



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Diogo Braga Mendes

**Estimação Bayesiana em Modelos Dicotômicos de Resposta ao Item
sob o framework de Modelos Mistos**

Orientador: Prof. Dra. Maria Regina Madruga Tavares
Co-orientador: Prof. Dr. Héilton Ribeiro Tavares

Belém
2012

Diogo Braga Mendes

**Estimação Bayesiana em Modelos Dicotômicos de Resposta ao Item
sob o framework de Modelos mistos**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará, como requisito parcial para a obtenção do grau de Mestre em Estatística.

Orientador: Prof. Dra. Maria Regina Madruga Tavares

Co-orientador: Prof. Dr. Héilton Ribeiro Tavares

Área de Concentração: Estatística

Belém

2012

Diogo Braga Mendes

**Estimação Bayesiana em Modelos Dicotômicos de Resposta ao Item
sob o framework de Modelos Mistos**

Esta Dissertação foi julgada e aprovada, para a obtenção do grau de Mestre em Estatística,
no Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará.

Belém, 08 Março de 2012.

Prof. Dr. Giovany de Jesus Malcher Figueiredo

Coordenador do Programa de Pós-Graduação em Matemática e Estatística da UFPA

Banca Examinadora

Prof. Dra. Maria Regina Madruga Tavares

Universidade Federal do Pará

Orientador

Prof. Dr. Héilton Ribeiro Tavares

Universidade Federal do Pará

Co-orientador

Prof. Dr. Caio Lucidios Naberezny Azevedo

Universidade Estadual de Campinas

Membro

Dedico esta dissertação à Deus, meus pais e minha esposa, pelo força, carinho, incentivo, paciência e confiança, sem os quais nada disso teria sido realizado.

Agradecimentos

- ★ À Deus por ter me dado saúde, força, coragem e sempre iluminando meu caminho;
- ★ À Professora Maria Regina pela orientação e apoio a esta dissertação de mestrado;
- ★ Ao Professor Héilton Tavares pela orientação e apoio a esta dissertação de mestrado;
- ★ Ao Professor Gyovany Figueiredo, coordenador do Programa de Pós-graduação em Matemática e Estatística, profissional exemplar e dedicado;
- ★ Aos professores do Programa de Pós-graduação em Matemática e Estatística pelo conhecimento transmitido durante meu curso de mestrado;
- ★ Aos demais colegas do Departamento de Estatística que contribuíram direta ou indiretamente para realização deste trabalho;
- ★ A minha querida esposa Charlene pelo apoio, confiança e dedicação;
- ★ Aos participantes da banca examinadora pelas sugestões;
- ★ À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro;

”O âmago da ciência, que sempre se apresenta como a mais importante no que refere a resultados práticos, é a pesquisa altamente teórica e abstrata, nascida da infatigável curiosidade, flexibilidade e força da razão humana”.

Andrei Sakharov In Análise Harmônica de Processos Estocásticos

Resumo

MENDES, Diogo Braga. Estimação bayesiana em Modelos dicotômicos de Resposta ao Item sob o framework Modelos Mistos. 2012. Dissertação (Mestrado em Matemática e Estatística), PPGME, UFPA, Belém, Pará, Brasil.

Neste trabalho discutimos a análise bayesiana de Modelos dicotômicos de Resposta ao Item sob o framework de Modelos Mistos. Alguns métodos de estimação baseado em simulações de Monte Carlo via Cadeias de Markov (MCMC) são propostos com a introdução de um modelo de regressão na variável latente e a partir de um esquema de dados aumentados obteve-se distribuições condicionais completas com formas analíticas conhecidas dos parâmetros de interesse. Estudos de simulação foram realizados indicando que a metodologia proposta conduz a estimativas mais fidedignas dos coeficientes de regressão quando comparados com a estimação realizada em duas etapas, ou seja, um modelo de regressão (bayesiano) aplicado aos escores obtidos por um programa tradicional da Teoria da Resposta ao Item (Bilog-MG), além disso constatou-se que o Erro Quadrático Médio (EQM) dos estimadores do traço latente reduz se variáveis adequadamente correlacionadas são incluídas no modelo de regressão para modelagem do traço latente. Por fim, uma aplicação a um banco de dados real, utilizando um dos modelos estudados indicou que a metodologia pode ser empregada com facilidade para a estimação conjunta dos parâmetros do Modelo de Resposta ao Item e dos coeficientes de regressão.

Palavras-chave: Inferência Bayesiana, Teoria da resposta ao item.

Abstract

MENDES , Diogo Braga. Bayesian estimation in Item Response dichotomous Models on framework Mixed Models. 2012. Dissertation (Master Science in Mathematics and Statistics), PPGME, UFPA, Belém, Pará, Brazil.

.

In this work we discuss the the Bayesian analysis of models for dichotomous item response under the framework of Mixed Models. Some estimation methods based on Monte Carlo of Markov Chain (MCMC) simulations are proposed with the introduction of a regression model in the latent variable and from a data augmentation scheme was possible to obtain full conditional distributions with known analytical forms of the parameters of interest. Simulation studies were performed indicating that the proposed methodology leads to more reliable estimates of the regression coefficients when compared with the estimation performed in two stages, for example, a regression model (Bayesian) applied to the scores obtained by a traditional program of the Theory of Item Response (Bilog-MG), Furthermore, it was found that the mean squared error (MSE) of estimators of the latent trait reduce when properly correlated variables are included in the regression model for modeling of the latent trait. Finally, an application to a real database, using one of the models studied indicated that the methodology can be used easily to estimate the parameters of item response model and of the regression coefficients.

Keywords: Bayesian inference, Item response models, Mixed models.

Índice

Resumo	vii
Abstract	viii
Lista de Tabelas	x
Lista de Figuras	xi
1 Introdução	1
1.1 Aspectos Gerais	1
1.2 Objetivos	4
1.2.1 Objetivo geral	4
1.2.2 Objetivos específicos	4
1.3 Estrutura da Dissertação	4
2 TRI sob um enfoque de Modelos Mistos	5
2.1 Introdução	5
2.2 Modelos dicotômicos da TRI como Modelos Mistos	6
2.3 Identificabilidade	8
2.4 Aspectos inferenciais	9
2.4.1 Inferência Bayesiana	10
2.5 Verossimilhança de dados aumentados	11
2.6 Seleção de modelos	12
2.7 Estimação via MCMC com DAGS	14

2.7.1	Modelo Probito-Normal de 2 parâmetros	15
2.7.2	Modelo probito-normal de 3 parâmetros	18
3	Aplicação a dados simulados	20
3.1	Introdução	20
3.1.1	Modelo probito-normal de 2 parâmetros	22
3.1.2	Modelo Probito-Normal de 3 parâmetros	30
4	Aplicações a dados reais	34
4.1	Introdução	34
4.2	Modelo e especificação de prioris	35
4.3	Análise das estimativas	37
5	Considerações Finais e Recomendações	45
5.1	Recomendações	46
A	Anexo A	47
A.1	Amostrador de Gibbs	47
B	Anexo B	49
B.1	Método de Geweke	49
B.2	Método de Gelman & Rubin	50

Lista de Tabelas

3.1	Prioris para análise de sensibilidade	21
3.2	Crítérios de ajuste para o modelo probito-normal de 2 parâmetros: réplica 1	26
3.3	<i>Corr</i> (%) e MEQM para os traços latentes, discriminação e dificuldade dos itens	27
3.4	MVAER para os coeficientes de regressão	27
3.5	Parâmetros dos itens, Estimativas, erros-padrão, IC(95%) e HPD(95%) - Priori A	29
3.6	<i>Corr</i> (%) e MEQM dos parâmetros dos itens estimados: Priori P1	31
3.7	Valores simulados, estimativas de Monte Carlo, erros-padrão, intervalos de confiança e HPD: modelo 3PN, Priori P1-réplica 1	32
3.8	Erro quadrático médio e tempo médio, em segundos (entre parênteses), por cadeia, a cada 1.000 iterações para um teste de 36 itens	33
4.1	Estatísticas descritivas	38
4.2	Comparação dos modelos para as prioris C1, C2 e C3	38
4.3	Resultados para os parâmetros do modelo de regressão da proficiência em Português	41
4.4	Resultados para os parâmetros de discriminação e dificuldade dos itens: Aplicação à dados reais - Priori C2	43
4.5	Resultados para os parâmetros de discriminação e dificuldade dos itens: Aplicação à dados reais - Priori C2	44

Lista de Figuras

3.1	Gráfico de diagnóstico de convergência de Gelman-Rubin, réplica 1 - Priori P1	23
3.2	Simulações para os parâmetros de regressão e dos itens 4 e 20, retidos a cada 30 iterações, réplica 1 - Priori P1	24
3.3	Auto-correlações dos valores simulados para os parâmetros de regressão e dos itens 4 e 20, réplica 1 - Priori P1	25
3.4	Qualidade do ajuste do modelo de regressão e parâmetros dos itens-2PN, réplica 1. (Priori A)	28
3.5	Qualidade do ajuste do modelo de regressão e parâmetros dos itens para o modelo probito-normal de 3 parâmetros: Priori 1, réplica 1	30
4.1	Prioris para o parâmetro de acerto casual	37
4.2	Gráfico dos quantis normal-normal dos resíduos latentes e a Densidade dos resíduos do modelo de regressão	39
4.3	Diagramas de caixa para os resíduos do modelo de regressão por sexo e raça	40
4.4	Gráfico de dispersão dos resíduos do modelo de regressão <i>vs.</i> indicadores I1 e I3	40

Capítulo 1

Introdução

1.1 Aspectos Gerais

Em diversas áreas do conhecimento há necessidade de medir características que não são diretamente observáveis, como por exemplo, nível de ansiedade, inteligência, personalidade, cognição ou impulsividade que são características implícitas do ser humano. Essas características são chamadas de traços latentes e medi-las exige instrumentos adequados como testes de habilidade (provas de conhecimento) ou testes para traços não cognitivos como medida de atitudes e níveis de depressão. Em ambos as questões ou itens devem expressar algum aspecto do processo latente de interesse, veja Skrondal e Rabe-Hesketh (2004) para diversas aplicações com variáveis latentes em ciências sociais.

Entre as diversas metodologias disponíveis para analisar variáveis latentes, a Teoria da Resposta ao Item (TRI) (Hambleton e Swaminathan, 1985) também chamada Teoria dos Traços Latentes, desenvolveu-se inicialmente para suprir deficiências na Teoria Clássica de Medidas (Lord e Novick, 1968) sendo um conjunto de modelos matemáticos, que têm por objetivo estimar estas características com base nas respostas obtidas de um instrumento de avaliação adequado (testes, questionários etc.). Esta teoria permite a modelagem destes construtos com uma função (Curva Característica do Item, CCI) que descreve o comportamento probabilístico da resposta de um sujeito submetido a um teste, relacionando o nível do traço latente e parâmetros (características) dos itens.

Segundo Baker (1992), muitos autores ajudaram na construção da teoria de resposta ao item (TRI), em especial dois desses autores, para os quais é importante citar seus nomes e respectivos trabalhos, Lord e Rasch. O trabalho de Lord (1952) ganhou força nos anos 50 com a aplicação em testes educacionais. Lord iniciou o desenvolvimento formal da teoria, e também contribuiu para o desenvolvimento de programas para computadores necessários

para pôr a teoria em prática, o que levou posteriormente a redigir um livro clássico junto com M. Novick (1968). O matemático dinamarquês Rasch (1960) reforçou os estudos da TRI criando o modelo de 1 parâmetro.

Os modelos da TRI propostos na literatura dependem basicamente da natureza dos itens, isto é, se eles são dicotômicos (corrigidos como certo ou errado) ou não dicotômicos (itens de resposta livre ou itens de múltipla escolha avaliados de forma graduada), da quantidade de traços latentes ou habilidades que estão sendo medidas (uma ou mais de uma) e do número de populações envolvidas (uma ou mais de uma). Todavia, atualmente, modelos mais complexos têm sido desenvolvidos e estes consideram não só as características citadas como também a forma funcional da CCI adotada, estudos longitudinais, etc. Neste trabalho serão considerados apenas aqueles que tratam os itens de forma dicotômica, que avaliam apenas um traço latente (os chamados modelos unidimensionais) e em uma única população. Além disso, supõem-se que as respostas oriundas de indivíduos diferentes são independentes e que os itens são respondidos de forma independente por cada indivíduo, fixada sua habilidade (suposição conhecida como Independência Local).

Após a estimação dos parâmetros dos itens e dos traços latentes várias metodologias têm sido empregadas para análise das estimativas destes parâmetros e especialmente dos traços latentes. Por exemplo, Pereira (2004) com o objetivo de criar alternativas ao Critério Brasil ou Critério de Classificação Econômica Brasil (CCEB) utilizou modelos da TRI de resposta gradual para construir indicadores da condição socioeconômica. A partir das estimativas dos traços latentes, o autor ajustou um modelo clássico de regressão linear tendo como variável dependente os escores do Critério Brasil e a variável independente os escores da TRI, concluindo que os escores produzidos pelo critério Brasil e os escores da TRI estavam altamente correlacionados.

Soares e Mendonça (2003) construíram um modelo de regressão hierárquico para os dados do SIMAVE/2000, onde a população foi composta pelos alunos da 4ª série que participaram da avaliação de matemática e português. Os autores concluíram que, entre outros, o desempenho escolar (medido pelos escores obtidos nestas disciplinas via TRI) sofre um impacto positivo de variáveis que medem o nível de escore socioeconômico, os anos de estudo da mãe, e um impacto negativo de variáveis que medem a defasagem escolar e anos de abandono da escola.

Outro trabalho que utiliza os escores estimados pela TRI é o de Jesus e Laros (2004), que construíram um modelo explicativo do desempenho acadêmico de alunos da 8ª série, na disciplina de português, através de um modelo regressão multinível, para contemplar a estrutura hierárquica dos dados do Sistema de Avaliação da Educação Básica (SAEB) do ano de 2001. Os autores concluíram que os fatores mais determinantes no desempenho escolar nesta disciplina foram o nível socioeconômico médio da escola e o nível de escolaridade médio dos pais.

Silva et al. (2010) calcularam escores relacionados a alguns fatores determinantes da proficiência média em português e matemática que foram estimadas pela Teoria da Resposta ao Item. Estes fatores foram construídos utilizando o Modelo de Resposta ao Item Gradual de Samejima (1969), a partir de respostas aos questionários contextuais (aplicado pelo INEP*/MEC) por alunos da 4ª série do ensino fundamental de escolas públicas municipais em Belém no estado do Pará, seus professores e diretores das escolas. Em um segundo momento, uma análise de agrupamentos foi aplicada para definir possíveis grupos homogêneos de escolas para os fatores considerados.

Pelos trabalhos supracitados é comum a análise posterior dos traços latentes estimados pelos modelos da TRI. Um ponto importante é o fato de que a variável latente utilizada é não observável, ou seja, a incerteza decorrente do seu processo de estimação não é incorporada nessas análises uma vez que as estimativas dos traços latentes são consideradas como verdadeiras. Neste trabalho será desenvolvido um método sob a perspectiva bayesiana com a **estimação conjunta** dos escores latentes e dos coeficientes de regressão para modelos dicotômicos da TRI e, desta forma, expressar um Modelo de Resposta ao Item (MRI) sob formas conhecidas de modelagens bem fundamentadas na literatura científica, especificamente uma estrutura de modelos mistos, pode facilitar sua aplicação em situações diversas.

* Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

1.2 Objetivos

1.2.1 Objetivo geral

- O objetivo geral deste trabalho é desenvolver métodos de estimação bayesiana para os parâmetros de alguns modelos de resposta ao item, introduzindo modelos de regressão na variável latente de interesse, considerando uma abordagem de modelos mistos.

1.2.2 Objetivos específicos

- Introduzir os Modelos Lineares Generalizados Mistos (MLGM) e os aspectos inferenciais para estimação dos parâmetros de interesse via Teoria da Resposta ao Item.
- Apresentar estudos de simulação que validem o processo de estimação proposto.
- Aplicar a metodologia em um banco de dados reais.

1.3 Estrutura da Dissertação

Esta dissertação encontra-se dividida em 5 capítulos, a saber:

Capítulo 1: Refere-se à introdução do trabalho, onde são abordados os aspectos gerais, importância do trabalho, objetivo geral e específicos, bem como a estrutura do trabalho.

