AIC e BIC se mostram extremamente próximos para os três modelos, não suportando nenhuma escolha de modelos. A diferença de qui-quadrados, entretanto, quando corrigida segundo Wilson et al. (1987)



Figura 6.6: Escores observados (pontos interligados), escores esperados obtidos pelo método MCMC (linha cheia) e intervalo de 95% de confiança (linha pontilhada)

é significativa ($p \le 0.01$, $\chi^2 = 77.4$, gl = 60) para a adição de um quarto fator e não significativa ($p \ge 0.99, \chi^2 = 33, gl = 60$) para a adição de um quinto fator. Contudo, selecionamos o modelo de 5 dimensões pela sua interpretação, como feito em Nojosa (2002).

	9		
p	3	4	5
AIC	362.255	362.148	362.175
BIC	363.898	364.201	364.639
χ^2	361.751	361.518	361.419
gl	4748	4688	4629

Tabela 6.4:Valores das estatísticas de seleção de modelos para diversos modelos
utilizados no ajuste dos dados do ENEM.

Supondo então 5 dimensões para o modelo estimamos os parâmetros dos itens por MVM, obtendo os valores disponíveis no Apêndice B. Avaliamos os itens em termos de seus poderes discriminativos e dificuldades, expressados pelas estatísticas (3.4) e (3.5) como visto na Figura 6.7.

Observa-se que as questões do ENEM têm em sua maioria um bom poder discriminativo e tem itens de níveis de dificuldade tanto baixos quanto elevados, fundamentais para uma boa avaliação.



Estatísticas-resumo dos Itens

Figura 6.7: Poder discriminativo (quadrados) e dificuldade (círculos) para os itens do ENEM

Os traços latentes foram estimados pelo método bayesiano da EAP descrito na seção 4.3, obtendo as estatísticas-resumo da tabela 6.5 e correlações entre os traços latentes menores que 0.1.

Tabela 6.5: Média(μ) e variância (σ^2) das estimativas obtidas para o vetor de

	traços latentes por LAP								
	$ heta_1$	$ heta_2$	$ heta_3$	$ heta_4$	θ_5				
$\hat{\mu}$	-0,27	$0,\!48$	$0,\!19$	0,06	0,02				
$\hat{\sigma}$	$0,\!49$	$0,\!43$	$0,\!40$	$0,\!41$	$0,\!38$				

Verificamos a premissa de normalidade através dos histogramas apresentados na Figura 6.8 e de Q-Q plots, na Figura 6.9. Observam-se frequências que sugerem uma distribuição normal e quantis próximos aos esperados pela premissa de normalidade com um desvio bastante leve nas caudas.

Assim como com os dados do BDI, verificamos a adequação do modelo aos dados pelos escores previstos, ilustrados na Figura 6.10. Pode-se observar que o ML2P não



Figura 6.8: Histogramas das componentes de $\hat{\theta}$ para estimativas por EAP

prevê tão bem os escores dos indivíduos como observado com os dados do BDI, possivelmente pela ausência de um parâmetro de acerto casual. Tal parâmetro não é tão relevante nas aplicações do BDI, mas é um parâmetro importante para provas como o ENEM.



Figura 6.9: Q-Q plots das componentes de $\widehat{\boldsymbol{\theta}}$ para estimativas por EAP



Figura 6.10: Escores esperados pelo M2LP (linha pontilhada) e observados (pontos interligados) para os dados do ENEM

7 Conclusão

Capítulo

Os modelos multidimensionais ampliam os limites das aplicações da TRI para múltiplos traços latentes e podem apresentar um poder explicativo superior aos modelos unidimensionais a um custo análogo em complexidade computacional. Ao longo desse trabalho, ilustramos tais vantagens e dificuldades para os modelos que tratam com respostas dicotômicas, pela relativa simplicidade de implementação de métodos de estimação tanto frequentistas quanto bayesianos.

Contudo, nem todos os testes resultam em respostas dicotômicas. O proprio BDI é aplicado normalmente em uma escala de Likert, com 4 graus denotando a intensidade do sintoma. Mais geralmente, é possível classificar a resposta do inidíduo ao item em categorias, demandando modelos da TRI que levem em conta tais nuances. Kelderman & Rijkes (1994) propôe modelos para respostas em categorias, chamados modelos politônicos e modelos para escalas como a do BDI, denominados modelos de crédito parcial. Kelderman (1996) propôe extensões de tais modelos para múltiplos traços latentes.

Os modelos multidimensionais podem também ser utilizados para a análise de dados longitudinais. te Marvelde, Glas, Landeghem & Van Damme (2006) utilizam uma extensão dos modelos multidimensionais para dados categóricos em um ajuste que leve medidas repetidas em consideração e possa verificar o desenvolvimento de um indivíduo ao longo do tempo.

Uma premissa utilizada ao longo desse trabalho foi a de que os traços latentes não são correlacionados. Tal premissa é presumidamente falsa em modelos que tomem o desenvolvimento de um indivíduo como o vetor de traços latentes, além de ser frequentemente uma premissa fracamente observada na prática. O impacto de tal correlação entre fraca e moderada nas estimativas dos parâmetros não é grande no caso do Modelo Logístico Compensatório de 2 Parâmetros que utilizamos nos ajustes (Bolt & Lall 2003), de modo que utilizamos tal premissa para preservar a identificabilidade do modelo.

Entretanto, também é visto no mesmo trabalho que as estimativas tornam-se bem menos confiáveis na medida que a correlação se mantém significativa, mas a complexidade do modelo aumenta. Albert (1992) impôe algumas restrições sobre os parâmetros dos itens para levar tais correlações em consideração, com considerável dano à interpretabilidade das estimativas, pois as limitações são meramente matemáticas, sem paralelo com o significado dos parâmetros. Outras tentativas de levar a correlação em consideração sem prejudicar a identificabilidade é utilizada no programa NOHARM que assume que alguns itens avaliam um e apenas um traço latente, fixando a discriminação relacionada aos demais traços em 0. Tal limitação parece de tal maneira simplista para certos testes, demandando novas propostas para a identificabilidade dos modelos da TRI multidimensional.

Outra premissa a ser verificada é a de normalidade multivariada do vetor de traços latentes.Bazán, Branco & Bolfarine (2006) por exemplo utilizam distribuições assimétricas que poderiam ser estendidas para o caso multidimensional, assim como modelagens utilizando DIF.

Apêndice

Inventário de Depressão de Beck

INVENTÁRIO DE DEPRESSÃO DE BECK (BDI)

Este questionário consiste em 21 grupos de afirmações. Depois de ler cuidadosamente cada grupo, faça um círculo em torno do número (0, 1, 2 ou 3) diante da afirmação, em cada grupo, que descreve melhor a maneira como você tem se sentido nesta semana, incluindo hoje. Se várias afirmações num grupo parecerem se aplicar igualmente bem, faça um círculo em cada uma. Tome o cuidado de ler todas as afirmações, em cada grupo, antes de fazer a sua escolha.