Capítulo 2: Apresenta uma revisão sob modelo probito-normal na Teoria da Resposta ao Item, com e sem modelos de regressão latente univariados, bem como sua estimação bayesiana via MCMC com esquema de dados aumentado.

Capítulo 3: Realização de estudos de simulação.

Capítulo 4: Aplicação a dados reais.

Capítulo 5: Considerações finais e propostas para trabalhos futuros;

Capítulo 2

TRI sob um enfoque de Modelos Mistos

2.1 Introdução

Os primeiros modelos da TRI surgiram por volta da década de 50, e eram modelos que consideravam que um único constructo, de um único grupo, estava sendo mensurado por um teste com itens corrigidos de forma dicotômica. Estes modelos foram primeiramente desenvolvidos baseados na distribuição acumulada da normal e, depois, foram descritos para uma forma matemática mais conveniente, a logística, pois sua função de distribuição acumulada é uma função explícita dos parâmetros do MRI.

Desde então, vários Modelos de Resposta ao Item foram propostos para contemplar situações cada vez mais complexas que vão além da estimação dos parâmetros dos itens e traços latentes. Desta forma, construir um modelo adequado para cada situação seria o ideal. Entretanto visualizar o MRI como membro de classes mais gerais de modelos pode trazer benefícios como, por exemplo, a família dos modelos de Rasch que pertencem à classe dos Modelos Lineares Generalizados Mistos (MLGM), veja, por exemplo, (McCulloch & Searle, 2001), o que possibilita a utilização de métodos de estimação, medidas de diagnóstico e pacotes estatísticos existentes para a sua aplicação.

Em muitas situações práticas o interesse não reside apenas no nível individual do traço latente, mas sim em relacionar as respostas de um teste aplicado a outras variáveis (características) dos indivíduos examinados ou mesmo com as características dos itens. Por exemplo, poderíamos estar interessados em quantificar a relação entre o nível socioeconômico de alunos de uma rede de ensino e o desempenho médio em determinada área do conhecimento.

2.2 Modelos dicotômicos da TRI como Modelos Mistos

Dados dicotômicos, ou seja, do tipo certo ou errado, concordo ou discordo etc. são frequentes em diversas áreas do conhecimento, e podem ser vistos com variáveis observáveis a partir do truncamento de variáveis latentes. Neste caso, modelos com um parâmetro por item podem ser aplicados, este parâmetro foi nomeado dificuldade do item, indicando o nível traço latente para qual o sujeito tem alta probabilidade de responder como esperado ao item. Rasch (1960) desenvolveu esse modelo considerando uma CCI logística e extensões desse modelo considerando outras características dos itens como discriminação (Birnbaum, 1957) e acerto casual foram desenvolvidas, este último muito utilizado em avaliações educacionais.

O modelo dicotômico da TRI com o parâmetro de acerto casual pode ser definido da seguinte forma:

$$Y_{ji}|\theta_j, \zeta_i \sim \text{Bernoulli}(P_{ji}), \quad (2.1)$$

onde Y_{ji} são valores correspondentes as respostas dicotômicas do indivíduo j no item i , assumindo o valor 1 para indicar o acerto e 0 caso contrário, P_{ji} é uma função dos parâmetros θ_j e ζ_i^t , em que $\zeta_i^t = (a_i, b_i, c_i)$, onde a_i , b_i e c_i são os parâmetros de discriminação, dificuldade e acerto casual do item respectivamente, e θ_j é o valor correspondente da variável latente, associada ao indivíduo j .

A probabilidade condicional de que o indivíduo j responda corretamente ao item i , $j=1,2,\dots,n$, $i=1,2,\dots,I$, é dada por:

$$P_{ji} = P(Y_{ji} = 1|\theta_j, \zeta_i) = c_i + (1 - c_i)F(\eta_{ji}), \quad (2.2)$$

com $\eta_{ji} = a_i\theta_j - b_i$.

A função $F(\cdot)$, em 2.2 é dada por uma função de distribuição acumulada chamada de Função de Resposta ao Item ou Curva Característica do Item, que neste modelo é estritamente não decrescente em θ_j . Caso $F(\cdot)$ seja a função de distribuição acumulada logística chamamos o modelo de **logito**-normal de 3 parâmetros e, se a distribuição é a função de distribuição acumulada normal padrão, será denominado de **probit**-normal, denominações consideradas em Bázan (2005).

A suposição de independência condicional é feita para a estimação dos parâmetros dos

itens, ou seja, para um indivíduo j , as respostas Y_{ji} , $i = 1, 2, \dots, I$, são condicionalmente independentes dados os valores das variáveis latentes θ_j , $j = 1, 2, \dots, n$, ou seja, o indivíduo não “aprende” ao longo da realização do teste. Adicionalmente, a independência das respostas de diferentes indivíduos também é considerada.

Os parâmetros dos itens estão descritos a seguir:

- a_i é o parâmetro de discriminação (ou, inclinação) do item i , com valor proporcional à inclinação da CCI no ponto b_i ;
- b_i é o parâmetro de dificuldade (ou de posição) do item i medido na mesma escala do traço latente θ ;
- c_i é o parâmetro do item que representa a probabilidade do indivíduo com baixo nível do traço latente responder corretamente ao item i , ou seja, c_i representa a probabilidade de acerto casual.

Muitos Modelos de Resposta ao Item podem ser analisados como membros de outras classes de modelos estatísticos, dentre elas, pode-se citar a estrutura de modelos não lineares mistos (MNLM). De acordo com Rijmen et al. (2003), Modelos Mistos são uma ferramenta útil em análise de dados agrupados como, por exemplo, dados de estudantes aninhados com escolas ou dados de medidas repetidas sobre o mesmo aluno.

Esses modelos pertencem a família de modelos lineares e não-lineares mistos, cuja a formulação geral pode ser

$$\begin{aligned} Y &\sim \mathbf{f}(\boldsymbol{\mu}, \phi) \\ g(\boldsymbol{\mu}) &= h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) \\ \mathbf{b} &\sim N(0, \sigma^2) \end{aligned} \tag{2.3}$$

Onde \mathbf{f} é uma função densidade de probabilidade conhecida, usualmente pertencente a família exponencial (FE) de modelos, μ é a esperança matemática da variável resposta Y , g é uma função de ligação, ϕ é o parâmetro de dispersão, $\boldsymbol{\beta}$ são os coeficientes de regressão (efeitos fixos), \mathbf{b} são os efeitos aleatórios, X e Z são matrizes de delineamento conhecidas.

2.3 Identificabilidade

Os modelos apresentados na seção 2.2 não são identificáveis no sentido de que a verossimilhança dos dados pode ser reproduzida por diferentes conjuntos de parâmetros, em outras palavras, os estimadores de máxima verossimilhança podem não ser únicos. Isto deve-se ao fato de que a escala do traço latente θ não está definida. No caso do modelo 2PN, a probabilidade de resposta correta é

$$\begin{aligned} P(Y_{ji} = 1 | \theta_j, \zeta_i) &= \Phi(a_i(\theta_j - b_i)) = \Phi\left(\frac{a_i}{\sigma_\theta}(\sigma_\theta\theta_j - \sigma_\theta b_i)\right) \\ &= \Phi\left(\frac{a_i}{\sigma_\theta}(\sigma_\theta\theta_j - k - \sigma_\theta b_i + k)\right) \\ &= \Phi(a_i^*(\theta_j^* - b_i^*)), \end{aligned}$$

com $a_i^* = \frac{a_i}{\sigma_\theta}$, $\theta_j^* = \sigma_\theta\theta_j - k$ e $b_i^* = \sigma_\theta b_i - k$. Assim, a locação e escala da distribuição de probabilidade para θ pode ser arbitrariamente especificada.

No caso de um modelo de regressão para variável latente, tem-se

$$\begin{aligned} \theta_j &\sim N(\mathbf{x}'_j\boldsymbol{\beta}, \sigma_\theta^2) \\ \theta_j^* &= \frac{\theta_j}{\sigma_\theta} \\ &\sim N\left(\frac{\mathbf{x}'_j\boldsymbol{\beta}}{\sigma_\theta}, 1\right). \end{aligned}$$

A probabilidade de resposta correta pode ser reescrita como

$$\begin{aligned} P(Y_{ji} = 1 | \theta_j, \zeta_i) &= \Phi(a_i(\theta_j - b_i)) = \Phi\left(\sigma_\theta a_i\left(\frac{\theta_j}{\sigma_\theta} - \frac{b_i}{\sigma_\theta}\right)\right) \\ &= \Phi(a_i^*(\theta_j^* - b_i^*)), \end{aligned}$$

com $a_i^* = \sigma_\theta a_i$, $\theta_j^* = \frac{\theta_j}{\sigma_\theta}$ e $b_i^* = \frac{b_i}{\sigma_\theta}$. Então, para o modelo com traço latente transformado θ^* , o parâmetro de discriminação é multiplicado por σ_θ e o parâmetro de dificuldade dividido por σ_θ e, desta forma, o modelo não é identificável a menos que um dos parâmetros de discriminação seja igualado a um ou que a variância do traço latente seja restrita a uma constante conhecida.

Na abordagem bayesiana via MCMC, a métrica pode ser garantida de diversas formas. Em modelos da TRI nulos, ou seja, sem modelos de regressão para o traço latente, é suficiente restringir os hiperparâmetros das prioris para uma métrica específica, usualmente

$\mu_\theta = 0$ e $\sigma_\theta^2 = 1$. Alternativamente, em todas as iterações do MCMC, por uma transformação linear simples, as amostras do traço latente podem ser reescaladas para uma métrica arbitrária, de forma que nos próximos passos da amostragem, porém na mesma iteração, o mesmo vetor reescalado é utilizado, garantindo dessa forma a identificabilidade do modelo. Este artifício é de grande ajuda principalmente em modelos complexos, onde a especificação de prioris em múltiplos estágios dificulta a identificabilidade do modelo.

2.4 Aspectos inferenciais

Os problemas de estimação podem ser agrupados em três categorias: estimação frequentista (ou clássica), estimação Bayesiana sem MCMC (*Monte Carlo Markov Chain*) ou não plena e a estimação Bayesiana com MCMC (análise bayesiana plena). Na abordagem clássica, a estimação por máxima verossimilhança têm sido a mais utilizada e várias propostas foram formuladas, tais como: a verossimilhança conjunta, verossimilhança marginal (MVM) e verossimilhança condicional (Andrade et al., 2000, Baker, 1992). Todavia, como Bock and Aitkin (1981) sinalizaram, estimativas de máxima verossimilhança para parâmetros estruturais não são consistentes na presença de parâmetros incidentais (Neyman e Scott, 1948). Como consequência, a abordagem mais usada é o método da verossimilhança marginal que usa um algoritmo pseudo-EM com quadraturas gaussianas para aproximar as integrais que são necessárias na implementação da etapa E do algoritmo (Book e Aitkin, 1981). Este procedimento de estimação está implementado em software específicos como, por exemplo, no programa BILOG-MG (Mislevy e Bock, 1990), o que usa uma restrição "heurística" do modelo (Du Toit, 2003). A estimação dos parâmetros de habilidade é executada numa segunda etapa com os parâmetros de itens substituídos pelas suas estimativas obtidas previamente. Limitações desta metodologia são discutidas em Patz e Junker (1999) e Sahu (2002). Para citar algumas, tem-se a impossibilidade de estimar os parâmetros do modelo no caso de escores extremos (acerto ou erro total, por parte dos itens e/ou sujeitos).

2.4.1 Inferência Bayesiana

Na estatística, toda informação que se tem sobre uma certa quantidade de interesse ϕ (não observável), se torna fundamental para continuação do estudo. Devido o verdadeiro valor de ϕ ser desconhecido, o objetivo é tentar reduzir esta nossa falta de conhecimento. Além disso, a intensidade do desconhecimento a respeito de ϕ pode assumir diferentes graus. Na inferência Bayesiana, estes diferentes graus de desconhecimento são representados através da adoção de modelos probabilísticos para ϕ . Assim, é natural que o grau de conhecimento a cerca de ϕ não seja o mesmo para todos os pesquisadores, ou seja, é comum a especificação de modelos distintos.

Considere que a informação de que dispomos sobre uma certa quantidade de interesse desconhecida ϕ , resumida probabilisticamente através de $\pi(\phi)$, pode ser aumentada observando-se uma quantidade aleatória Θ relacionada com ϕ . A distribuição amostral $p(\theta|\phi)$ define esta relação. A idéia de que após observarmos $\Theta = \tau$, a quantidade de informação a respeito de ϕ aumento é bastante intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação,

$$p(\phi|\tau) = \frac{p(\tau|\phi)\pi(\phi)}{p(\tau)} = \frac{p(\tau|\phi)\pi(\phi)}{\int p(\tau, \phi)d\phi} \quad (2.4)$$

onde $p(\tau) = [\int p(\tau, \phi)d\phi]^{-1}$ é a distribuição marginal de θ .

Para um valor fixo de τ , a função $L(\phi) = p(\tau|\phi)$ fornece a plausibilidade ou a verossimilhança de cada um dos valores de ϕ , enquanto $\pi(\phi)$ é a chamada distribuição a priori de ϕ . Estas duas fontes de informação, priori e verossimilhança, são combinadas levando à distribuição a posteriori de ϕ , $p(\phi|\tau)$, a qual será usada para fazer inferências. Assim, a forma usual do teorema de Bayes é

$$p(\phi|\tau) \propto L(\phi) \times \pi(\phi) \quad (2.5)$$

A estimação bayesiana consiste em estimar os parâmetros com base em alguma característica da distribuição a posteriori, como por exemplo, a média, mediana, moda, máximo, mínimo e etc. Na TRI, para a estimação bayesiana sem MCMC, o principal método é a Estimação Bayesiana Marginal (EBM) que é uma extensão da proposta de Bock and Aitken (1981). Dos métodos de estimação bayesiana baseado em simulações, existem dois principais métodos de amostragem MCMC: Metropolis-Hastings (M-H) desenvolvido por

Metropolis et al. (1953) e generalizado por Hastings (1970), tendo sido considerado como um dos dez algoritmos que mais contribuíram ao desenvolvimento da ciência e tecnologia do Século XX (Beichl e Sullivan, 2000) e o amostrador de Gibbs (Geman e Geman, 1984).

Métodos MCMC são um conjunto de algoritmos de simulação que constroem uma cadeia de Markov cuja distribuição limite (estacionária) é uma distribuição multivariada de interesse, comumente desconhecida.

O algoritmo de Gibbs, é essencialmente um esquema iterativo de amostragem de uma cadeia de Markov cujo núcleo é formado pelas distribuições condicionais completas (distribuições marginais da posteriori com forma funcional conhecida), e sob certas condições de regularidade, a distribuição limite é a densidade de interesse, veja Gamerman e Lopes (2006).

Quando as distribuições marginais da posteriori não tem forma funcional conhecida ou de fácil simulação, o algoritmo de Gibbs não pode ser diretamente aplicável, desta forma uma solução é o algoritmo M-H que explora a superfície da posteriori de interesse em dois passos: (1) a partir da especificação de uma densidade proposta, candidatos são simulados e devem estar próximos da distribuição alvo e (2) estes são aceitos ou rejeitados de acordo com um critério. No presente trabalho, utilizaremos o amostrador de Gibbs (veja o apêndice A) com um esquema de dados aumentados proposto inicialmente por Albert (1992) utilizando variáveis auxiliares.

2.5 Verossimilhança de dados aumentados

A verossimilhança desempenha papel fundamental em qualquer análise estatística e todos os processos citados anteriormente baseiam-se nesta função. Basicamente, ela pode ser definida para o tipo de dados e modelos considerados neste trabalho como

$$L(\theta, \zeta | D_{obs}; Y) = \prod_{i=1}^k \prod_{j=1}^n P_{ji}^{Y_{ji}} (1 - P_{ji})^{1 - Y_{ji}}, \quad (2.6)$$

com P_{ji} sendo modelada de acordo com a CCI adotada, ademais a verossimilhança pode ser vista como a densidade conjunta das observações condicionada nos parâmetros do modelo e θ é o parâmetro de interesse.

Considere o preditor linear $\eta_{ji} = a_i\theta_j - b_i$ e $Y_{ji} \sim \text{Bernoulli}(\Phi(\eta_{ji}))$ com

$$Y_{ji}|\theta_j, b_i = \begin{cases} 1, & \text{se } Z_{ji} > 0, \\ 0, & \text{se } Z_{ji} \leq 0, \end{cases} \quad (2.7)$$

sendo $Z_{ji}|\theta_j, \zeta_i \sim \mathcal{N}(\eta_{ji}, 1)$ e $\epsilon_{ji} = Z_{ji} - \eta_{ji}$, então

$$\begin{aligned} P(Z_{ji} > 0|\theta_j, b_i) &= P(\epsilon_{ji} > -\eta_{ji}|\theta_j, \zeta_i) \\ &= 1 - \Phi(-\eta_{ji}) \\ &= \Phi(\eta_{ji}). \end{aligned} \quad (2.8)$$

Desta forma, as probabilidades de sucesso de Y_{ji} são iguais e portanto as inferências baseadas em Z_{ji} serão as mesmas.

Para o modelo 3PN, a verossilhança aumentada é obtida da forma similar. Entretanto é necessário definir uma variável auxiliar adicional para representar o estado de ignorância do sujeito com respeito ao item apresentado no teste. O desenvolvimento aqui apresentado segue os mesmos procedimentos adotados por Béguin & Glass (2001), considere uma nova variável auxiliar

$$S_{ji} = \begin{cases} 1, & \text{se o sujeito } j \text{ sabe a resposta correta do item } i, \\ 0, & \text{se não sabe.} \end{cases} \quad (2.9)$$

A probabilidade de resposta correta pode ser calculada a partir de

$$\begin{aligned} P(Y_{ji} = 1|\zeta_i, \theta_j) &= P(Y_{ji} = 1|\zeta_i, \theta_j, S_{ji} = 1)P(S_{ji} = 1) \\ &+ P(Y_{ji} = 1|\zeta_i, \theta_j, S_{ji} = 0)P(S_{ji} = 0) \\ &= \Phi(\eta_{ji}) + c_i(1 - \Phi(\eta_{ji})), \end{aligned} \quad (2.10)$$

isto é, a probabilidade é a mesma na equação ?? e, portanto, espera-se que as inferências baseadas na verossilhança aumentada e original sejam próximas.

2.6 Seleção de modelos

Quando vários modelos estão disponíveis para mesma aplicação é importante compará-los para selecionar o mais parcimonioso. Na estimação clássica os critérios de informação

de Akaike-AIC (Akaike, 1973) e o de informação bayesiana, BIC (Schwarz), são os mais utilizados e, basicamente, penalizam a verossimilhança considerando a complexidade do modelo. Na estimação, pode-se dizer que estes métodos são aproximações do conhecido Fator de Bayes bastante utilizado em testes de hipóteses bayesianos, mas que não pode ser aplicado diretamente quando prioris impróprias ou vagas são consideradas, situação comum em modelos complexos com esquemas hierárquicos.

Para superar a limitação do Fator de Bayes (BF), outros critérios foram sugeridos, dentre eles o critério de informação de desvio (DIC), descrito em Spiegelhalter et al. (2002) que independe das prioris do modelo e do número de parâmetros, tendo uma abordagem preditiva a posteriori e é facilmente obtido a partir do amostras geradas pelos métodos MCMC.

Antes de defini-lo, é necessário introduzir uma medida da bondade de ajuste de modelos criada por Dempster (1977), chamada **desvio bayesiano** definido pelo simétrico do logaritmo da verossimilhança do modelo. Por exemplo, no modelo probito de 2 parâmetros é calculado da seguinte forma:

$$\begin{aligned}
 D(\boldsymbol{\theta}, \boldsymbol{\zeta}) &= -2\log L(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\zeta}) \\
 &= -2\log \prod_{j,i} [\Phi(\eta_{ji})^{y_{ji}} (1 - \Phi(\eta_{ji}))^{1-y_{ji}}] \\
 &= -2 \sum_{j,i} [y_{ji} \log(\Phi(\eta_{ji})) + (1 - y_{ji}) \log((1 - \Phi(\eta_{ji})))]
 \end{aligned} \tag{2.11}$$

com $\eta_{ji} = a_i \theta_j - b_i$.

Para uma amostra da distribuição a posteriori conjunta obtida por MCMC, pode-se usar o conceito desvio bayesiano esperado $E[D(\boldsymbol{\theta}, \boldsymbol{\zeta})]$, que neste caso pode ser aproximado por

$$\bar{D} = E[D(\boldsymbol{\theta}, \boldsymbol{\zeta})] \approx \sum_{m=1}^M D(\boldsymbol{\theta}^m, \boldsymbol{\zeta}^m) / M$$

Com m sendo a m -ésima amostra simulada a partir da posteriori conjunta de um total de M , segundo Bazán (2005) está implementada por default no Winbugs.

Possivelmente modelos mais complexos (mais parâmetros) terão menor desvio esperado,

Carlin e Louis (2000) e Brooks (2002) propuseram penalizá-lo conforme os critérios AIC e BIC, resultando no critério de informação de Akaike esperado (EAIC):

$$EAIC = E[D(\boldsymbol{\theta}, \boldsymbol{\zeta})] + 2p,$$

e o critério de bayesiano esperado (EBIC)

$$EBIC = E[D(\boldsymbol{\theta}, \boldsymbol{\zeta})] + p \log(N),$$

onde p é o número de parâmetros no modelo e N é o número total de amostras. As respectivas aproximações baseadas em \bar{D} são

$$E\hat{AIC} = \bar{D} + 2p \text{ e } E\hat{BIC} = \bar{D} + p \log(N)$$

Finalmente o DIC penaliza a verossimilhança a partir do número de parâmetros efetivos $p_D = E[D(\boldsymbol{\theta}, \boldsymbol{\zeta})] - D(E(\boldsymbol{\theta}), E(\boldsymbol{\zeta}))$ sendo o segundo termo o desvio bayesiano nas médias a posteriori $E(\boldsymbol{\theta})$ e $E(\boldsymbol{\zeta})$ que pode ser aproximado por

$$D(E(\boldsymbol{\theta}), E(\boldsymbol{\zeta})) \approx D\left(\sum_{m=1}^M \boldsymbol{\theta}^{(m)} / M, \sum_{m=1}^M \boldsymbol{\zeta}^{(m)} / M\right)$$

Segundo Spiegelhalter et al. (2002), o DIC pode ser usado concomitantemente com BF, uma vez que, o BF pode detectar conflitos entre a informação a priori e verossimilhança, fato muito importante, principalmente na Teoria da Resposta ao Item, cujos modelos frequentemente possuem prioris hierárquicas que podem influenciar substancialmente os resultados.

2.7 Estimação via MCMC com DAGS

Frequentemente as posteriores dos parâmetros de interesse são analiticamente intratáveis e métodos numéricos são necessários para obtenção de estimativas. Albert e Chibbs (1992) foram pioneiros em utilizar dados aumentados para obter distribuições condicionais completas conhecidas que possibilitassem a utilização do algoritmo de Gibbs naturalmente.

2.7.1 Modelo Probito-Normal de 2 parâmetros

A seguinte estrutura hierárquica foi considerada para estimação dos parâmetros do modelo probito-normal de 2 parâmetros com um modelo de regressão para o traço latente com p covariáveis:

$$\begin{aligned} \eta_{ji} &= a_i\theta_j - b_i, \\ \theta_j | \sigma_\theta^2 &\sim \mathcal{N}(\mathbf{x}_j^t \boldsymbol{\beta}, \sigma_\theta^2), \\ p(\beta_k) &\propto \text{const. ou } \boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \\ \boldsymbol{\zeta}_i = (a_i, b_i)^t | \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta &\sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0,5 & 0 \\ 0 & 2 \end{bmatrix}\right) \mathbb{I}(a_i > 0), \end{aligned} \quad (2.12)$$

onde $\mathbb{I}(a_i > 0)$ indica que os valores de a_i assumem somente valores positivos, *const.* é uma constante, $\mathbf{x}_j^t = (x_{j1}, x_{j2}, \dots, x_{jp})$ e $\boldsymbol{\beta}^t = (\beta_1, \beta_2, \dots, \beta_p)$.

Conforme o modelo acima, a priori proposta por Sahu (2002), Patz e Junker (1999) para os parâmetros dos itens foi utilizada (observe que a_i e b_i são independentes), Bázan (2005) após um estudo de robustez à especificação de 5 prioris para os parâmetros dos itens, encontrou resultados similares no que tange a estimação dos parâmetros, entretanto esta priori apresentou resultados ligeiramente superiores. A abordagem matricial será considerada aqui para subsidiar outras opções como uma reparmetrização do modelo fazendo $a_i = e^{\tilde{a}_i}$ com $\tilde{a}_i \sim \mathcal{N}(\mu_a, \sigma_a)$ para o parâmetro de discriminação que possibilita a modelagem de correlação entre os parâmetros de cada item, veja Fox (2010) para mais detalhes. Para estimação bayesiana dos parâmetros do modelo 2PN via dados aumentados, considere os passos:

Passo 1: Amostrar Z_{ji} a partir de:

$$Z_{ji} | \theta_j, \zeta_i \sim \begin{cases} \mathcal{N}(a_i\theta_j - b_i, 1) \mathbb{I}(Z_{ji} > 0), & \text{se } Y_{ji} = 1, \\ \mathcal{N}(a_i\theta_j - b_i, 1) \mathbb{I}(Z_{ji} \leq 0), & \text{se } Y_{ji} = 0. \end{cases} \quad (2.13)$$

Passo 2: Amostrar θ_j a partir de:

$$\theta_j | \mathbf{Z}_j, \boldsymbol{\beta}, \mathbf{b}, \sigma_\theta^2 \sim \mathcal{N}(\mu_\theta^*, \sigma_\theta^*)$$

onde

$$\sigma_\theta^* = \left(\sum_{i=1}^I a_i^2 + \frac{1}{\sigma_\theta^2} \right)^{-1} \quad (2.14)$$

$$\mu_\theta^* = \sigma_\theta^* \left(\sum_{i=1}^I a_i (Z_{ji} + b_i) + \frac{\mathbf{x}_j^t \boldsymbol{\beta}}{\sigma_\theta^2} \right) \quad (2.15)$$

Passo 3: Amostrar $\boldsymbol{\beta}$. Duas priors foram consideradas aqui, uma imprópria e outra normal multivariada, para a priori imprópria

$$\boldsymbol{\beta} | \boldsymbol{\theta}, \mathbf{X} \sim \mathcal{N}((\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\theta}, (\mathbf{X}^t \mathbf{X})^{-1} \sigma_\theta^2) \quad (2.16)$$

Para a priori normal multivariada

$$\boldsymbol{\beta} | \boldsymbol{\theta}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_\beta^*, \boldsymbol{\Sigma}_\beta^*)$$

onde

$$\boldsymbol{\Sigma}_\beta^* = \left(\frac{\mathbf{X}^t \mathbf{X}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \right)^{-1} \quad (2.17)$$

$$\boldsymbol{\mu}_\beta^* = \boldsymbol{\Sigma}_\beta^* \left(\frac{\mathbf{X}^t \boldsymbol{\theta}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta \right) \quad (2.18)$$

Passo 4: Amostrar os parâmetros de discriminação e dificuldade dos itens, $\zeta_i = (a_i, b_i)^T$

$$\zeta_i | \mathbf{Z}_i, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_\zeta^*, \boldsymbol{\Sigma}_\zeta^*) I(a_i > 0)$$

Onde

$$\mathbf{w} = (\boldsymbol{\theta}, -\mathbf{1}_n) \quad (2.19)$$

$$\boldsymbol{\Sigma}_\zeta^* = (\mathbf{w}^t \mathbf{w} + \boldsymbol{\Sigma}_\zeta^{-1})^{-1} \quad (2.20)$$

$$\boldsymbol{\mu}_\zeta^* = \boldsymbol{\Sigma}_\zeta^* (\mathbf{w}^t \mathbf{Z}_i + \boldsymbol{\Sigma}_\zeta^{-1} \boldsymbol{\mu}_\zeta) \quad (2.21)$$

$$\mathbf{Z}_i^t = (Z_{1i}, Z_{2i}, \dots, Z_{Ni})$$

A seguir as demonstrações dos passos 1 a 4, que correspondem as condicionais completas

para o modelo probito-normal de 2 parâmetros.

$$\begin{aligned}
p(\theta_j | \mathbf{Z}_j, \boldsymbol{\beta}, \mathbf{b}, \sigma_\theta^2) &\propto p(\mathbf{Z}_j | \theta_j, \boldsymbol{\beta}, \mathbf{b}) p(\theta_j | \sigma_\theta^2) \\
&\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^I (Z_{ji} - (a_i \theta_j - b_i))^2 \right\} \\
&\times \exp \left\{ -\frac{1}{2\sigma_\theta^2} (\theta_j - \mathbf{x}_j^t \boldsymbol{\beta})^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\left(\sum_{i=1}^I a_i^2 + \frac{1}{\sigma_\theta^2} \right) \theta_j^2 - 2\theta_j \left(\sum_{i=1}^I a_i (Z_{ji} + b_i) + \frac{\mathbf{x}_j^t \boldsymbol{\beta}}{\sigma_\theta^2} \right) \right] \right\} \\
&\propto \mathcal{N}(\mu_\theta^*, \sigma_\theta^{*2})
\end{aligned} \tag{2.22}$$

Onde

$$\sigma_\theta^* = \left(\sum_{i=1}^I a_i^2 + \frac{1}{\sigma_\theta^2} \right)^{-1} \tag{2.23}$$

$$\mu_\theta^* = \sigma_\theta^* \left(\sum_{i=1}^I a_i (Z_{ji} + b_i) + \frac{\mathbf{x}_j^t \boldsymbol{\beta}}{\sigma_\theta^2} \right) \tag{2.24}$$

$$\begin{aligned}
p(\boldsymbol{\beta} | \mathbf{Z}, \boldsymbol{\theta}, \mathbf{b}, \mathbf{X}) &\propto p(\boldsymbol{\theta} | \boldsymbol{\beta}) p(\boldsymbol{\beta}) \\
&\propto \exp \left\{ -\frac{1}{2\sigma_\theta^2} (\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta})^t (\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta}) \right\} \\
&\propto \mathcal{N}((\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\theta}, (\mathbf{X}^t \mathbf{X})^{-1} \sigma_\theta^2)
\end{aligned} \tag{2.25}$$

especificando uma priori normal multivariada para $\boldsymbol{\beta}$, sua distribuição condicional completa fica:

$$\begin{aligned}
p(\boldsymbol{\beta} | \mathbf{Z}, \boldsymbol{\theta}, \mathbf{b}, \mathbf{X}) &\propto p(\boldsymbol{\theta} | \boldsymbol{\beta}) p(\boldsymbol{\beta}) \\
&\propto \exp \left\{ -\frac{1}{2\sigma_\theta^2} (\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta})^t (\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta}) \right\} \\
&\times \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^t \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\beta}^t \left(\frac{\mathbf{X}^t \mathbf{X}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \right) \boldsymbol{\beta} - 2\boldsymbol{\beta}^t \left(\frac{\mathbf{X}^t \boldsymbol{\theta}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta \right) \right) \right\} \\
&\propto \mathcal{N}(\boldsymbol{\mu}_\beta^*, \boldsymbol{\Sigma}_\beta^*)
\end{aligned} \tag{2.26}$$

onde

$$\boldsymbol{\Sigma}_\beta^* = \left(\frac{\mathbf{X}^t \mathbf{X}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \right)^{-1} \tag{2.27}$$

$$\boldsymbol{\mu}_\beta^* = \boldsymbol{\Sigma}_\beta^* \left(\frac{\mathbf{X}^t \boldsymbol{\theta}}{\sigma_\theta^2} + \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta \right) \tag{2.28}$$

$$\begin{aligned}
p(\zeta_i | \mathbf{Z}_i, \boldsymbol{\theta}) &\propto p(\mathbf{Z}_i | \boldsymbol{\theta}, \zeta_i) p(\zeta_i | \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta) I(a_i > 0) \\
&\propto \exp \left\{ -\frac{1}{2} (\mathbf{Z}_i - \mathbf{w} \zeta_i)^t (\mathbf{Z}_i - \mathbf{w} \zeta_i) \right\} \\
&\times \exp \left\{ -\frac{1}{2} (\zeta_i - \boldsymbol{\mu}_\zeta)^t \boldsymbol{\Sigma}_\zeta^{-1} (\zeta_i - \boldsymbol{\mu}_\zeta) \right\} I(a_i > 0) \\
&\propto \exp \left\{ -\frac{1}{2} [\zeta_i^t (\mathbf{w}^t \mathbf{w} + \boldsymbol{\Sigma}_\zeta^{-1}) \zeta_i - 2 \zeta_i^t (\mathbf{w}^t \mathbf{Z}_i + \boldsymbol{\Sigma}_\zeta^{-1} \boldsymbol{\mu}_\zeta)] \right\} I(a_i > 0) \\
&\propto \mathcal{N}(\boldsymbol{\mu}_\zeta^*, \boldsymbol{\Sigma}_\zeta^*) I(a_i > 0)
\end{aligned} \tag{2.29}$$