1.

- 0~ Não me sinto triste.
- 1 Eu me sinto triste.
- 2 Estou sempre triste e não consigo sair disso.
- 3 Estou tão triste ou infeliz que não consigo suportar.

- 0 Não estou especialmente desanimado quanto ao futuro.
- 1 Eu me sinto desanimado quanto ao futuro.
- 2 Acho que nada tenho a esperar.

3 Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar.

3.

- 0 Não me sinto um fracasso.
- 1 Acho que fracassei mais do que uma pessoa comum.
- 2 Quando olho para trás, na minha vida, tudo o que posso ver é um monte de fracassos.
- 3 Acho que, como pessoa, sou um completo fracasso.

4.

- 0 Tenho tanto prazer em tudo como antes.
- 1 Não sinto mais prazer nas coisas como antes.
- 2 Não encontro um prazer real em mais nada.
- 3 Estou insatisfeito ou aborrecido com tudo.

5.

- 0 Não me sinto especialmente culpado.
- 1 Eu me sinto culpado às vezes.
- 2 Eu me sinto culpado na maior parte do tempo.
- 3 Eu me sinto sempre culpado.

6.

- 0 Não acho que esteja sendo punido.
- 1 Acho que posso ser punido.
- 2 Creio que vou ser punido.
- 3 Acho que estou sendo punido.

- 0 Não me sinto decepcionado comigo mesmo.
- 1 Estou decepcionado comigo mesmo.

- 2 Estou enojado de mim.
- 3 Eu me odeio.

8.

- 0 Não me sinto de qualquer modo pior que os outros.
- 1 Sou crítico em relação a mim devido a minhas fraquezas ou meus erros.
- 2 Eu me culpo sempre por minhas falhas.
- 3 Eu me culpo por tudo de mal que acontece.

9.

- 0 Não tenho quaisquer idéias de me matar.
- 1 Tenho idéias de me matar, mas não as executaria.
- 2 Gostaria de me matar.
- 3 Eu me mataria se tivesse oportunidade.

10.

- 0 Não choro mais que o habitual.
- 1 Choro mais agora do que costumava.
- 2 Agora, choro o tempo todo.
- 3 Costumava ser capaz de chorar, mas agora não consigo mesmo que o queira.

11.

- 0 Não sou mais irritado agora do que já fui.
- 1 Fico molestado ou irritado mais facilmente do que costumava.
- 2 Atualmente me sinto irritado o tempo todo.
- 3 Absolutamente não me irrito com as coisas que costumavam irritar-me.

- 0 Não perdi o interesse nas outras pessoas.
- 1 Interesso-me menos do que costumava pelas outras pessoas.
- 2 Perdi a maior parte do meu interesse nas outras pessoas.

3 Perdi todo o meu interesse nas outras pessoas.

13.

- 0 Tomo decisões mais ou menos tão bem como em outra época.
- 1 Adio minhas decisões mais do que costumava.
- 2 Tenho maior dificuldade em tomar decisões do que antes.
- 3 Não consigo mais tomar decisões.

14.

- 0 Não sinto que minha aparência seja pior do que costumava ser.
- 1 Preocupo-me por estar parecendo velho ou sem atrativos.
- 2 Sinto que há mudanças permanentes em minha aparência que me fazem parecer sem atrativos.
- 3 Considero-me feio.

15.

- 0 Posso trabalhar mais ou menos tão bem quanto antes.
- 1 Preciso de um esforço extra para começar qualquer coisa.
- 2 Tenho de me esforçar muito até fazer qualquer coisa.
- 3 Não consigo fazer nenhum trabalho.

16.

- 0 Durmo tão bem quanto de hábito.
- 1 Não durmo tão bem quanto costumava.
- 2 Acordo uma ou duas horas mais cedo do que de hábito e tenho dificuldade para voltar a dormir.
- 3 Acordo várias horas mais cedo do que costumava e tenho dificuldade para voltar a dormir.

- 0 Não fico mais cansado que de hábito.
- 1 Fico cansado com mais facilidade do que costumava.

2 Sinto-me cansado ao fazer quase qualquer coisa.

3 Estou cansado demais para fazer qualquer coisa.

18.

- 0 Meu apetite não está pior do que de hábito.
- 1 Meu apetite não é tão bom quanto costumava ser.
- 2 Meu apetite está muito pior agora.
- 3 Não tenho mais nenhum apetite.

19.

- 0 Não perdi muito peso, se é que perdi algum ultimamente.
- 1 Perdi mais de 2,5 Kg.
- 2 Perdi mais de 5,0 Kg.
- 3 Perdi mais de 7,5 Kg.

Estou deliberadamente tentando perder peso, comendo menos: SIM () NÃO ()

20.

- 0 Não me preocupo mais que o de hábito com minha saúde.
- 1 Preocupo-me com problemas físicos como dores e aflições ou perturbações no estômago ou prisão de ventre.
- 2 Estou muito preocupado com problemas físicos e é difícil pensar em outra coisa que não isso.
- 3 Estou tão preocupado com meus problemas físicos que não consigo pensar em outra coisa.

- 0 Não tenho observado qualquer mudança recente em meu interesse sexual.
- 1 Estou menos interessado por sexo que costumava.
- 2 Estou bem menos interessado em sexo atualmente.
- 3 Perdi completamente o interesse por sexo.