Onde

$$\mathbf{w} = (\boldsymbol{\theta}, -\mathbf{1}_n) \tag{2.30}$$

$$\boldsymbol{\Sigma}_\zeta^* = (\mathbf{w}^t \mathbf{w} + \boldsymbol{\Sigma}_\zeta^{-1})^{-1} \tag{2.31}$$

$$\boldsymbol{\mu}_\zeta^* = \boldsymbol{\Sigma}_\zeta^* (\mathbf{w}^t \mathbf{Z}_i + \boldsymbol{\Sigma}_\zeta^{-1} \boldsymbol{\mu}_\zeta) \tag{2.32}$$

$$\mathbf{Z}_i^t = (Z_{1i}, Z_{2i}, \dots, Z_{Ni})$$

2.7.2 Modelo probito-normal de 3 parâmetros

A partir dos resultados apresentados na seção 2.7.1 e considerando agora uma priori para o parâmetro de acerto casual, \mathbf{c} , tem-se a seguinte estrutura hierárquica para o modelo probito-normal de 3 parâmetros com modelo de regressão para o traço latente:

$$\begin{aligned}
\eta_{ji} &= a_i \theta_j - b_i, \\
\theta_j | \sigma_\theta^2 &\sim \mathcal{N}(\mathbf{x}_j^t \boldsymbol{\beta}, \sigma_\theta^2), \\
p(\beta_k) &\propto \text{const. ou } \boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \\
\zeta_i = (a_i, b_i)^t | \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta &\sim \mathcal{N} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0,5 & 0 \\ 0 & 2 \end{bmatrix} \right) \mathbb{I}(a_i > 0), \\
c_i &\propto \text{Beta}(\alpha_1, \alpha_2),
\end{aligned} \tag{2.33}$$

onde $\mathbb{I}(a_i > 0)$ indica que os valores de a_i assumem somente valores positivos, *const.* é uma constante, $\mathbf{x}_j^t = (x_{j1}, x_{j2}, \dots, x_{jp})$ e $\boldsymbol{\beta}^t = (\beta_1, \beta_2, \dots, \beta_p)$. Para determinar a distribuição

condicional completa de S_{ji} , observe que

$$\begin{aligned} P(S_{ji} = 1|Y_{ji} = 1) &\propto P(Y_{ji} = 1|S_{ji} = 1)P(S_{ji} = 1), \\ &\propto \Phi(\eta_{ji}), \end{aligned} \quad (2.34)$$

$$\begin{aligned} P(S_{ji} = 0|Y_{ji} = 1) &\propto P(Y_{ji} = 1|S_{ji} = 0)P(S_{ji} = 0) \\ &\propto c_i(1 - \Phi(\eta_{ji})), \end{aligned} \quad (2.35)$$

$$P(S_{ji} = 1|Y_{ji} = 0) = 0, \quad (2.36)$$

$$P(S_{ji} = 0|Y_{ji} = 0) = 1. \quad (2.37)$$

Note que a distribuição de S_{ji} condicional a $Y_{ji} = 1$ é binomial para uma constante normalizadora apropriada, ou seja,

$$S_{ji} \sim \text{Bernoulli}(p_{ji}) \text{ com } p_{ji} = \frac{\Phi(\eta_{ji})}{\Phi(\eta_{ji}) + c_i(1 - \Phi(\eta_{ji}))}. \quad (2.38)$$

Amostrar S_{ji} conforme distribuição discreta acima. A condicional completa para Z_{ji} deve levar em conta o passo anterior, desta forma tem-se:

$$Z_{ji}|S_{ji}, \theta_j, b_i \sim \begin{cases} \mathcal{N}(a_i\theta_j - b_i, 1)\mathbb{I}(Z_{ji} > 0), & \text{se } S_{ji} = 1, \\ \mathcal{N}(a_i\theta_j - b_i, 1)\mathbb{I}(Z_{ji} \leq 0), & \text{se } S_{ji} = 0. \end{cases} \quad (2.39)$$

A amostragem de θ e ζ seguem os passos do modelo 2PN apresentado. Para amostrar \mathbf{c} , defina $t_i = \sum_{j=1}^n I(S_{ji} = 0)$, ou seja, o total de sujeitos que não sabem a resposta correta do item e chutam a resposta. A probabilidade de resposta correta de um sujeito j ao item i , por chute, é $P(Y_{ji} = 1|S_{ji} = 0) = c_i$, sendo o total dado por $s_i = \sum_{i|S_{ji}=0}^n Y_{ji}$ tem distribuição Binomial(t_i, c_i). Com uma priori conjugada Beta(α_1, α_2) para c_i , sua distribuição posterior é

$$c_i|\mathbf{S}, \mathbf{Y} \sim \text{Beta}(s_i + \alpha_1, t_i - s_i + \alpha_2). \quad (2.40)$$

Capítulo 3

Aplicação a dados simulados

3.1 Introdução

Neste capítulo serão conduzidos estudos de simulação com o objetivo de avaliar o método de estimação proposto. Como a metodologia empregada neste trabalho é a inferência bayesiana via MCMC, estudos de simulação são necessários para avaliar três aspectos da estimação. Primeiro, a robustez do modelo em relação à escolha de prioris que podemos chamar de Análise de sensibilidade, segundo, a consistência dos estimadores, que pode ser verificada a partir de estudos de replicação dos dados e, por último, a convergência e auto-correlações das cadeias geradas pelo procedimento MCMC.

Para a análise de sensibilidade foram consideradas 4 prioris diferentes para os coeficientes de regressão β , listadas na Tabela 3.1. A priori P1 é imprópria, P2-P4 são prioris normais bivariadas vagamente informativas, de forma que para todas as prioris mencionadas, foram rodadas três cadeias paralelas com valores iniciais distintos, sendo em uma delas as estimativas da TCM para os traços latentes e parâmetros dos itens conforme Andrade et al. (2000) e nas demais, os valores foram escolhidos aleatoriamente. Um modelo de regressão bayesiano com as mesmas prioris da Tabela 3.1 foi ajustado aos escores obtidos pelo Bilog-MG (Mislevy e Bock, 1990) na métrica (0,1) e, seus resultados foram posteriormente comparados com a estimação conjunta do método bayesiano proposto.

Para Azevedo (2010), um método de estimação produz estimativas consistentes se as esperanças dos estimadores convergirem para seus verdadeiros valores e se suas respectivas variâncias tenderem para zero, de forma que vícios e variâncias tenham magnitudes desprezíveis. Desta forma, foram simuladas dez réplicas de um banco de dados composto por 24 itens com base nos modelos probito-normal de 2 e 3 parâmetros.

Tabela 3.1 *Prioris para análise de sensibilidade*

Priori	
P1	imprópria
P2	$N(\mathbf{0}, \mathbf{diag}\{10, 10\})$
P3	$N(\mathbf{0}, \mathbf{diag}\{100, 100\})$
P4	$N([-1 \ 1]^T, \mathbf{diag}\{100, 100\})$

A primeira réplica foi utilizada para a análise de sensibilidade dos modelos e o conjunto de dez réplicas para avaliar a consistência dos estimadores através das correlações entre as estimativas médias e os verdadeiros valores simulados, além da média do Erro Quadrático Médio (EQM) dos estimadores em cada réplica.

Para checar a consistência dos estimadores dos parâmetros de modelos estatísticos é frequente conduzir estudos de replicação para o cálculo empírico dessas características, o expediente neste trabalho seguirá algumas medidas descritas a seguir:

- Corr: correlação entre as médias das estimativas do parâmetro de interesse, obtidas em cada réplica e os seus verdadeiros valores.
- MEQM: a média dos Erros Quadráticos Médios das estimativas do parâmetro de interesse, obtidas em cada réplica.
- MVAER: a média do Valor Absoluto do Erro Relativo (VAER) entre a média das estimativas do parâmetro de interesse, obtidas em cada réplica e os seus verdadeiros valores.

Para ilustrar a metodologia apresentada, um pequeno estudo simulação foi conduzido. Para o modelo 2PN, foram simuladas respostas binárias de $n = 500$ indivíduos e 2 covariáveis $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2)$ com \mathbf{x}_i tendo média 0 e variância 1, ou seja, isto equivale a padronização das covariáveis, o teste simulado foi composto de 24 itens. Em seguida, com $\boldsymbol{\beta} = (0, 5; 0, 8)^T$, as habilidades foram obtidas a partir de $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \sqrt{1 - \boldsymbol{\beta}^T \boldsymbol{\beta}} \boldsymbol{\epsilon}$, sendo $\boldsymbol{\epsilon} \sim N(0, \sigma_\theta^2 I)$ e os parâmetros dos itens foram escolhidos aleatoriamente de forma que $\mathbf{a} \in (0, 6; 1, 4)$ e $\mathbf{b}^* \in (-2, 4; 2, 4)$, segundo Azevedo et al. (2010) estes valores garantem boas propriedades para os itens. Para o modelo 3PN, o tamanho da amostra foi de

$n = 1.000$ sujeitos com o mesmo número de itens e covariáveis simulados da mesma forma que no modelo 2PN e, observando a independência a priori dos parâmetros dos itens, o parâmetro de acerto casual c foi escolhido aleatoriamente no intervalo $(0; 0,35)$.

As respostas aos itens do teste foram simuladas considerando o algoritmo da transformação inversa, disponível no software R (R Development Core Team, 2011).

3.1.1 Modelo probito-normal de 2 parâmetros

Para diagnosticar a convergência das cadeias, foram calculadas as estatísticas de Gelman e Rubin (1992), e após a determinação do período de aquecimento (*burn-in*), as estatísticas de Geweke (1992) e gráficos das médias ergódicas e variâncias foram calculadas a cada 1.000 iterações, para checar a adequação do período de aquecimento. A convergência das cadeias foi atingida a partir de 25.000 iterações (veja Figura 3.1, pois o intervalo de 95% confiança para mediana converge para 1 a partir deste período considerado como aquecimento (*burn-in*), além disso, as estatísticas de Geweke foram não significativas e os gráficos de médias e variâncias mostraram-se estáveis a partir deste período para os parâmetros dos itens e para os coeficientes de regressão.

Na Figura 3.2, os valores simulados para as três cadeias, retidos a cada 30 observações, ao longo de 30.000 iterações são apresentados, comportamentos similares estão presentes em todos os outros parâmetros do modelo (exceto os traços latentes), note que visualmente as cadeias misturam-se bem, entretanto conforme ressaltado, uma das desvantagens do esquema de dados aumentados é produzir amostras mais correlacionadas sendo necessário determinar um salto a partir do qual as amostras sejam aproximadamente independentes. Segundo Bazán (2005), defasagens com auto-correlações inferiores à 0,4 são suficientes para amostragem, desta forma, amostras sem e com espaçamento de trinta (30) unidades após o aquecimento são mostradas na Figura 3.3, note que as auto-correlações elevadas nas amostras sem espaçamento tanto para os parâmetros de regressão ($\beta[1]$ e $\beta[2]$) como para os parâmetros dos itens 4 e 20, tornando-se praticamente não auto-correlacionadas com espaçamento (amostra válida), pois todas as auto-correlações são inferiores a 0,4 e a maioria são inferiores a 0,2 para todos os parâmetros de regressão e dos itens.

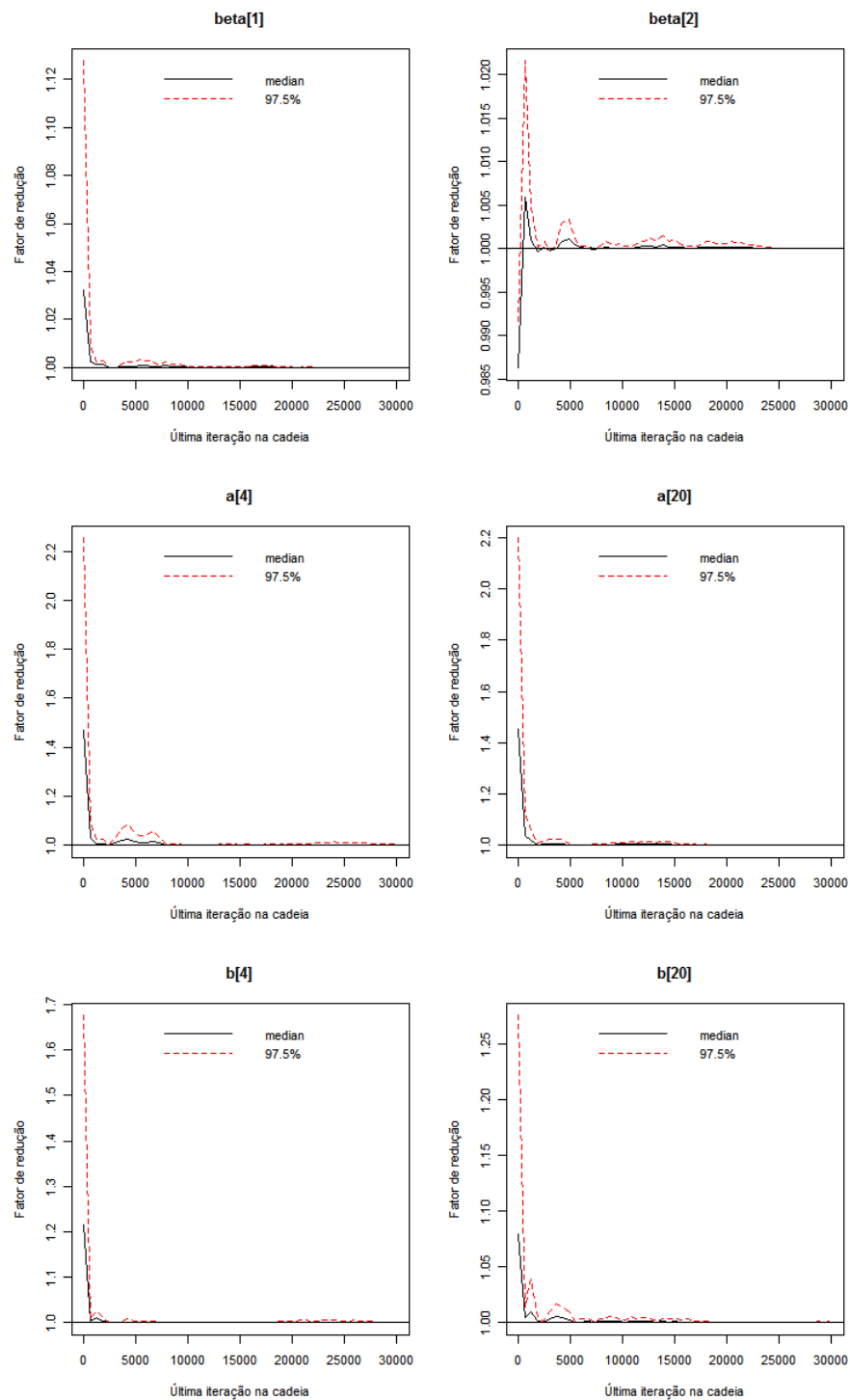


Figura 3.1 Gráfico de diagnóstico de convergência de Gelman-Rubin, réplica 1 - Priori P1