Apêndice

Estimativas dos Parâmetros dos Itens do ENEM

Tabela B.1:	Estimativas	para os	s parâmetros	dos	itens	do
ENEM						

Item	$\mathbf{a} = a_1$		a_2		a_3		a_4		a ₅		b	
1	0,28	(0,08)	0,53	(0,09)	0,18	(0,08)	-0,04	(0,08)	0,31	(0,08)	-2,18	(0,11)
2	1,42	(0,09)	1,59	(0,10)	0,44	(0,06)	0,78	(0,07)	$0,\!25$	(0,06)	-1,84	(0,10)
3	0,34	(0,04)	0,09	(0,03)	0,09	(0,03)	$0,\!41$	(0,04)	-0,09	(0,03)	-0,26	(0,03)
4	0,76	(0,05)	0,40	(0,04)	0,26	(0,04)	$0,\!51$	(0,04)	$0,\!13$	(0,04)	0,47	(0,04)
5	0,22	(0,04)	0,56	(0,05)	0,13	(0,04)	0,20	(0,04)	-0,03	(0,04)	-0,99	(0,05)
6	$0,\!58$	(0,05)	0,32	(0,05)	0,31	(0,05)	$0,\!18$	(0,04)	-0,26	(0,04)	-1,16	(0,05)
7	0,51	(0,06)	1,02	(0,07)	0,37	(0,05)	0,21	(0,05)	0,06	(0,05)	-1,65	(0,08)
8	1,72	(0,10)	0,70	(0,07)	0,15	(0,06)	$0,\!53$	(0,06)	$0,\!17$	(0,06)	1,87	(0,10)
9	-0,14	(0,07)	0,33	(0,07)	0,20	(0,07)	0,00	(0,07)	$0,\!10$	(0,07)	-1,84	(0,08)
10	0,95	(0,05)	0,55	(0,04)	0,33	(0,04)	$0,\!23$	(0,04)	-0,17	(0,04)	0,24	(0,04)
11	1,08	(0,06)	0,47	(0,04)	0,21	(0,04)	0,02	(0,04)	-0,23	(0,04)	0,16	(0,04)
12	0,31	(0,04)	0,21	(0,04)	0,21	(0,04)	$0,\!13$	(0,04)	-0,04	(0,04)	-0,70	(0,04)
13	0,28	(0,04)	-0,04	(0,04)	-0,03	(0,04)	0,92	(0,05)	-0,01	(0,04)	-0,51	(0,04)
14	0,14	(0,05)	0,44	(0,05)	0,17	(0,05)	$0,\!07$	(0,05)	0,06	(0,05)	-1,40	(0,06)
15	$0,\!68$	(0,04)	0,66	(0,04)	0,34	(0,04)	$0,\!12$	(0,04)	0,01	(0,04)	-0,19	(0,04)
16	0,28	(0,04)	0,03	(0,04)	0,17	(0,04)	0,78	(0,04)	0,01	(0,04)	-0,12	(0,03)
17	0,29	(0,04)	0,22	(0,04)	0,07	(0,04)	$0,\!10$	(0,04)	0,08	(0,04)	-0,96	(0,04)
18	1,02	(0,06)	0,22	(0,04)	0,78	(0,05)	$0,\!16$	(0,04)	$0,\!29$	(0,04)	0,56	(0,04)
19	$0,\!67$	(0,05)	0,68	(0,05)	0,42	(0,04)	0,16	(0,04)	0,14	(0,04)	-0,70	(0,04)
20	1,12	(0,06)	0,49	(0,04)	0,27	(0,04)	$0,\!28$	(0,04)	0,32	(0,04)	-0,09	(0,04)

Item	a_1		0	a_2		a_3		a_4				b	
21	0,50	(0,04)	0,11	(0,03)	0,22	(0,04)	0,29	(0,04)	0,10	(0,03)	0,40	(0,03)	
22	0,30	(0,04)	0,01	(0,04)	0,37	(0,04)	0,17	(0,04)	0,19	(0,04)	-0,90	(0,04)	
23	0,30	(0,04)	0,29	(0,04)	0,03	(0,03)	0,09	(0,03)	-0,07	(0,03)	-0,54	(0,03)	
24	0,14	(0,05)	0,31	(0,05)	$0,\!24$	(0,05)	0,27	(0,05)	0,07	(0,04)	-1,18	(0,05)	
25	0,75	(0,05)	0,37	(0,04)	$0,\!27$	(0,04)	0,38	(0,04)	-0,02	(0,04)	0,18	(0,04)	
26	0,91	(0,05)	0,32	(0,04)	$0,\!43$	(0,04)	0,23	(0,04)	-0,14	(0,04)	-0,03	(0,04)	
27	1,04	(0,05)	0,46	(0,04)	$0,\!17$	(0,04)	0,21	(0,04)	0,04	(0,04)	-0,15	(0,04)	
28	0,82	(0,05)	0,52	(0,04)	$0,\!54$	(0,04)	0,40	(0,04)	0,02	(0,04)	-0,48	(0,04)	
29	0,19	(0,05)	0,60	(0,06)	$0,\!27$	(0,05)	-0,14	(0,05)	0,38	(0,05)	-1,55	(0,07)	
30	0,88	(0,05)	0,38	(0,04)	$0,\!17$	(0,04)	0,06	(0,04)	0,78	(0,05)	-0,66	(0,04)	
31	0,21	(0,04)	0,27	(0,04)	$0,\!25$	(0,04)	0,01	(0,04)	-0,04	(0,04)	-0,85	(0,04)	
32	$0,\!65$	(0,04)	0,52	(0,04)	0,22	(0,04)	0,20	(0,04)	-0,04	(0,04)	-0,56	(0,04)	
33	0,86	(0,05)	0,09	(0,04)	$0,\!14$	(0,04)	0,12	(0,04)	-0,16	(0,04)	$0,\!30$	(0,04)	
34	0,90	(0,05)	0,09	(0,04)	$0,\!28$	(0,04)	0,32	(0,04)	0,32	(0,04)	$0,\!50$	(0,04)	
35	0,52	(0,04)	-0,15	(0,04)	$0,\!27$	(0,04)	$0,\!55$	(0,04)	0,21	(0,04)	0,24	(0,03)	
36	0,28	(0,05)	0,73	(0,06)	$0,\!31$	(0,05)	0,12	(0,05)	0,16	(0,05)	-1,49	(0,07)	
37	-0,29	(0,07)	0,01	(0,07)	$0,\!05$	(0,07)	-0,02	(0,07)	-0,02	(0,07)	-1,74	(0,07)	
38	0,86	(0,05)	0,55	(0,04)	0,36	(0,04)	0,03	(0,04)	-0,04	(0,04)	-0,17	(0,04)	
39	1,41	(0,09)	0,27	(0,06)	$0,\!68$	(0,07)	0,55	(0,07)	0,05	(0,06)	2,13	(0, 11)	
40	0,48	(0,04)	0,68	(0,05)	$0,\!27$	(0,04)	0,19	(0,04)	0,00	(0,04)	-0,89	(0,05)	
41	0,80	(0,05)	0,03	(0,04)	0,20	(0,04)	-0,06	(0,04)	0,05	(0,04)	0,71	(0,04)	
42	1,11	(0,06)	0,52	(0,05)	$0,\!48$	(0,05)	0,33	(0,04)	0,41	(0,04)	0,21	(0,04)	
43	0,61	(0,05)	0,38	(0,04)	$0,\!18$	(0,04)	0,17	(0,04)	0,18	(0,04)	-1,08	(0,05)	
44	1,39	(0,08)	0,36	(0,05)	$0,\!35$	(0,05)	0,46	(0,05)	0,44	(0,05)	1,13	(0,06)	
45	0,40	(0,07)	0,83	(0,08)	$0,\!28$	(0,07)	0,17	(0,07)	0,04	(0,07)	-2,14	(0,11)	
46	0,63	(0,04)	0,20	(0,04)	$0,\!40$	(0,04)	0,18	(0,04)	0,06	(0,04)	-0,32	(0,03)	
47	0,10	(0,04)	-0,08	(0,04)	$0,\!12$	(0,04)	0,12	(0,04)	0,16	(0,04)	-0,67	(0,03)	
48	0,33	(0,04)	0,40	(0,04)	0,20	(0,04)	0,14	(0,04)	-0,07	(0,04)	-1,05	(0,05)	
49	0,41	(0,04)	0,22	(0,03)	$0,\!18$	(0,03)	0,18	(0,03)	0,01	(0,03)	-0,31	(0,03)	
50	0,11	(0,04)	0,70	(0,05)	0,23	(0,04)	0,17	(0,04)	-0,06	(0,04)	-1,08	(0,05)	
51	0,29	(0,04)	0,53	(0,05)	0,28	(0,04)	0,09	(0,04)	-0,31	(0,04)	-1,13	(0,05)	
52	0,06	(0,03)	0,22	(0,04)	0,38	(0,04)	0,30	(0,04)	0,11	(0,04)	-0,54	(0,04)	
53	0,48	(0,04)	0,57	(0,04)	0,24	(0,04)	0,10	(0,04)	-0,02	(0,04)	-0,76	(0,04)	
54	0,92	(0,05)	0,25	(0,04)	0,75	(0,05)	0,00	(0,04)	0,10	(0,04)	0,49	(0,04)	
55	0,28	(0,04)	0,24	(0,04)	0,58	(0,04)	0,17	(0,04)	-0,24	(0,04)	-0,63	(0,04)	
56	0,31	(0,04)	-0,09	(0,04)	0,64	(0,04)	0,20	(0,04)	0,00	(0,03)	-0,32	(0,03)	
57	0,55	(0,06)	0,83	(0,06)	0,29	(0,05)	0,05	(0,05)	-0,06	(0,05)	-1,49	(0,07)	
58	-0,44	(0,11)	1,48	(0,17)	0,47	(0,12)		(0,11)		(0,11)	-3,67	(0,29)	
59		(0,04)	0,28	(0,04)	0,41	(0,04)	0,28	(0,04)	-0,05	(0,04)	-0,15	(0,03)	
60		(0,04)	0,28	(0,04)	0,39	(0,04)		(0,04)	-0,09	(0,04)	-0,69	(0,04)	
61 co	0,98	(0,05)	0,47	(0,04)	0,70	(0,05)		(0,04)	0,09	(0,04)	0,20	(0,04)	
62	0,56	(0,05)	0,13	(0,04)	0,97	(0,06)	0,27	(0,04)	0,28	(0,04)	-0,67	(0,04)	
63	0,53	(0,04)	0,42	(0,04)	0,20	(0,04)	0,07	(0,04)	0,06	(0,04)	-0,60	(0,04)	

Tabela B.1 – Cont.

Apêndice

Algorítmos

C.1 Estimação dos Parâmetros dos Itens por MVM

```
library(BB); #Resolução de sistemas não lineares para o Passo M
#Leitura dos dados
N <- nrow(X); #Numero de elementos
Q <- ncol(X);#Numero de itens</pre>
p <- 2; #numero de dimensões do modelo
ITS <- 150; #Numero de iterações do Algoritmo EM
#Padrões de resposta
pad <- matrix(nrow=N);</pre>
for(i in 1:N)
ł
  pad[i] <- paste(X[i,],collapse=" ");</pre>
}
AUX <- table(pad);
S <- length(AUX); #numero de padrões de resposta;
R <- matrix(nrow=S);#ocorrencia de cada padrão de resposta;
U <- matrix(nrow=S,ncol=Q);#padrões de resposta
for(i in 1:S)
{
 R[i] <- AUX[[i]];</pre>
 aux1 <- strsplit(names(AUX)[[i]],split=" ");</pre>
```

```
aux1 <- aux1[[1]];</pre>
 U[i,] <- as.numeric(aux1);</pre>
 }
#Valores iniciais obtidos por Analise Fatorial
ai <- matrix(nrow=Q,ncol=p); #Vetor de discriminações
bi <- matrix(nrow=Q); #Vetor de dificuldades</pre>
Fac <- factanal(X,p); #analise fatorial por máxima verossimilhança
props <- (apply(X,2,sum)/N);</pre>
for(i in 1:p)
{
  for(j in 1:Q)
  ſ
  SS1 <- sqrt(1+sum(Fac$loadings[j,]^2));</pre>
  ai[j,i] <- Fac$loadings[j,i]/SS1;</pre>
  bi[j] <- -qnorm(props[j])/SS1;</pre>
  }
}
#Nós de quadratura
delta <- 1;
coords <- seq(-3,3,by=delta);</pre>
coords <- list(coords);</pre>
eixos <- rep(coords,p);</pre>
theta <- expand.grid(eixos);</pre>
theta <- as.matrix(theta);</pre>
g <- matrix(nrow=nrow(theta)); #densidade da normal multivariada e peso
                                  #de quadratura associado ao nó
g <- apply(dnorm(theta),1,prod);</pre>
for(int in 1:ITS)
{
print(c("Iteração",int));
#Passo E:
#Cálculo das probabilidades de acordo com o modelo logístico de 2 parâmetros
P <- matrix(nrow=nrow(theta),ncol=Q);</pre>
for(i in 1:nrow(theta))
ſ
```

```
for(j in 1:Q)
  {
    P[i,j] <- 1/(1+ exp(-(ai[j,]%*%theta[i,] +bi[j])));</pre>
  }
}
#Função verossimilhança
L <- matrix(nrow=nrow(theta),ncol=S);</pre>
  for(i in 1:nrow(theta))
    ł
     for(j in 1:S)
       {
           L[i,j] <- prod((P[i,]^U[j,])*((1-P[i,])^(1-U[j,])));
       }
    }
#Probabilidades marginais e densidade a posteriori
gs <- matrix(nrow=nrow(theta),ncol=S);</pre>
Pm <- matrix(nrow=1,ncol=S);</pre>
for(j in 1:S)
{
 Pm[j] <- L[,j]%*%g; # Integral ~ sum(L[,1]*g)</pre>
 for(i in 1:nrow(theta))
 {
  gs[i,j] <- L[i,j]/Pm[j];</pre>
 }
}
#Número esperado e acertos esperados
N0 <- matrix(nrow=nrow(theta));</pre>
r0 <- matrix(nrow=nrow(theta),ncol=Q);</pre>
for(i in 1:nrow(theta))
{
NO[i] <- sum(R*gs[i,]);</pre>
 for(j in 1:Q)
 {
```

```
r0[i,j] <- U[,j]%*%(R*gs[i,]);
 }
}
#Passo M:
    for(j in 1:Q){
    pars <- matrix(ncol=p+1);</pre>
    pars[,1:p] <- ai[j,];</pre>
    pars[,p+1] <- bi[j];</pre>
    pars <- c(t(pars));</pre>
    grad <- function(x,k) #Gradiente da log-verossimilhança</pre>
    {
       dlog <- matrix(nrow=(p+1));</pre>
       P <- matrix(nrow=nrow(theta),ncol=1);</pre>
           for(i in 1:nrow(P))
           {
                {
                  P[i] <- 1/(1 + exp(-(x[1:p]%*%theta[i,] + x[p+1])));</pre>
                }
           }
              {
                dl <- matrix(nrow=p+1);</pre>
                aux2 <- r0[,k] - N0*P;
                for(i in 1:p)
                {
                  difa <- -theta[,i];</pre>
                  dl[i] <- sum(difa*aux2*g);</pre>
                }
                difb <- -1;
                dl[p+1] <- sum(difb*aux2*g);</pre>
                dlog <- dl;</pre>
                }
       return(as.numeric(dlog));
    }
  print(c("Item",j));
  erro <- t(grad(pars,j))%*%grad(pars,j);</pre>
```

```
if(erro>10^-5)
{
pars <- dfsane(pars,</pre>
 grad,
k=j,
method=3,
 control=list(trace=TRUE,maxit=10000,M=10))$par;
}
  {
   ai[j,] <- pars[1:p];
  bi[j] <- pars[p+1];</pre>
  }
}
print(c("-2xLoglik",2*sum(-t(R)*log(Pm))));
}#Fim das iterações
#Chi-quadrado
Gsquared <- -2*sum(t(R)*log(N*Pm/t(R)));</pre>
DF <- (S-1) - ((p+1)*Q -p*(p-1)/2);
aic <- 2*Q*(2*p+1) - 2*sum(t(R)*log(Pm));
bic <- Q*(2*p+1)*log(N) - 2*sum(t(R)*log(Pm));</pre>
print(ai);
print(bi);
print(c("Qui-Quadrado",Gsquared,"DF",DF));
print(c("AIC",aic));
print(c("BIC",bic));
ac <- ai;
bc <- bi;</pre>
MDISC <- sqrt(apply(ai^2,1,sum)); #poder discriminativo de cada item</pre>
for(i in 1:Q)
{
if(p>1){ac[i,] <- ai[i,]/MDISC[i];}</pre>
bc[i] <- -bi[i]/MDISC[i];</pre>
}
print("Valores Reajustados dos parâmetros dos itens");
print("Discriminação");
print(ac);
```

```
print("Dificuldade");
print(bc);
```