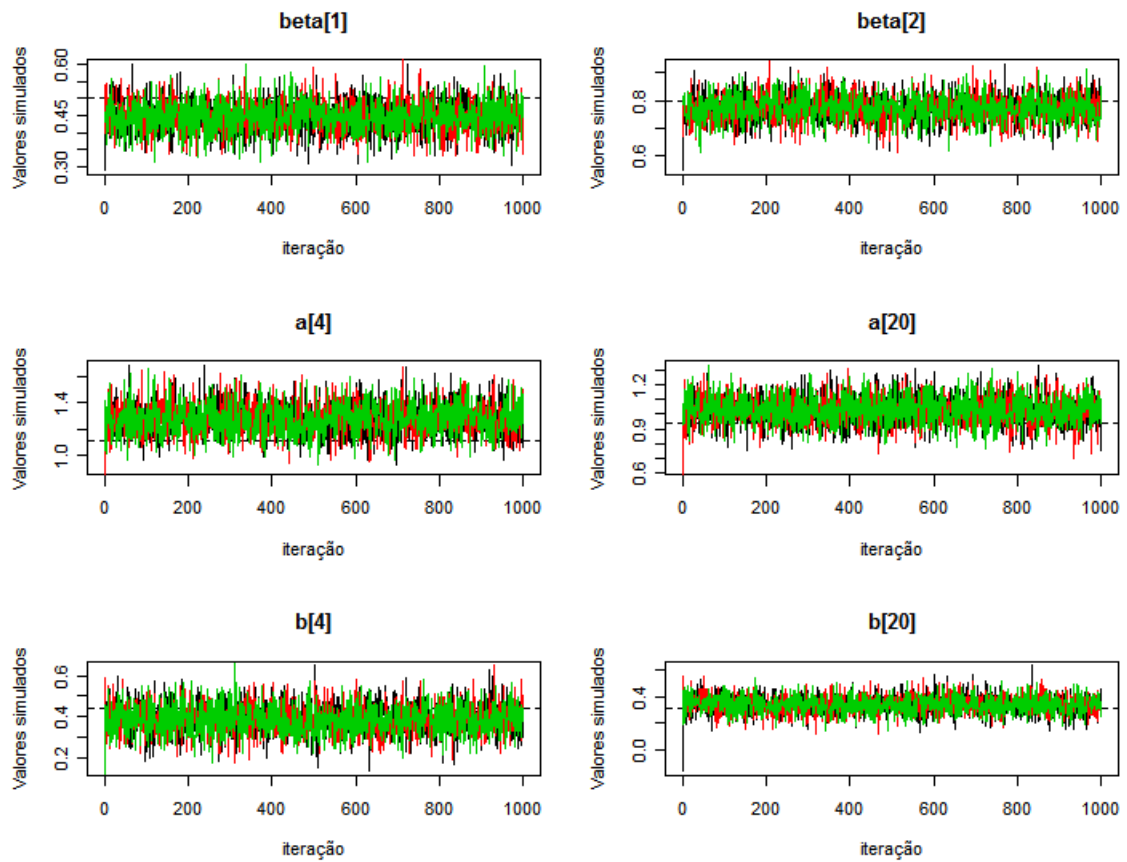


Figura 3.2 *Simulações para os parâmetros de regressão e dos itens 4 e 20, retidos a cada 30 iterações, réplica 1 - Priori P1*

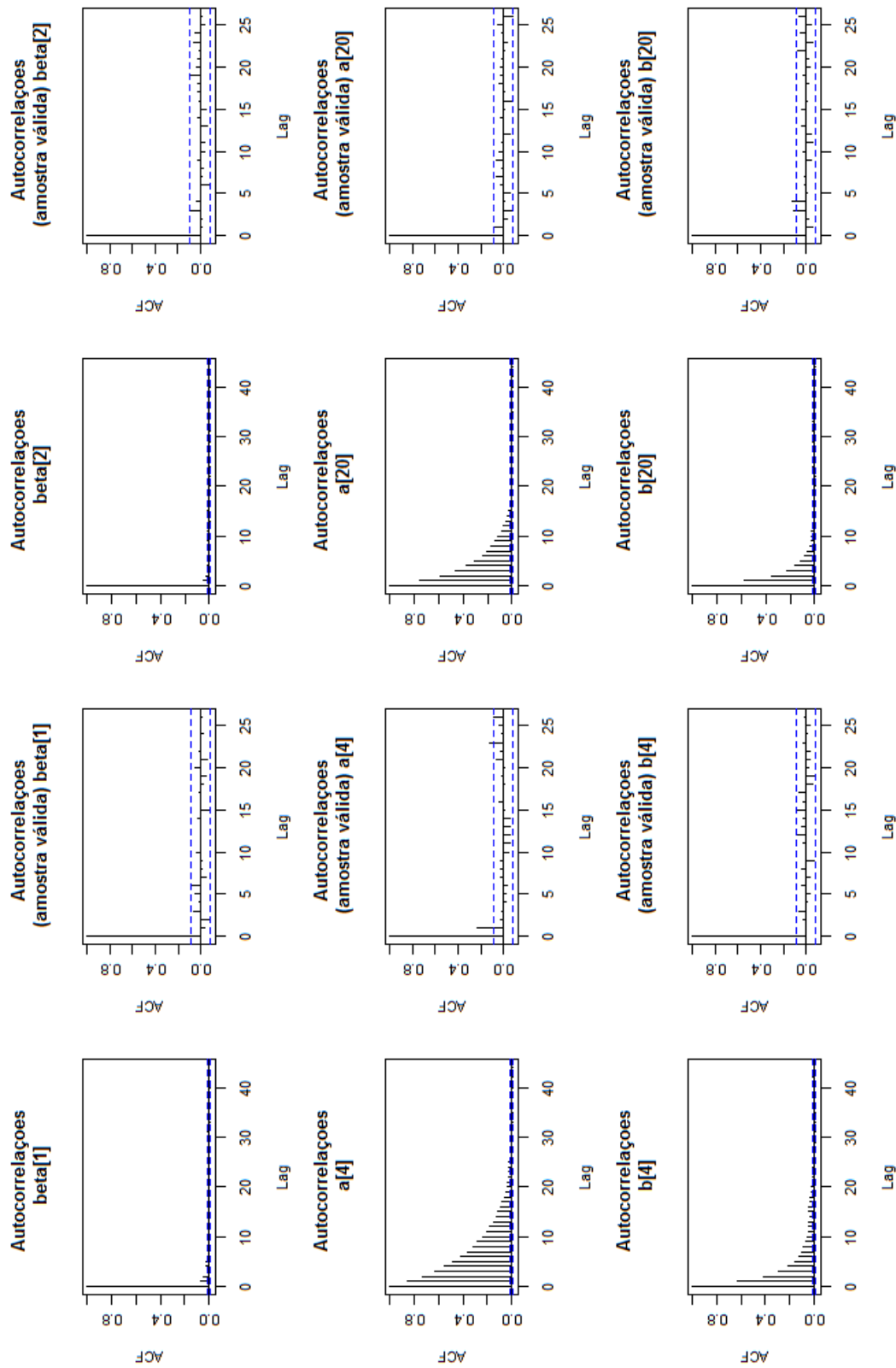


Figura 3.3 Auto-correlações dos valores simulados para os parâmetros de regressão e dos itens 4 e 20, réplica 1 - Priori P1

A análise do DIC marginal e outros critérios na Tabela 3.2 mostra que não há diferenças significativas entre as prioris consideradas para β . Entretanto a soma dos quadrados da resíduos latentes (SQRL) mostra que as prioris vagamente informativas (P2-P4) foram ligeiramente superiores na modelagem de θ . A priori P3 apresentou menor SQRL, enquanto a priori P2 teve os melhores resultados nos critérios DIC, EAIC e EBIC.

Tabela 3.2 *Crítérios de ajuste para o modelo probito-normal de 2 parâmetros: réplica 1*

Priori	ρ_D	DIC	EAIC	EBIC	SQRL
P1	459.87	12154.47	12790.59	15100.20	64.10
P2	457.56	12151.49	12789.94	15099.54	63.90
P3	459.40	12154.27	12790.87	15100.47	63.75
P4	459.56	12155.25	12791.69	15101.30	63.91

A Tabela 3.3 mostra a *corr* e o *MEQM* para as prioris propostas pela estimação conjunta e para estimação via Bilog-MG (ou seja, sem o modelo de regressão para o traço latente). As correlações entre as estimativas obtidas e o verdadeiros valores do traço latente e parâmetros dos itens também foram calculadas, sendo que para todas as prioris as correlações foram maiores do que as correlações a partir dos escores estimados a partir do Bilog-MG, indicando que, ao menos empiricamente, a especificação de um modelo de regressão melhora as estimativas dos parâmetros de interesse.

Os erros de estimação dos parâmetros de regressão foram mensurados a partir do MVAER. Observa-se que em todas as prioris especificadas, os coeficientes β_i foram menores quando comparados com a regressão baseada nos escores obtidos pelo ajuste do Bilog-MG.

A Figura 3.4 apresenta o ajuste do modelo aos verdadeiros parâmetros, para a priori imprópria (A), réplica 1. Note que o modelo é bem comportado exceto para alguns parâmetros de discriminação, e que os resíduos da parte estrutural do modelo de regressão estão normalmente distribuídos sem indícios de outliers. Vale ressaltar que os resultados aqui obtidos restringem-se ao conjunto de prioris utilizadas e a especificação de prioris para os parâmetros de regressão não influencia significativamente a estimação das habilidades, tão pouco a estimação dos itens, outrora, a estimação de θ não apresenta erro

Tabela 3.3 *Corr (%) e MEQM para os traços latentes, discriminação e dificuldade dos itens*

Método	Prioris	Correlação			EQM		
		θ	a	b	θ	a	b
MCMC							
	P1	95,65	95,99	99,15	0,01	<0,01	1,83
	P2	95,58	96,00	99,07	0,01	<0,01	1,80
	P3	95,57	95,95	99,04	0,01	<0,01	1,78
	P4	95,48	95,87	99,04	0,01	<0,01	1,80
Bilog-MG							
	P1	94.30	95.45	96.10	0.01	0.02	2.32
	P2	94,31	95.42	95.92	0.01	0.02	2.20
	P3	95.01	95.33	99.81	0.01	0.01	2.18
	P4	94.93	95.15	95.80	0.01	0.01	2.20

Tabela 3.4 *MVAER para os coeficientes de regressão*

Coefficiente	P1	P2	P3	P4
MCMC				
beta[1]	11.61	11.32	11.12	11.07
beta[2]	4.01	3.45	3.02	3.33
Bilog-MG				
beta[1]	12.01	11.94	11.90	11.88
beta[2]	5.81	6.03	5.89	5.85

quadrático médio significativamente inferior na estimação conjunta com os parâmetros de regressão.

Na Tabela 3.5, os verdadeiros valores, suas estimativas, intervalos de confiança e intervalos HPD são apresentados. Os intervalos HPD foram calculados usando a técnica descrita em Chen e Shao (1999), implementada no pacote BOA (Smith, 2007). Observe

que todos os intervalos de confiança e HPD's contém os verdadeiros valores com exceção dos itens 1 e 12 (parâmetro **b**), 14 e 17 (parâmetro **a**), bem como os erros-padrão foram relativamente baixos exceto para alguns parâmetros de dificuldade próximos de zero.

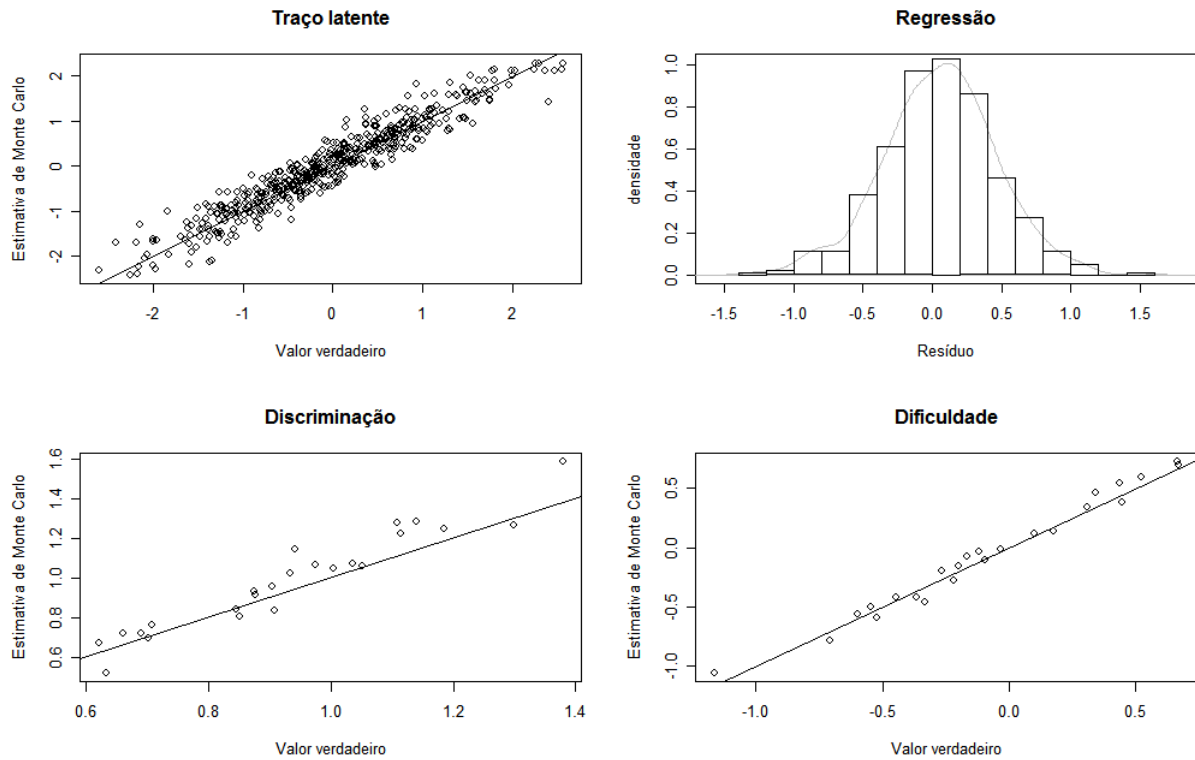


Figura 3.4 *Qualidade do ajuste do modelo de regressão e parâmetros dos itens-2PN, réplica 1. (Priori A)*

Tabela 3.5 *Parâmetros dos itens, Estimativas, erros-padrão, IC(95%) e HPD(95%) - Priori A*

Item	Verd.		Discriminação				Dificuldade			
	a	b	\hat{a}	EP	IC	HPD	\hat{b}	EP	IC	HPD
1	1.18	0.34	1.24	0.12	[1.06;1.46]	[1.04;1.50]	0.46	0.07	[0.34;0.58]	[0.32;0.60]
2	0.85	-0.60	0.84	0.09	[0.68;0.99]	[0.69;1.01]	-0.56	0.07	[-0.67;-0.45]	[-0.68;-0.42]
3	1.04	-0.55	1.09	0.11	[0.91;1.27]	[0.89;1.28]	-0.50	0.07	[-0.62;-0.39]	[-0.63;-0.37]
4	1.11	0.44	1.28	0.13	[1.09;1.49]	[1.04;1.50]	0.39	0.07	[0.26;0.50]	[0.27;0.54]
5	0.66	-0.52	0.72	0.08	[0.58;0.86]	[0.55;0.88]	-0.59	0.07	[-0.70;-0.48]	[-0.72;-0.46]
6	1.00	-0.22	1.06	0.10	[0.90;1.23]	[0.87;1.26]	-0.27	0.07	[-0.39;-0.16]	[-0.40;-0.14]
7	0.88	-0.20	0.91	0.09	[0.77;1.07]	[0.74;1.11]	-0.15	0.06	[-0.24;-0.04]	[-0.27;-0.03]
8	0.85	0.44	0.80	0.09	[0.66;0.94]	[0.64;0.97]	0.54	0.06	[0.43;0.65]	[0.41;0.67]
9	0.87	0.18	0.93	0.09	[0.78;1.06]	[0.75;1.10]	0.14	0.06	[0.04;0.24]	[0.01;0.25]
10	1.30	-0.45	1.28	0.12	[1.09;1.49]	[1.02;1.47]	-0.41	0.07	[-0.53;-0.30]	[-0.53;-0.26]
11	1.14	0.67	1.28	0.12	[1.11;1.47]	[1.03;1.50]	0.70	0.08	[0.57;0.83]	[0.51;0.82]
12	0.62	-0.33	0.66	0.08	[0.54;0.79]	[0.53;0.82]	-0.46	0.06	[-0.56;-0.35]	[-0.58;-0.34]
13	0.69	-0.12	0.73	0.08	[0.60;0.86]	[0.53;0.85]	-0.03	0.06	[-0.13;0.07]	[-0.14;0.08]
14	0.63	-0.27	0.52	0.07	[0.41;0.63]	[0.39;0.67]	-0.19	0.06	[-0.29;-0.09]	[-0.29;-0.09]
15	1.05	0.10	1.06	0.10	[0.90;1.24]	[0.84;1.25]	0.12	0.07	[0.01;0.23]	[-0.01;0.23]
16	1.38	-0.71	1.57	0.16	[1.30;1.85]	[1.32;1.87]	-0.78	0.10	[-0.94;-0.64]	[-0.96;-0.61]
17	0.94	0.66	1.14	0.11	[0.96;1.34]	[0.92;1.31]	0.73	0.07	[0.61;0.85]	[0.59;0.88]
18	0.91	-0.36	0.83	0.09	[0.68;0.97]	[0.63;0.98]	-0.41	0.07	[-0.52;-0.29]	[-0.53;-0.29]
19	0.90	-1.16	0.96	0.10	[0.79;1.12]	[0.72;1.15]	-1.06	0.08	[-1.19;-0.93]	[-1.24;-0.90]
20	0.93	0.31	1.03	0.10	[0.88;1.20]	[0.84;1.25]	0.35	0.07	[0.25;0.46]	[0.22;0.46]
21	1.11	-0.09	1.22	0.12	[1.04;1.43]	[1.00;1.47]	-0.11	0.07	[-0.22;0.00]	[-0.25;0.02]
22	0.98	0.52	1.06	0.10	[0.88;1.23]	[0.87;1.25]	0.59	0.07	[0.48;0.71]	[0.45;0.73]
23	0.70	-0.03	0.68	0.07	[0.56;0.81]	[0.53;0.84]	-0.01	0.06	[-0.11;0.09]	[-0.15;0.10]
24	0.71	-0.17	0.76	0.08	[0.64;0.89]	[0.61;0.91]	-0.08	0.06	[-0.18;0.01]	[-0.18;0.06]

3.1.2 Modelo Probito-Normal de 3 parâmetros

O estudo de simulação e as medidas de diagnóstico de convergência para este modelo seguem basicamente os mesmos passos da seção 3.1.1, ou seja, três cadeias paralelas foram rodadas, diagnosticando-se a convergência para os parâmetros dos itens e para os coeficientes de regressão a partir de 20.000 iterações com espaçamento de 40 unidades, gerando uma amostra de tamanho $n = 250$ indivíduos. A partir da Figura 3.5, os verdadeiros traços latentes apresentam-se correlacionados com as estimativas de Monte Carlo.