C.2 Erros-Padrão dos Estimadores de MVM

```
Infs <- array(dim=c((p+1),(p+1),Q));</pre>
Fi <- matrix(nrow=(p+1),ncol=(p+1));</pre>
for(k in 1:Q)
{
    for(i in 1:p)
    {
       for(j in 1:p)
       {
             Fi[i,j] <- N*sum(P[,k]*(1-P[,k])*theta[,i]*theta[,j]*g);</pre>
       }
       Fi[i,(p+1)] <- N*sum(P[,k]*(1-P[,k])*theta[,i]*g);</pre>
       Fi[(p+1),i] <- Fi[i,(p+1)];</pre>
    }
    Fi[(p+1),(p+1)] <- N*sum(P[,k]*(1-P[,k])*g);</pre>
    Infs[,,k] <- solve(Fi);</pre>
}
Sds <- array(dim=c((p+1),1,Q));</pre>
for(i in 1:Q)
{
 Sds[,,i] <- sqrt(diag(Infs[,,i]));</pre>
}
```

C.3 Estimação Bayesiana dos Parâmetros dos Itens por MCMC

```
model M2PL
{
    for (k in 1:Q) {
        b[k] ~ dnorm(0, 1)
        for (j in 1:p) {
```

```
a[k, j] ~ dlnorm(1, 0.5)
        }
    }
    for (j in 1:N) {
        for (k in 1:Q) {
             P[j, k] <- (1/(1 + exp(-inprod(a[k, ], theta[j, ]) -</pre>
                 b[k])))
            X[j, k] \sim dbern(P[j, k])
        }
        theta[j, 1:p] ~ dmnorm(Mt[1:p], Vt[1:p, 1:p])
    }
    }
}
#Modelo não-compensatório bidimensional
MNC2 <- function()</pre>
{
#prioris
for(k in 1:Q){
for(j in 1:p)
{
a[k,j] ~ dlnorm(1,.5);
b[k,j] ~ dnorm(0,1);
}
}
for(j in 1:N)
{
  theta[j,1:p] ~ dmnorm(Mt[1:p],Vt[1:p,1:p]);
}
#Probabilidade de acerto
  for(i in 1:Q)
  {
   for(j in 1:N)
   {
    P[j,i] <- (1/( 1 + exp(-(a[i,1]*theta[j,1]) + b[i,1])))*(1/( 1 + exp(-(a[i,2]*t</pre>
    X[j,i] ~ dbern(P[j,i]);
```