Comparando com as estimativas dos parâmetros de discriminação e dificuldade do modelo 2PN, o modelo 3PN teve desempenho inferior no parâmetro de discriminação, mas visualmente ajustou bem o parâmetro de dificuldade. Nota-se também que itens com acerto casual próximo de zero não são bem estimados, porém itens com acerto casual acima de 0,1 são mais bem estimados e, de acordo com Figura 3.5, todos itens com acerto casual abaixo de 0,1 são pobremente estimados, isto deve-se provavelmente ao conflito com a priori $c_i \sim \text{Beta}(6, 16)$.

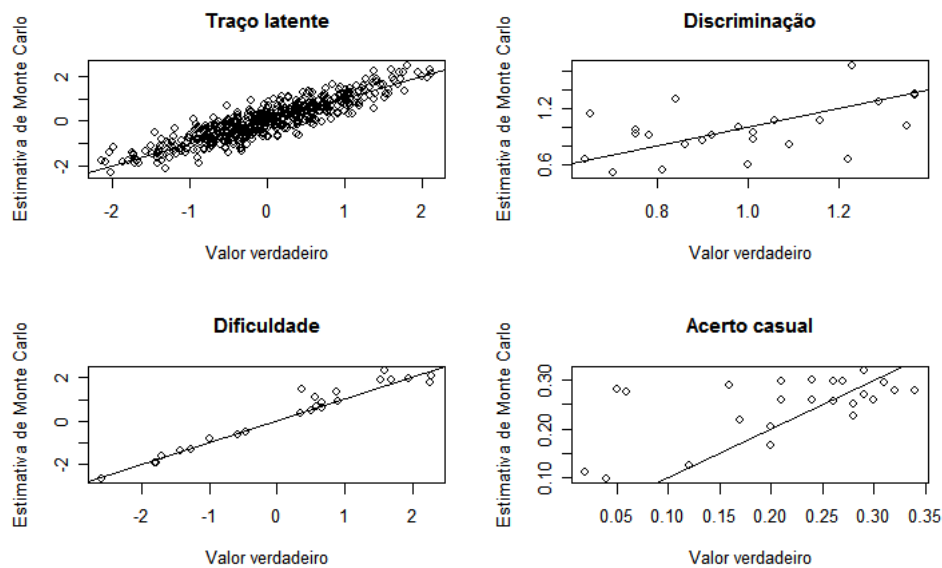


Figura 3.5 *Qualidade do ajuste do modelo de regressão e parâmetros dos itens para o modelo probito-normal de 3 parâmetros: Priori 1, réplica 1*

Os principais resultados do ajuste são mostrados na Tabela 3.7, observe que o parâmetro c dos itens 1, 12, 14 e 19 foi estimado com erro elevado indentificando os itens discutidos na

Figura 3.5 com baixo acerto casual. Os intervalos de confiança e HPD para estes itens não incluem os verdadeiros valores deste parâmetro e também para o parâmetro de dificuldade (1, 14 e 19) e somente no item 18, os intervalos bayesianos não continham o valor simulado para discriminação.

Na Tabela 3.6, as correlações com os verdadeiros valores do parâmetros e o EQM em sua estimação são apresentados. Nota-se que, de forma geral, as estimativas entre o Bilog-MG e método MCMC são próximas entre si sendo que o desempenho menos satisfatório foi para o parâmetro de discriminação, com baixa correlação seguido do parâmetro de acerto casual confirmando a inspeção visual da Figura 3.5.

Tabela 3.6 *Corr (%) e MEQM dos parâmetros dos itens estimados: Priori P1*

Parâmetro	Bilog-MG		MCMC	
	Cor	MEQM	Cor	MEQM
a	39.878	0.063	50.851	0.065
b	97.409	0.213	97.657	0.113
c	63.926	0.006	59.414	0.007
θ	85.415	0.210	89.485	0.158

O MVAER também foi calculado para os coeficientes de regressão estimados a partir do modelo probito-normal de 3 parâmetros, sendo posteriormente comparado com os resultados baseados no modelo de regressão linear aplicado no escores obtidos pelo Bilog-MG. Como no modelo 2PN, o modelo bayesiano via dados aumentados estimou melhor os coeficientes de regressão com um MVAER para $\beta[1]$ e $\beta[2]$ de 0,08% e 4,01% respectivamente, enquanto que um MVAER de 10,01% e 6,72% foi obtido com os escores do Bilog-MG.

Tabela 3.7 Valores simulados, estimativas de Monte Carlo, erros-padrão, intervalos de confiança e HPD: modelo β PN, Priori $P1$ -réplica 1

Item	Verd.			Discriminação			Dificuldade			Acerto casual					
	a	b	c	\hat{a}	EP	IC	HPD	\hat{b}	EP	IC	HPD	\hat{c}	EP	IC	HPD
1	0,65	0,38	0,06	1,15	0,34	[0,64;1,81]	[0,54;1,84]	1,44	0,47	[0,78;2,37]	[0,55;2,37]	0,28	0,04	[0,20;0,33]	[0,20;0,35]
2	0,86	-1,81	0,27	0,81	0,21	[0,49;1,16]	[0,45;1,23]	-1,93	0,22	[-2,31;-1,60]	[-2,34;-1,54]	0,30	0,10	[0,16;0,47]	[0,13;0,51]
3	1,35	0,66	0,34	1,01	0,26	[0,65;1,48]	[0,54;1,52]	0,60	0,29	[0,15;1,12]	[0,10;1,14]	0,28	0,06	[0,17;0,38]	[0,17;0,40]
4	0,64	-1,26	0,21	0,66	0,16	[0,43;0,94]	[0,41;1,01]	-1,32	0,13	[-1,53;-1,11]	[-1,57;-1,10]	0,30	0,10	[0,15;0,48]	[0,13;0,51]
5	1,29	-1,42	0,21	1,27	0,23	[0,89;1,65]	[0,87;1,74]	-1,36	0,18	[-1,66;-1,06]	[-1,70;-1,02]	0,26	0,08	[0,13;0,38]	[0,10;0,40]
6	1,16	-2,59	0,16	1,07	0,31	[0,60;1,55]	[0,54;1,67]	-2,65	0,37	[-3,31;-2,08]	[-3,38;-2,01]	0,29	0,10	[0,15;0,47]	[0,11;0,47]
7	1,37	0,34	0,29	1,37	0,33	[0,91;2,00]	[0,77;2,05]	0,34	0,25	[-0,06;0,75]	[-0,16;0,77]	0,27	0,07	[0,16;0,37]	[0,12;0,38]
8	0,81	-1,71	0,24	0,54	0,16	[0,30;0,82]	[0,23;0,83]	-1,62	0,16	[-1,88;-1,37]	[-1,92;-1,32]	0,30	0,10	[0,15;0,46]	[0,14;0,51]
9	0,92	1,94	0,20	0,91	0,27	[0,53;1,43]	[0,45;1,50]	1,94	0,46	[1,28;2,72]	[1,19;2,81]	0,17	0,03	[0,12;0,22]	[0,11;0,22]
10	1,37	1,69	0,24	1,35	0,42	[0,78;2,11]	[0,66;2,17]	1,87	0,55	[1,09;2,89]	[0,93;2,93]	0,26	0,04	[0,20;0,31]	[0,19;0,32]
11	1,09	-0,46	0,26	0,82	0,17	[0,59;1,12]	[0,54;1,14]	-0,47	0,16	[-0,72;-0,18]	[-0,75;-0,16]	0,30	0,08	[0,16;0,44]	[0,14;0,45]
12	0,75	-1,79	0,05	0,92	0,24	[0,60;1,36]	[0,51;1,42]	-1,94	0,24	[-2,38;-1,61]	[-2,39;-1,54]	0,28	0,09	[0,14;0,45]	[0,10;0,45]
13	0,70	-1,00	0,29	0,51	0,14	[0,31;0,78]	[0,29;0,85]	-0,80	0,16	[-1,04;-0,56]	[-1,11;-0,55]	0,32	0,11	[0,16;0,51]	[0,15;0,55]
14	0,84	1,60	0,04	1,31	0,34	[0,82;1,95]	[0,77;2,00]	2,32	0,50	[1,64;3,26]	[1,46;3,30]	0,10	0,02	[0,07;0,13]	[0,06;0,13]
15	1,01	0,60	0,26	0,94	0,18	[0,69;1,24]	[0,63;1,26]	0,68	0,21	[0,36;1,06]	[0,31;1,10]	0,26	0,05	[0,17;0,32]	[0,16;0,34]
16	1,06	-0,57	0,30	1,07	0,18	[0,81;1,36]	[0,76;1,37]	-0,62	0,15	[-0,83;-0,39]	[-0,94;-0,38]	0,26	0,07	[0,14;0,38]	[0,12;0,40]
17	1,23	0,66	0,31	1,67	0,36	[1,11;2,33]	[1,05;2,44]	0,86	0,29	[0,44;1,37]	[0,35;1,40]	0,30	0,04	[0,23;0,36]	[0,21;0,37]
18	1,22	2,29	0,20	0,66	0,29	[0,23;1,17]	[0,18;1,26]	2,06	0,51	[1,26;2,93]	[1,13;3,03]	0,20	0,03	[0,14;0,26]	[0,14;0,27]
19	0,98	0,56	0,02	1,00	0,19	[0,75;1,36]	[0,72;1,42]	1,08	0,22	[0,78;1,46]	[0,69;1,48]	0,11	0,03	[0,06;0,16]	[0,06;0,17]
20	0,75	0,89	0,17	0,97	0,25	[0,60;1,42]	[0,52;1,45]	1,35	0,33	[0,87;1,93]	[0,81;2,09]	0,22	0,04	[0,16;0,27]	[0,15;0,28]
21	0,90	0,50	0,28	0,86	0,19	[0,61;1,20]	[0,56;1,26]	0,50	0,23	[0,18;0,93]	[0,10;0,96]	0,23	0,06	[0,13;0,32]	[0,11;0,33]
22	1,00	2,26	0,32	0,60	0,31	[0,21;1,21]	[0,11;1,22]	1,77	0,59	[0,99;3,00]	[0,86;3,08]	0,28	0,04	[0,20;0,34]	[0,19;0,34]
23	0,78	1,54	0,12	0,91	0,27	[0,58;1,44]	[0,55;1,54]	1,87	0,40	[1,37;2,59]	[1,20;2,61]	0,13	0,02	[0,09;0,17]	[0,09;0,17]
24	1,01	0,90	0,28	0,87	0,25	[0,52;1,25]	[0,37;1,25]	0,93	0,34	[0,45;1,61]	[0,32;1,67]	0,25	0,06	[0,15;0,34]	[0,13;0,35]

A Tabela 3.8 apresenta resultados a um estudo de simulação para verificar a redução do erro quadrático médio à medida que variáveis adequadamente correlacionadas fossem incluídas no modelo de regressão e o tempo médio necessário em segundos, por cadeia, para gerar 1.000 amostras. Os procedimentos para simular os parâmetros e as 4 covariáveis foram os mesmos adotados na seção 2.1 em um computador Intel(R) Core(TM) 2 Duo (2,20 GHz e 4,00 GB de memória RAM).

Cada variável foi incluída de acordo com a magnitude de correlação (crescente) com θ sendo observado a redução sistemática do EQM à medida da adição de covariáveis no modelo. O acréscimo no tempo de simulação em 1.000 iterações do amostrador não apresentou variações significativas fixado o número de sujeitos, mas a cada acréscimo de 500 sujeitos, fixado o número de covariáveis, trouxe um incremento de pelo menos 28% no tempo de processamento anterior.

Tabela 3.8 *Erro quadrático médio e tempo médio, em segundos (entre parênteses), por cadeia, a cada 1.000 iterações para um teste de 36 itens*

n	número de covariáveis			
	1	2	3	4
500	0,155 (12,77)	0,156 (12,80)	0,146 (13,03)	0,126 (13,17)
1000	0,162 (24,83)	0,155 (26,70)	0,141 (28,23)	0,122 (29,87)
1500	0,150 (33,70)	0,135 (36,50)	0,120 (39,43)	0,116 (41,60)
2000	0,149 (50,33)	0,129 (52,67)	0,120 (53,60)	0,101 (53,40)

Capítulo 4

Aplicações a dados reais

4.1 Introdução

O conjunto de dados reais que será utilizado neste trabalho é do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP. O SARESP é uma avaliação externa em larga escala dos alunos do Ensino Fundamental e Médio da Rede Estadual de São Paulo e de escolas municipais e particulares que aderiram ao sistema de avaliação, sendo aplicada a cada ano desde 1996 pela SEE (Secretaria da Educação do Estado) de São Paulo, com a finalidade de produzir um diagnóstico da situação da escolaridade básica na rede pública de ensino paulista, visando orientar os gestores do ensino no monitoramento das políticas voltadas para a melhoria da qualidade educacional.

Anualmente, o SARESP avalia os alunos do 3º, 5º, 7º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio, por meio de provas cognitivas nas áreas de Língua Portuguesa com Redação e Matemática, com alternância entre as disciplinas das áreas de Ciências Humanas (Geografia e História) e Ciências da Natureza (Biologia, Física e Química) aos alunos do 7º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio.

São aplicados também questionários socioeconômicos aos pais e alunos participantes de todas as redes de ensino. As escolas estaduais respondem ainda a questionários de contexto dirigidos aos diretores, coordenadores e professores das disciplinas avaliadas. O questionário socioeconômico do SARESP 2010 foi composto de quarenta e uma (41) questões, das quais a 1ª foi reservada a identificação do responsável, as questões de 2 a 31 deveriam ser respondidas pelos pais e o restante pelos alunos. Como as categorias de resposta das questões (variáveis) são ordinais, ou seja, podem ser ordenadas, e com intuito de resumir as variáveis com características comuns, criou-se alguns índices:

- I1: índice sócioeconômico familiar dos pais e alunos;

- I2: índice de realização das lições de casa;
- I3: índice de fatores favoráveis à aprendizagem;
- I4: índice de qualidade de ensino em Língua Portuguesa;
- I5: índice de qualidade de ensino em Matemática;
- I6: índice de qualidade de ensino em ciências;
- I7: índice de qualidade geral dos professores;
- I8: índice de clima escolar conforme percebido pelos pais;
- I9: índice de confiança no trabalho da escola conforme percebido pelos pais;
- I10: índice de informação fornecida pela escola conforme percebida pelos pais.