} }

}

C.4 Estimação dos Traços Latentes por Máxima Verossimilhança e EAP

```
print('Estimação dos Traços Latentes');
P <- matrix(nrow=nrow(theta),ncol=Q);</pre>
for(i in 1:nrow(theta))
{
  for(j in 1:Q)
  {
    P[i,j] <- 1/(1+ exp(-1.702*(ai[j,]%*%theta[i,] +bi[j])));</pre>
  }
}
#Verossimilhança
L <- matrix(nrow=nrow(theta),ncol=N);</pre>
  for(i in 1:nrow(theta))
    {
     for(j in 1:N)
       {
          L[i,j] <- prod((P[i,]^X[j,])*((1-P[i,])^(1-X[j,])));</pre>
       }
    }
#Probabilidades marginais e densidade a posteriori
Pm <- matrix(nrow=1,ncol=N); #probabilidade marginal para um padrão de resposta
for(j in 1:N)
{
 Pm[j] <- L[,j]%*%g; # Integral ~ sum(L[,1]*g)</pre>
}
    EAP <- matrix(nrow=N,ncol=p);#Estimativas por EAP para cada ind
    for(i in 1:N)
```

```
{
   if((sum(X[i,])!=Q)&((sum(X[i,])!=0)))#Escores nulos ou perfeitos não tem
   {
                                           #valores definidos.
    for(j in 1:p)
      {
       aux <- (theta[,j]*L[,i]*g);</pre>
       EAP[i,j] <- sign(sum(aux)/Pm[i])*min(abs(sum(aux)/Pm[i]),3.9);</pre>
      }
    }
    if((sum(X[i,])==Q)){EAP[i,] <- rep(4,p)};</pre>
    if((sum(X[i,])==0)){EAP[i,] <- rep(-4,p)}
  }
dtheta <- function(x,k) #Gradiente da log-verossimilhança para a habilidade
  ſ
  dl <- matrix(ncol=1,nrow=p);</pre>
    P <- matrix(nrow=1,ncol=Q);</pre>
        for(i in 1:Q)
        {
               P[i] <- 1/(1 + exp(-1.7*(ai[i,]%*%x + bi[i])));</pre>
        }
            aux <- -(1-X[k,])*P + (1-P)*(X[k,]);
             for(i in 1:p)
             {
               difP <- -1.702*ai[,i];
               dl[i] <- sum(difP*aux);</pre>
             }
    return(as.numeric(sum(dl^2)));
  }
MLE <- matrix(nrow=N,ncol=p);</pre>
for(j in 1:N)
{
 if((sum(X[j,])!=Q)&((sum(X[,])!=0)))
 {
  MLE[j,] <- spg(rep(0,p),dtheta,k=j,lower=rep(-3,p),upper=rep(3,p),control=list(</pre>
 }
```

```
if((sum(X[j,])==Q)){MLE[j,] <- rep(4,p)};
if((sum(X[j,])==0)){MLE[j,] <- rep(-4,p)}
}
print("Médias EAP");
print(apply(EAP,2,mean));
print("Médias MLE");
print(apply(MLE,2,mean));
print("Variâncias EAP");
print(apply(EAP,2,var));
print("Variâncias MLE");
print(apply(MLE,2,var));
print("Correlações, MLE");
print(cor(MLE[,1],MLE[,2]));
print("Correlações, EAP");
print(cor(EAP[,1],EAP[,2]));
print("Correlação entre estimativas");
print(cor(MLE[,1],EAP[,1]));
print(cor(MLE[,2],EAP[,2]));
```

C.5 Resíduos Padronizados

```
Es <- matrix(nrow=N,ncol=Q);#Valor Esperado da resposta ao item i pelo padrão de respo
Var <- matrix(nrow=N,ncol=Q); #Variância da resposta s ao item i
Zscore <- matrix(nrow=N,ncol=Q); #Residuo padronizado
for(j in 1:N)
{
    for(i in 1:Q)
    {
        Es[j,i] <- exp(1.7*(ai[i,]%*%EAP[j,]+bi[i]))/(1+exp(1.7*(ai[i,]%*%EAP[j,]+bi[i])));
        Var[j,i] <- Es[j,i]*(1-Es[j,i]);
        Zscore[j,i] <- (X[j,i]-Es[j,i])/(sqrt(Var[j,i]));
    }
}
Pfit <- apply(Zscore^2,1,mean);
Ifit <- apply(Zscore^2,2,mean);</pre>
```

C.6 Escore Esperado

```
library(Matrix);
Sk <- function(P)
{
  aux <- matrix(0,nrow=(length(P)+1));</pre>
  aux[1] <- 1-P[1];
  aux[2] <- P[1];
  for(i in 2:length(P))
  {
   Pk <- rep(P[i],i);</pre>
   Qk <- (1-P[i]);
   Qk <- rep(Qk,i);</pre>
   aux1 <- bandSparse(n=(i+1),</pre>
             m=i ,
             k=c(-1,0),
             diagonals=list(Pk,Qk), symmetric = FALSE);
   aux1 <- as.matrix(aux1);</pre>
   aux[1:(i+1)] <- aux1%*%aux[1:i];</pre>
  }
  return(aux);
  }
  Esp <- matrix(nrow=(Q+1),ncol=nrow(P));#escores esperados para um nó</pre>
  for(j in 1:nrow(P))
  {
   Esp[,j] <- Sk(P[j,]);</pre>
  }
  fr <- matrix(nrow=(Q+1)); #Escore marginal para cada padrão</pre>
  for(j in 1:(Q+1))
  {
   fr[j] <- N*Esp[j,]%*%g;</pre>
  }
  Sobs <- apply(X,1,sum);</pre>
  Nobs <- matrix(0,nrow=(Q+1));</pre>
  for(i in 1:(Q+1))
```

{
 Nobs[i] <- length(which(Sobs==(i-1)));</pre>

}

print(data.frame("Escore"=0:Q,"Observados"=Nobs,"Previstos"=round(fr)));

C.7 Cálculo da Estatística $s - \chi^2$

```
Oik <- matrix(0,nrow=Q,ncol=(Q)); #prop. de acertos observados
Eik <- matrix(0,nrow=Q,ncol=(Q)); #prop esperada de respostas</pre>
Sobs <- apply(X,1,sum);</pre>
Sums <- data.frame(X,Sobs);</pre>
for(i in 1:Q)
{
 Oik[(i),] <- apply(X[which(Sums$Sobs==i),1:Q],2,sum)/Nobs[i+1];</pre>
}
Esps <- array(dim=c(nrow(P),Q,Q));#No x Cat x item deletado</pre>
for(i in 1:Q)
{
  for(r in 1:nrow(P))
 {
 Esps[r,,i] \leq Sk(P[r,-i]);
 }
 Esps[,,i] <- P[,i]*Esps[,,i];</pre>
}
Num <- matrix(nrow=Q,ncol=Q);</pre>
for(i in 1:Q)
{
 for(j in 1:Q)
 {
  Num[i,j] <- Esps[,i,j]%*%g;</pre>
 }
}
for(i in 1:Q)
{
```

```
for(j in 1:Q)
{
   Eik[i,j] <- (Num[i,j])/(fr[i+1]/N);
}

Eik <- Eik[1:(Q-1),1:Q];
Oik <- Oik[1:(Q-1),1:Q];
Schisq <- apply(((Nobs[2:Q])*(Oik-Eik)^2)/(Eik*(1-Eik)),1,sum);
DFc <- Q + 1 - (p+1);
print(c("s-X2",round(Schisq,3),"df",DFc));</pre>
```

C.8 Checagens Preditivas da Posteriori e DIC

```
CPP <- function(A,B,Th,p)#Argumentos: A=amostra par. a, B=b Th = traços, p = dim
{
Rs <- niter(A);</pre>
Reps <- array(dim=c(N,Q,Rs));</pre>
AUX <- array(dim=c(1,1,Rs));
fr <- matrix(nrow=(Q+1),ncol=Rs);</pre>
X20 <- matrix(nrow=Rs);</pre>
X2r <- matrix(nrow=Rs);</pre>
for(r in 1:Rs)
{
    if((r%%100)==0)
    {print(c("Replicação",r));}
    aux1 <- A[r,];
    aux2 <- B[r,];
    aux3 <- Th[r,];
    am <- matrix(data.frame(aux1)$aux1,nrow=Q,ncol=p,byrow=TRUE);</pre>
    bm <- matrix(data.frame(aux2)$aux2,nrow=Q,ncol=1,byrow=TRUE);</pre>
    tr <- matrix(data.frame(aux3)$aux3,nrow=N,ncol=p,byrow=TRUE);</pre>
    Pm <- matrix(nrow=N,ncol=Q);</pre>
    Xr <- matrix(nrow=N,ncol=Q);</pre>
    for(i in 1:Q)
    {
     for(j in 1:N)
     {
```

```
#Modelo compensatorio
  Pm[j,i] <- exp(am[i,]%*%tr[j,] + bm[i])/(1+exp(am[i,]%*%tr[j,] + bm[i]));</pre>
  #Modelo não-compensatorio
  #Pm[j,i] <- exp(am[i,1]%*%tr[j,1] + bm[i,1])/(1+exp(am[i,1]%*%tr[j,1] + bm[i,1])</pre>
  #exp(am[i,2]%*%tr[j,2] + bm[i,2])/(1+exp(am[i,2]%*%tr[j,2] + bm[i,2]));
  Xr[j,i] <- rbinom(1,1,Pm[j,i]);</pre>
 }
}
Reps[,,r] <- Xr;</pre>
#Dbar
AUX[r] <- -2*sum(log((Pm<sup>X</sup>)*((1-Pm)<sup>(1-X</sup>))));
#Escores
frm <- matrix(nrow=(Q+1)); #Freq na repetição</pre>
for(j in 1:(Q+1))
{
  frm[j] <- length(which(apply(Xr,1,sum)==(j-1)));</pre>
 }
 fr[,r] <- frm;
#p-values
  #Valores marginais
      #Nós de quadratura
      delta <- 1;
      coords <- seq(-3,3,by=delta);</pre>
      coords <- list(coords);</pre>
      eixos <- rep(coords,p);</pre>
      theta <- expand.grid(eixos);</pre>
      theta <- as.matrix(theta);</pre>
      g <- matrix(nrow=nrow(theta));</pre>
      g <- apply(dnorm(theta),1,prod);</pre>
                P <- matrix(nrow=nrow(theta),ncol=Q);</pre>
                for(i in 1:nrow(theta))
                {
                  for(j in 1:Q)
                  {
                    #Modelo compensatorio
  P[j,i] <- exp(am[i,]%*%tr[j,] + bm[i])/(1+exp(am[i,]%*%tr[j,] + bm[i]));</pre>
  #Modelo não-compensatorio
```

```
#P[j,i] <- exp(am[i,1]%*%tr[j,1] + bm[i,1])/(1+exp(am[i,1]%*%tr[j,1] + bm[i,1</pre>
      #exp(am[i,2]%*%tr[j,2] + bm[i,2])/(1+exp(am[i,2]%*%tr[j,2] + bm[i,2]));
                      }
                    }
                   Esp <- matrix(nrow=(Q+1),ncol=nrow(P));</pre>
                   for(j in 1:nrow(P))
                    {
                    Esp[,j] <- Sk(P[j,]);</pre>
                    }
                    fre <- matrix(nrow=(Q+1)); #Escore marginal para cada padrão</pre>
                    for(j in 1:(Q+1))
                    {
                     fre[j] <- N*Esp[j,]%*%g;</pre>
                    }
                   Sobs <- apply(X,1,sum);</pre>
                   Nobs <- matrix(0,nrow=(Q+1));</pre>
                   for(i in 1:(Q+1))
                    {
                     Nobs[i] <- length(which(Sobs==(i-1)));</pre>
                   }
                   X20[r] <- sum(((Nobs-fre)^2)/fre);</pre>
                   X2r[r] <- sum(((frm-fre)^2)/fre);</pre>
}#Fim das repetições
#Dhat
    Ph <- matrix(nrow=N,ncol=Q);</pre>
    ah <- matrix(summary(A)$statistics[,1],nrow=Q,ncol=p,byrow=TRUE);</pre>
    bh <- matrix(summary(B)$statistics[,1],nrow=Q,ncol=1,byrow=TRUE);</pre>
    th <- matrix(summary(Th)$statistics[,1],nrow=N,ncol=p,byrow=TRUE);</pre>
    for(i in 1:Q)
    ſ
     for(j in 1:N)
     {
      #Modelo compensatorio
      Ph[j,i] <- exp(am[i,]%*%tr[j,] + bm[i])/(1+exp(am[i,]%*%tr[j,] + bm[i]));</pre>
      #Modelo não-compensatorio
      #Ph[j,i] <- exp(am[i,1]%*%tr[j,1] + bm[i,1])/(1+exp(am[i,1]%*%tr[j,1] + bm[i,</pre>
      #exp(am[i,2]%*%tr[j,2] + bm[i,2])/(1+exp(am[i,2]%*%tr[j,2] + bm[i,2]));
```

```
}
    }
    Dhat <- -2*sum(log((Ph^X)*((1-Ph)^(1-X))));</pre>
Dbar <- mean(AUX);</pre>
DIC <- 2*Dbar - Dhat;
low <- apply(fr,1,quantile, probs=.05);</pre>
escs <- apply(fr,1,mean);</pre>
up <- apply(fr,1,quantile,probs=.95);</pre>
pvalue <- length(which(X20<X2r))/Rs;</pre>
print("Estatistica de Teste: Escores");
print(c("p-value",round(pvalue,2)));
print(data.frame("Observados"=Nobs,"Média MCMC"=round(escs),
"IC"=cbind("5"=round(low),"95"=round(up))));
print(c("DIC",DIC));
invisible(list( "DIC"=DIC,
"Escores"=cbind(Nobs,round(escs),round(low),round(up)),pvalue));
}
```

Referências Bibliográficas

- Adams, R. J., Wilson, M. & Wang, W. (1997). The multidimensional random coefficients multinomial logit model, *Applied Psychological Measurement* 21: 1–23.
- Albert, H. J. (1992). Bayesian estimation of normal ogive item response curves using gibbs sampling, *Journal of Educational Statistics* 17(3): 251–269.
- Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000). Teoria de resposta ao item: conceitos e aplicações, Associação Brasileira de Estatística, São Paulo.
- Ayreshi, A. (1993). Computing conditional maximum likelihood estimates fir generalized rasch model ysubg sunple loglinear models with diagonal parameters, *Scandinavian Actuarial Journal* 20: 63–71.
- Baker, F. B. & Kim, S. (2004). Item Response Theory Parameter Estimation Techniques, 2nd edn, Marcel Dekker, Inc, New York.
- Bazán, J. L., Branco, D. M. & Bolfarine, H. (2006). A skew item response model., Bayesian Analysis 1(4): 861–892.
- Béguin, A. A. & Glas, C. A. W. (2001). Mcmc estimation and some model-fit analysis in multidimensional irt models, *Psychometrika* 66(4): 541–562.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the em algorithm, *Psychometrika* **46**: 443–459.
- Bock, R. D., Gibbons, R. & Muraki, E. (1988). Full information item factor analysis, Applied Psychological Measurement 12: 261–280.
- Bock, R. D. & Schilling, S. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptative quadrature, *Psychometrika* **70**(3): 533–555.