Estes índices foram calculados a partir da Teoria da Resposta ao Item, especificamente com o Modelo Samejima (1969) para respostas graduais. Neste modelo, indivíduos com baixo escore na característica latente tem probabilidade maior de responder a categoria menor do item, e indivíduos com altos escores tem probabilidade menor de responder a categoria menor do item. Para fins de análise, apenas os alunos da 4ª série/5º ano do ensino fundamental que responderam ao 1º caderno de questões foram considerados. Do total de 8.053 alunos, 1.500 alunos foram selecionados aleatoriamente com apoio do software SPSS.

Dos dez índices citados somente os índices I1, I3, I8, I9 e I10 foram calculados para os alunos do 5º ano do Ensino Fundamental e para esta aplicação somente os índices I1 e I3 serão considerados.

4.2 Modelo e especificação de prioris

O modelo proposto para estes dados será o probito-normal de 3 parâmetros com um modelo de regressão para a proficiência em língua português dos alunos da 4ª série/5º ano do Ensino Fundamental. O modelo é:

$$P(Y_{ji} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i)\Phi(a_i\theta_j - b_i),$$

$$\theta_j \sim N(\beta_1 \times (Sexo)_j + \beta_2 \times (Etnia)_j + \beta_3 \times (I1^*)_j + \beta_4 \times (I3^*)_j, 1).$$

Para a resposta de um aluno j ao item i , os termos do modelo são descritos a seguir:

- Y_{ji} assume o valor 1 se o aluno j respondeu corretamente ao item i , e 0 caso contrário;
- θ_j é a proficiência do aluno em língua portuguesa;
- $Sexo_j$ assume o valor 1 se o aluno for mulher e 0 se for homem;
- $Etnia_j$ assume o valor 1 se o aluno se auto-declararou branco e 0 caso contrário;
- $I1_j^*$ índice socioeconômico familiar padronizado do aluno;
- $I3_j^*$ índice de fatores favoráveis à aprendizagem padronizado do aluno;
- $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ são os coeficientes de regressão.

Para ajustar o modelo foram especificados três conjuntos de priors para os parâmetros dos itens. As priors são:

- C1: $a_i \sim N(1; 0, 5)I(a_i \geq 0)$, $b_i \sim N(0, 2)$, $c_i \sim Beta(6, 16)$, $i=1, 2, \dots, 24$;
- C2: $a_i \sim N(1; 0, 5)I(a_i \geq 0)$, $b_i \sim N(0, 2)$, $c_i \sim Beta(1, 3)$, $i=1, 2, \dots, 24$;
- C3: $a_i \sim N(1; 0, 5)I(a_i \geq 0)$, $b_i \sim N(0, 2)$, $c_i \sim Beta(10, 30)$, $i=1, 2, \dots, 24$;

Como nos estudos de simulação, as priors para os parâmetros de discriminação (**a**) e dificuldade (**b**) dos itens foram aquelas consideradas por Sahu (2002) e Patz e Junker (1999), que também foram utilizadas em uma aplicação a dados reais por Bázan (2005). A escolha das priors para o acerto casual leva em consideração o fato de que as questões de múltipla escolha que compõem o caderno da prova de língua portuguesa tem 4 alternativas e apenas uma correta, o que resultaria em uma probabilidade de 0,25 de acerto com chute aleatório, desta forma, as priors C2 e C3 tiveram seus parâmetros ajustados para que a sua distribuição à priori tenha essa probabilidade como valor esperado, diferindo apenas na variância, que é maior para priori C2, ou seja, menos informativa, veja a Figura 4.1.

No conjunto de priors C1, o acerto casual tem uma média de aproximadamente 0,27, o que é realístico tendo em vista que para um sujeito com uma baixa proficiência, leva-se em conta não apenas a habilidade, mas também algumas estratégias como a eliminação de opções por absurdo.

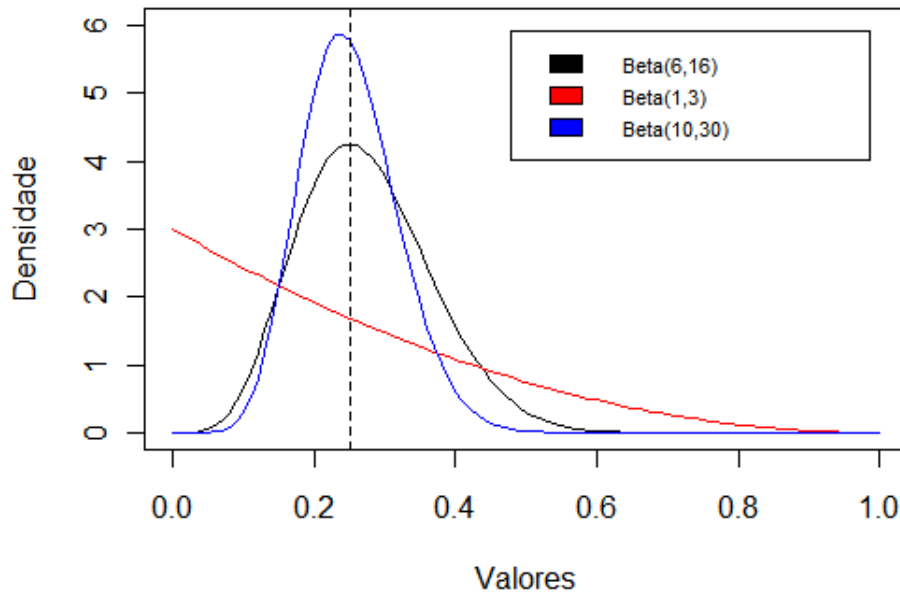


Figura 4.1 *Prioris para o parâmetro de acerto casual*

Os mesmos procedimentos realizados nos estudos de simulação foram considerados nesta aplicação, ou seja, número de cadeias, cálculo dos valores iniciais e diagnóstico de convergência. Uma priori imprópria para os parâmetros de regressão foi utilizada, e a partir de 30.000 simulações com descarte das primeiras 17.500 iterações, com salto de 50 para reduzir as autocorrelações nas cadeias, o que gerou uma amostra de tamanho 250.

4.3 Análise das estimativas

Algumas medidas descritivas simples para o escore bruto obtido pelos alunos na amostra estão presentes na Tabela 4.1. Observa-se que a média e a mediana dos escores estão próximas, ambas em torno de 13 acertos, isto revela uma distribuição aproximadamente simétrica, o que também pode ser concluído a partir do coeficiente de assimetria próximo de zero. A partir dos quartis, nota-se que três quartos dos alunos obtiveram escores acima

e abaixo de 9 e 17 respectivamente, e que o desvio-padrão é relativamente moderado, com um coeficiente de variação de aproximadamente 38,51%.

Tabela 4.1 *Estatísticas descritivas*

Média	13,19	Mediana	13,00
1º Quartil	9,00	3º Quartil	17,00
Variância	25,17	Desvio-padrão	5,08
Assimetria	-0,05	Curtose	-0,81

Pela Tabela 4.2, os resultados são similares entre as prioris C1 e C3 com poucas diferenças em todos os critérios. A priori C2 teve melhor desempenho pelos critérios DIC, EAIC e EBIC, mas de acordo com o critério SQRes, com as prioris C1 e C3, houve melhor explicação da proficiência em Português pelo modelo de regressão considerado. Ainda assim, optou-se pela priori C2 para exibição e análise dos resultados referentes ao parâmetros dos itens, uma vez que as diferenças encontradas em SQres são desprezíveis.

Tabela 4.2 *Comparação dos modelos para as prioris C1, C2 e C3*

	ρ_D	DIC	EAIC	EBIC	SQRes
C1	1324.792	38536.88	40308.09	48532.96	1090.131
C2	1332.989	38518.64	40281.65	48506.52	1099.979
C3	1318.409	38534.66	40312.26	48537.12	1089.571

Na Figura 4.2, os resíduos latentes e os resíduos do modelo de regressão estimado para as proficiências são apresentados. Percebe-se que os resíduos das proficiências não mostram grandes desvios de normalidade e os resíduos do modelo de regressão tem um comportamento ligeiramente assimétrico à esquerda. Para identificar alguma heterogeneidade dos resíduos com relação as covariáveis, bem como uma relação não linear da Proficiência com os indicadores I1 e I3 considerados na modelagem, diagramas de caixa (*Box-plots*) e de dispersão foram construídos, veja as Figuras 4.3 e 4.4. Não há indícios de heterocedastici-

dade grave tanto para as variáveis binárias *sexo* e *Raça/Etnia*, como para os indicadores I1 e I3.

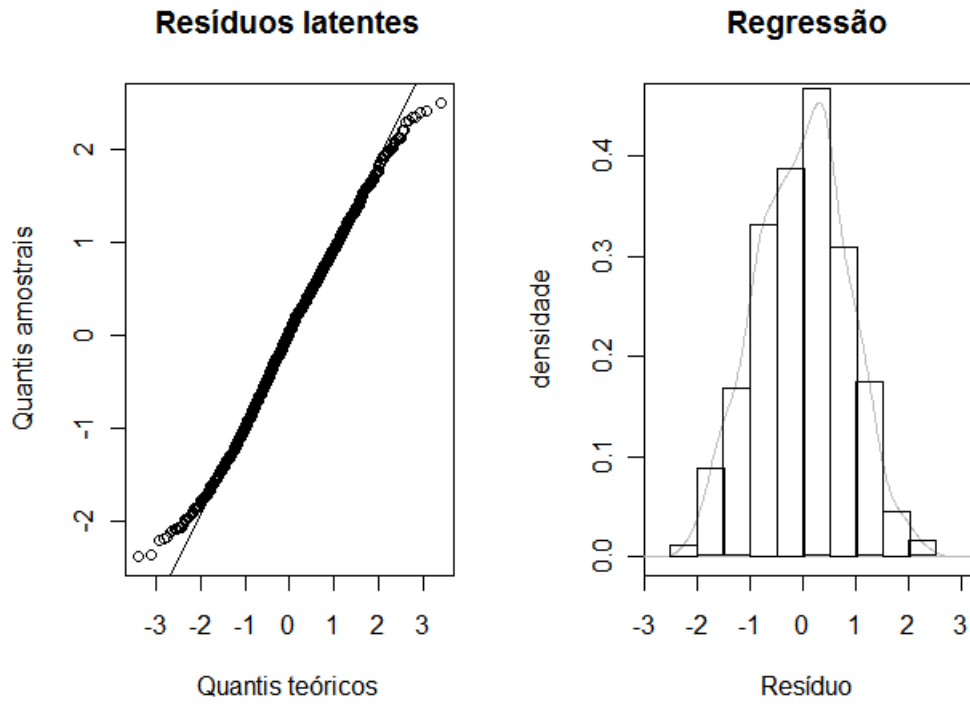


Figura 4.2 *Gráfico dos quantis normal-normal dos resíduos latentes e a Densidade dos resíduos do modelo de regressão*

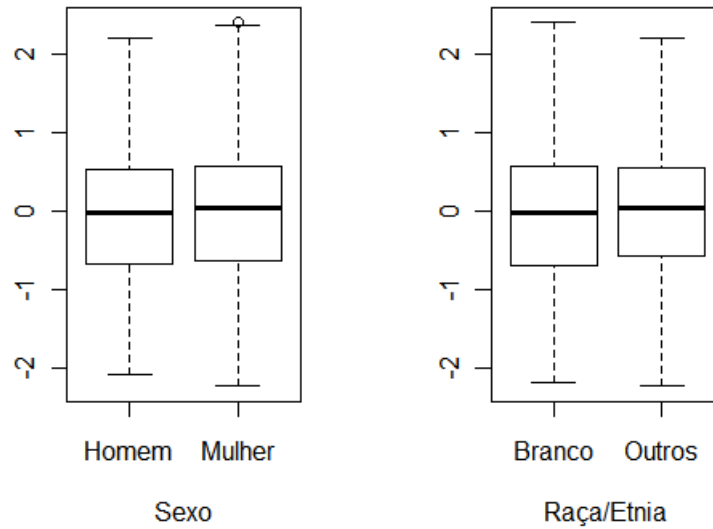


Figura 4.3 Diagramas de caixa para os resíduos do modelo de regressão por sexo e raça

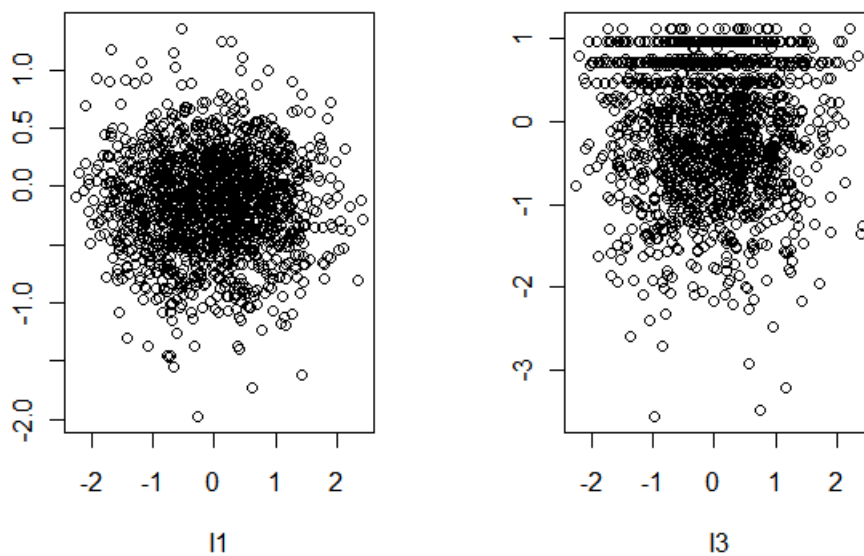


Figura 4.4 Gráfico de dispersão dos resíduos do modelo de regressão vs. indicadores I1 e I3

As estimativas, erros-padrão, intervalos de confiança (CI) e intervalos de HPD para os parâmetros de regressão do modelo são mostrados na Tabela 4.3. Nota-se que a estimação dos coeficientes de regressão é pouco sensível à escolha de prioris para o acerto casual, com diferenças apenas na segunda ou terceira casa decimal das estimativas, bem como nos IC's e HPD's.

Todos os coeficientes de regressão tiveram estimativas positivas, entretanto os IC's e HPD's para os coeficientes de regressão para as variáveis *Sexo* e *I3** contém o valor 0, decidiu-se então optar pela não significância estatística destes coeficientes. Ou seja, que não há diferenças significativas na proficiência média em português entre homens e mulheres, bem como o índice de fatores favoráveis a aprendizagem. É digno de nota que estes resultados podem ter ocorrido devido a flutuações decorrentes do processo de amostragem e, neste caso, uma amostra maior e um estudo mais minucioso deve ser realizado principalmente por que os zeros estão situados na fronteira dos intervalos.

Tabela 4.3 *Resultados para os parâmetros do modelo de regressão da proficiência em Português*

Variável	priori	Est.	EP	IC (95%)	HPD
<i>Sexo</i>	C1	0,088	0,006	[-0,020;0,195]	[-0,004;0,178]
	C2	0,084	0,005	[-0,031;0,179]	[-0,006;0,168]
	C3	0,091	0,006	[-0,006;0,177]	[0,009;0,173]
<i>Etnia</i>	C1	0,182	0,012	[0,098;0,269]	[0,104;0,255]
	C2	0,183	0,012	[0,093;0,261]	[0,107;0,258]
	C3	0,181	0,011	[0,092;0,266]	[0,110;0,256]
<i>I1*</i>	C1	0,757	0,048	[0,621;0,860]	[0,649;0,854]
	C2	0,741	0,047	[0,617;0,838]	[0,635;0,835]
	C3	0,744	0,047	[0,647;0,859]	[0,650;0,838]
<i>I3*</i>	C1	0,047	0,003	[-0,022;0,110]	[-0,010;0,104]
	C2	0,049	0,003	[-0,016;0,122]	[-0,007;0,112]
	C3	0,050	0,003	[-0,022;0,125]	[-0,016;0,113]

As Tabela 4.4 e 4.5 apresentam os resultados da estimação dos parâmetros de discri-

minação (**a**), dificuldade (**b**) e acerto casual (**c**). Observa-se que metade (12) itens tiveram discriminação acima de um, entretanto todos estão dentro dos intervalo de 0,4 a 1,6 (exceto o item 18, que é o mais discriminante do caderno com $a_{18} = 1,7$) e, portanto, gozam de boas propriedades neste aspecto. Lembrando que a escala da proficiência está na métrica (0,1), pode-se dizer que os itens mais difíceis no caderno são os itens 4, 10 e 14 que são, entretanto, pouco discriminativos. Para a probabilidade de resposta correta aos itens por sujeitos com baixa proficiência, os itens com menor probabilidade correspondem aos itens numerados 2, 6, 13 e 16 com chance inferior a 10%. Ademais, excetuando-se o item 1, todos os outros itens tiveram probabilidade de acerto casual inferior um chute ao acaso que neste estudo é 25

Além disso, de forma geral, as estimativas destes parâmetros estão próximas de suas medianas (não mostradas aqui) e do ponto médio dos extremos dos intervalos de confiança indicando que suas distribuições a posteriori são, ao menos, aproximadamente simétricas. Uma exceção seria a discriminação e dificuldade do item 1.