- Bolt, D. M. & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using markov chain monte carlo, Applied Psychological Measurement 27(6): 395414.
- Carlson, J. E. (1987). Multidimensional item response theory estimation: A computer program, *Technical report*, Iowa City, IA: American College Testing.
- Cohen, A. (2008). The underlying structure of the beck depression inventory ii: A multidimensional scaling approach, *Journal of Research in Personality* 42: 779– 786.
- Divgi, D. (1979). Calculation of univariate and bivariate normal probability functions, The Annals of Statistics **9**: 903–910.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*, Lawrence Erlbaum Associates, Inc., Publishers, New Jersey.
- Frasier, C. (1988). Noharm ii: A fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory, *Technical report*, Armisdale, Australia: The University of New England.
- Gelman, A., Carlin, J. B. & Rubin, D. B. (2004). Bayesian Data Analysis (2nd ed.), Chapman and Hall/CRC, Boca Raton - Florida.
- Gelman, A. & Rubin, D. B. (1992). Inference for iterative simulation using multiple sequences, *Statistical Science* 7: 457–511.
- Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Volume PAMI-6* 6: 721–741.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in J. M. Bernardo, J. O. Berger, A. P. Dawid & S. A. F. M. (eds), *Bayesian Statistics* 4, Clarendon Press, Oxford.
- Johnson, R. & Wichern, D. (2003). *Applied Multivariate Analysis*, Prentice Hall, New Jersey.
- Kang, T. & Cohen, A. S. (2007). Irt model selection methods for dichotomous items, Applied Psychological Measurement 31(4): 331358.
- Kelderman, H. (1996). Loglinear multidimensional item response model for polytomously scored items, in W. J. van der Linden & R. K. Hambleton (eds), Handbook of Modern Item Response Throry, Springer, New York.
- Kelderman, H. & Rijkes, C. P. M. (1994). Loglinear multidimensional irt models for polytomously scored items, *Psychometrika* 59: 144–147.
- Lawley, D. & Maxwell, A. (1971). Factor Analysis as a Statistical Method, Butterworths, London.
- Lee, S. & Terry, R. (2005). Mdirt-fit: Sas® macros for fitting multidimensional item response, *SUGI 31st Conference*.
- Li, Y. H. & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional irt linking, *Applied Psychological Measurement* **24**: 115–138.
- Linden, W. J. v. d. & Hambleton, R. K. (1996). Handbook of modern item response theory, Springer, New York.
- Masters, G. N. & Wright, B. D. (1996). The partial credit model, in W. J. van der Linden & R. K. Hambleton (eds), Handbook of Modern Item Response Throry, Springer, New York.
- McKinley, R. L. (1987). User's guide to multidim, *Technical report*, Princeton, NJ: Educational Testing Service.
- McKinley, R. L. & Reckase, M. D. (1980). The use of the general rasch model with multidimensional item response data, *Technical report*, Iowa City, IA: American College Testing.
- McKinley, R. L. & Reckase, M. D. (1983). Maxlog: A computer program for the estimation of the parameters of a multidimensional logistic model, *Behavior Research Methods and Instrumentation* **91**: 389–390.
- Mislevy, R. J. & Wu, P. K. (1996). Missing responses and irt ability estimation: Omits, choice, time limits and adaptative testing, *Technical report*, Educational Testing Service.
- Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrica* **16**(1): 1–32.
- Nojosa, R. T. (2002). Teoria da reposta ao item modelos multidimensionais, *Estudos de Avaliação Educacional* 25.

- Orlando, M. & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models, *Applied Psychological Measurement* **24**: 50–64.
- Patz, R. J. & Junker, B. W. (1999). A straightforward approach to markov chain monte carlo methods for item response theory models, *Journal of Educational and Behavioral Statistics* 24: 146–178.
- Raftery, A. E. & Lewis, S. M. (1995). The number of iterations, convergence diagnostics and generic metropolis algorithms, in D. S. W.R. Gilks & S. Richardson (eds), *Practical Markov Chain Monte Carlo*, Chapman and Hall, Londres.
- Reckase, M. D. (1996). A linear logistic multidimensional model, *Handbook of Modern Item Response Theory*, Springer, New York.
- Reckase, M. D. (1997). The past and the future of multidimensional item response theory, *Applied Psychological Measurement* **21**: 25–36.
- Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer-Verlag, New York.
- Sheu, C. F., Chen, C. T., Su, Y. H. & Wang, W. C. (2005). Using sas proc nlmixed to fit item response models, *Behavior Research Models* **37**(2): 202–218.
- Sinharay, S. & Lu, Y. (2008). A further look at the correlation between item parameters and item fit statistics, *Journal of Educational Measurement* **45**(1): 115.
- Steer, R. A., Ball, R. & Ranieri, W. F. Beck, A. T. (1999). Dimensions of the beck depression inventory-ii in clinically depressed outpatients, *Journal of Clinical Psychology* 55(1): 117–128.
- Sympson, J. (1977). A model for testing with multidimensional items, *Proc. of the* 1977 Computerized Adaptive Testing Conference.
- te Marvelde, J. M., Glas, C. A., Landeghem, G. & Van Damme, J. (2006). Application of multidimensional item response theory models to longitudinal data, *Educational* and Psychological Measurement 66(1): 5–34.
- Varadhan, R. & Gilbert, P. D. (2009). Bb: An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function, *Journal of Statistical Software* **32**(4).

- Varadhan, R. & Roland, C. (2008). Simple and globally-convergent methods for accelerating the convergence of any em algorithm, *Scandinavian Journal of Statistics* 35(2): 335 – 353.
- Whitely, S. (1980). Measuring aptitude processes with multicomponent latent trait models, *Technical report*, Lawrence: University of Kansas.
- Wilson, D., Wood, R. & Gibbons, R. D. (1987). TESTFACT: Test scoring, item statistics and item factor analysis, Scientific Software, Mooresville.
- Zhang, B. & Stone, C. A. (2008). Evaluating item fit for multidimensional item response models, Applied Psychological Measurement 68(2): 181–196.