Tabela 4.4 *Resultados para os parâmetros de discriminação e dificuldade dos itens: Aplicação à dados reais - Priori C2*

Item	Discriminação				Dificuldade orig.			
	<i>Est.</i>	EP	IC	HPD	<i>Est.</i>	EP	IC	HPD
1	0,412	0,008	[0,244;0,667]	[0,215;0,669]	0,28	0,044	[-0,973;1,146]	[-1,164;1,217]
2	0,807	0,004	[0,717;0,912]	[0,705;0,935]	-1,18	0,008	[-1,373;-0,953]	[-1,416;-0,941]
3	1,052	0,010	[0,810;1,320]	[0,730;1,325]	0,50	0,007	[0,318;0,673]	[0,315;0,729]
4	0,781	0,013	[0,448;1,156]	[0,393;1,182]	2,07	0,014	[1,747;2,484]	[1,694;2,515]
5	1,098	0,007	[0,941;1,302]	[0,896;1,313]	-0,79	0,008	[-0,988;-0,567]	[-1,020;-0,551]
6	0,694	0,004	[0,608;0,790]	[0,604;0,823]	-1,06	0,009	[-1,252;-0,805]	[-1,316;-0,785]
7	1,259	0,009	[1,038;1,519]	[0,978;1,523]	-0,27	0,007	[-0,488;-0,106]	[-0,507;-0,075]
8	1,548	0,010	[1,314;1,858]	[1,295;1,902]	0,24	0,003	[0,166;0,325]	[0,153;0,330]
9	0,765	0,010	[0,526;1,042]	[0,466;1,050]	0,82	0,011	[0,509;1,070]	[0,403;1,078]
10	0,963	0,010	[0,743;1,253]	[0,700;1,309]	1,63	0,008	[1,439;1,858]	[1,405;1,874]
11	1,368	0,010	[1,136;1,675]	[1,107;1,703]	1,06	0,004	[0,972;1,155]	[0,958;1,165]
12	1,121	0,008	[0,939;1,327]	[0,904;1,379]	-0,07	0,007	[-0,236;0,098]	[-0,274;0,127]
13	0,655	0,004	[0,558;0,769]	[0,533;0,773]	-1,18	0,010	[-1,420;-0,880]	[-1,462;-0,794]
14	0,956	0,009	[0,725;1,206]	[0,705;1,270]	1,56	0,008	[1,372;1,798]	[1,332;1,830]
15	0,749	0,011	[0,480;1,048]	[0,408;1,052]	1,08	0,011	[0,741;1,347]	[0,713;1,387]
16	1,000	0,005	[0,884;1,125]	[0,871;1,141]	-0,83	0,005	[-0,955;-0,688]	[-1,002;-0,673]
17	1,476	0,015	[1,121;1,904]	[1,024;1,939]	-0,53	0,009	[-0,781;-0,310]	[-0,821;-0,281]
18	1,724	0,013	[1,413;2,097]	[1,394;2,157]	1,12	0,003	[1,036;1,222]	[1,029;1,239]
19	1,060	0,008	[0,855;1,297]	[0,824;1,310]	-0,24	0,008	[-0,461;-0,054]	[-0,462;0,016]
20	1,019	0,006	[0,882;1,189]	[0,838;1,213]	-0,77	0,009	[-0,966;-0,526]	[-0,973;-0,473]
21	1,023	0,008	[0,834;1,251]	[0,794;1,281]	0,97	0,005	[0,851;1,097]	[0,836;1,113]
22	0,935	0,010	[0,696;1,220]	[0,652;1,246]	0,64	0,008	[0,423;0,815]	[0,397;0,856]
23	0,762	0,006	[0,628;0,918]	[0,602;0,934]	-0,01	0,010	[-0,300;0,227]	[-0,315;0,309]
24	0,841	0,006	[0,709;1,032]	[0,681;1,044]	-0,54	0,011	[-0,779;-0,218]	[-0,831;-0,214]

Tabela 4.5 *Resultados para os parâmetros de discriminação e dificuldade dos itens: Aplicação à dados reais - Priori C2*

Item	Acerto casual			
	<i>Est.</i>	EP	IC	HPD
1	0,284	0,009	[0,028;0,472]	[0,026;0,496]
2	0,077	0,004	[0,007;0,192]	[0,000;0,195]
3	0,222	0,003	[0,154;0,285]	[0,151;0,304]
4	0,155	0,002	[0,095;0,191]	[0,094;0,200]
5	0,121	0,005	[0,015;0,253]	[0,009;0,257]
6	0,077	0,004	[0,007;0,190]	[0,001;0,193]
7	0,199	0,003	[0,102;0,285]	[0,095;0,292]
8	0,130	0,002	[0,088;0,170]	[0,080;0,172]
9	0,153	0,004	[0,034;0,248]	[0,025;0,255]
10	0,178	0,001	[0,146;0,211]	[0,143;0,215]
11	0,137	0,001	[0,109;0,162]	[0,102;0,164]
12	0,193	0,003	[0,116;0,268]	[0,111;0,288]
13	0,081	0,004	[0,008;0,204]	[0,001;0,206]
14	0,191	0,001	[0,158;0,222]	[0,151;0,224]
15	0,218	0,004	[0,097;0,298]	[0,089;0,308]
16	0,050	0,003	[0,003;0,132]	[0,001;0,136]
17	0,249	0,005	[0,112;0,361]	[0,112;0,397]
18	0,144	0,001	[0,124;0,167]	[0,119;0,172]
19	0,151	0,004	[0,041;0,236]	[0,010;0,236]
20	0,117	0,004	[0,014;0,239]	[0,004;0,242]
21	0,162	0,002	[0,118;0,198]	[0,113;0,201]
22	0,187	0,003	[0,093;0,254]	[0,087;0,261]
23	0,147	0,004	[0,024;0,246]	[0,018;0,261]
24	0,107	0,005	[0,011;0,256]	[0,001;0,257]

Capítulo 5

Considerações Finais e Recomendações

Nesta dissertação descrevemos os modelos dicotômicos da TRI com a função de ligação proibito com a inclusão de modelos de regressão linear no traço latente. Para as estimativas de interesse nesse modelo desenvolvemos um programa escrito em C++ que foi implementado utilizando as classes do pacote Rcpp (Eddelbuettel e François, 2011) e RcppArmadillo (François et al., 2011).

Alguns estudos de simulação também foram realizados indicando que, de fato, há uma melhoria nas estimativas do traço latente, ou seja, o erro quadrático médio diminuiu e a magnitude desta redução depende diretamente do quão correlacionadas estão as covariáveis que compõem o modelo de regressão. Adicionalmente, a estimação conjunta mostrou-se mais eficiente na estimação dos coeficientes de regressão que apresentaram relativamente um erro absoluto menor quando comparadas com a estimação em duas etapas, ou seja, o ajuste de um modelo de regressão aos escores obtidos via TRI.

Na aplicação aos dados do SARESP 2010 para alunos da 4ª série/5º ano do Ensino Fundamental verificou-se que das variáveis utilizadas, a saber: *Sexo*, *Etnia* (Branco ou não branco), índice de fatores sócioeconômico familiar do aluno (*I1*) e índice de fatores favoráveis a aprendizagem (*I3*), as variáveis com impacto na proficiência em língua portuguesa foram somente *Etnia* e *I1*. Para o *sexo* este resultado não é surpreendente devido as diferenças entre homens e mulheres no desempenho em provas de língua portuguesa acentuar-se principalmente em séries mais elevadas, mas o fato de que *I3* não teve um impacto significativo, com as ressalvas do capítulo anterior, deve ser melhor investigada com, por exemplo, um aumento do tamanho amostral que foi de 1.500 alunos.

5.1 Recomendações

Recomenda-se para trabalhos futuros:

- i)* Desenvolver uma metodologia mais afinada para avaliar a significância estatística dos coeficientes de regressão, como um p-valor bayesiano específico para a regressão;
- ii)* Considerar outros métodos de estimação menos exigentes computacionalmente;
- iii)* Avaliar a capacidade preditiva dos modelos de regressão para o traço latente;
- iv)* Considerar um modelo de regressão com erros normais assimétricos para o traço latente.

Apêndice A

Aspectos computacionais

A.1 Amostrador de Gibbs

O amostrador de Gibbs (Gelman e Geman, 1984; Gelfand and Smith, 1990) permite obter amostras da distribuição conjunta e marginais a posteriori, obtendo-se desta maneira um resumo estatístico de tais distribuições, utilizando a estrutura Markoviana. Suponha que a distribuição de $p(\Theta|D) = p(\theta_1, \theta_2, \dots, \theta_n|D)$, ou seja, $(\Theta|D)$ tem distribuição n-dimensional e D é toda a informação disponível. A distribuição de $p_j(\theta_j|\{\theta_i\}_{j \neq i}, D)$ é chamada **condicional completa** de θ_j . Tais condicionais completas são sempre proporcionais ao produto verossimilhança dos dados observados com a distribuição à priori conjunta, isto é, $p_j(\theta_j|\{\theta_i\}_{j \neq i}, D) \propto p(D|\Theta)p(\Theta), \forall j$. Dado um conjunto de valores iniciais $(\theta_1^0, \theta_2^0, \dots, \theta_n^0)$, as amostras são obtidas da seguinte maneira:

$$\begin{aligned}\theta_1^k &\sim p_1(\theta_1|\theta_2^{k-1}, \dots, \theta_n^{k-1}, D) \\ \theta_2^k &\sim p_2(\theta_2|\theta_1^k, \theta_3^{k-1}, \dots, \theta_n^{k-1}, D) \\ &\vdots \\ \theta_n^k &\sim p_n(\theta_n|\theta_1^k, \theta_2^k, \dots, \theta_{n-1}^k, D)\end{aligned}$$

O esquema acima deve ser repetido para $k=1,2,\dots$. Sob certas condições de regularidade (Gelfan e Smith, 1990; Tierney, 1994) a cadeia Markov acima converge para uma distribuição estacionária e esta é a distribuição a posteriori $p(\theta_1, \theta_2, \dots, \theta_n|D)$.

As condições de regularidade sobre as cadeias são: 1) irreduzibilidade, isto é, ela pode alcançar qualquer conjunto não vazio com probabilidade positiva, e todos os estados no espaço amostral são alcançáveis de qualquer outro por amostragem repetida a partir das

condicionais descritas. 2) Aperiódica, ou seja, não existe periodicidades, de tal forma que estados possam ser atingidos em movimentos periódicos regulares. 3) a cadeia é recorrente positiva. Uma cadeia é recorrente positiva quando um valor inicial se distribui conforme a distribuição alvo então todos os valores subsequentes também se distribuem de acordo com ela (veja, por exemplo, Roberts, 1995).

Apêndice B

Diagnóstico de convergência dos métodos de MCMC

Os algoritmos MCMC permitiram uma enorme expansão da classe de modelos candidatos para um conjunto de dados, entretanto há a necessidade de averiguar a partir de que iteração se pode considerar que a convergência para a distribuição alvo foi atingida. A avaliação da convergência é um problema teórico complexo e existem diversos métodos de diagnóstico de convergência de métodos MCMC (Cowles & Carlin, 1996). Há técnicas para avaliar o diagnóstico de uma única cadeia como, por exemplo, os métodos descritos em Heidelberger & Welch (1983), Geweke (1992) e em Raftery & Lewis (1992). Há também métodos que baseiam-se na simulação de várias cadeias como aqueles propostos por Gelman & Rubin (1992) e Liu, Liu, & Rubin (1992), este último aplicado somente ao método de amostragem de Gibbs. Neste trabalho foram usados os métodos de Geweke (1992) e Gelman & Rubin (1992) para o diagnóstico de convergência dos algoritmos propostos, além do monitoramento das médias ergódicas para cada cadeia.

B.1 Método de Geweke

Geweke (1992) propôs um diagnóstico de convergência para cadeias de Markov baseado em métodos de séries temporais para verificar a convergência do algoritmo de Gibbs. Seja g uma função do parâmetro θ cuja a média pretende-se estimar, a partir dos valores θ^t simulados em uma cadeia de Markov. A estimativa do valor esperado de $g(\theta)$ é então dada pela média ergódica dos $g(\theta^{(j)})$ calculados em todas as iterações do amostrador de Gibbs resultando em uma série temporal.

Observa-se a série ao longo de um número N suficientemente longo de iterações e calcula-

se a média $\bar{g}_a = \frac{\sum g(\theta^t)}{n_a}$ à custa das n_a primeiras iterações, bem como $\bar{g}_b = \frac{\sum g(\theta^t)}{n_b}$, a partir das n_b últimas iterações. O método de Geweke baseia-se na suposição de que o processo natural MCMC e a função g implicam na existência de uma densidade espectral $S_g(w)$ que não dever ter descontinuidades na frequência 0 e, caso esta suposição esteja satisfeita, o estimador de $E[g(\theta)]$ é $\bar{g}_N = \frac{\sum g(\theta^t)}{N}$ cuja variância assintótica é dada por $S_g(0)/N$. A raiz quadrada da variância assintótica pode ser usada para estimar o erro padrão da média, estimativa denominada por Geweke como NSE*.

Se a cadeia é estacionária, então a média da 1ª parte e a média da 2ª parte devem ser semelhantes. Admitindo que n_a/N e n_b/N são fixos e $N \rightarrow \infty$, mostra-se que

$$\frac{g_a - g_b}{\sqrt{(s_a^2/n_a + s_b^2/n_b)}} \rightarrow N(0, 1),$$

onde s_a^2 e s_b^2 são estimativas das variâncias assintóticas de g_a e g_b calculadas a partir das densidades espectrais estimadas para cada parte da cadeia.

Entre as desvantagens do método de Geweke está a sensibilidade a escolha das frações utilizadas, Geweke (1992) recomenda que a primeira parte seja as 10% primeiras amostras e a segunda parte as 50% últimas que correspondem as opções *default* deste diagnóstico disponível no pacote CODA (Plummer et al., 2006).

B.2 Método de Gelman & Rubin

Gelman & Rubin (1992) sugerem uma abordagem para averiguar a convergência do amostrador de Gibbs usando componentes de variância de sequências de múltiplas cadeias simuladas a partir de uma variedade de pontos iniciais dispersos, este método envolve dois passos. O primeiro, é realizado antes do início da amostragem para a obtenção de estimativas sobredispersas da distribuição alvo e gerar, a partir desta, pontos iniciais para $m \geq 2$ cadeias independentes (no caso de múltiplas modas, deve-se aumentar o valor de m).

O segundo passo é simular a quantidade escalar de interesse (U) pelo amostrador de Gibbs para digamos $2n$ iterações. Utilizando as n últimas iterações para re-estimar a distribuição alvo de U como uma distribuição t-student conservativa cujo parâmetro de

* Sigla em inglês para “*Numeric Standard Error*”

escala envolve as variâncias entre e dentro das cadeias simuladas, a convergência é monitorada pela estimação de um fator de redução de escala (o fator de redução que sofreria o parâmetro de escala da distribuição t-student para U se as simulações continuassem indefinidamente), nomeadamente

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{mn} \frac{B}{W}\right) \frac{gl}{gl-2}},$$

onde B é a variância entre as médias das m cadeias paralelas, W é a média das variâncias dentro das cadeias e gl é os graus de liberdade da densidade t-student aproximada. Esta razão decresce para 1 quando $n \rightarrow \infty$, de forma que os valores de $\hat{R} \approx 1$ sugerem que o período de aquecimento pode terminar e supondo que todas as cadeias tenham sido rodadas por um total de $2n$ iterações, as inferências podem ser realizadas usando uma combinação de valores das cadeias nas n últimas iterações.

Este método requer que os valores iniciais sejam sobredispersos relativamente à distribuição a posteriori. Os autores sugerem que se obtenham valores de uma mistura de distribuições t-multivariadas, cada uma delas centrada em uma moda da distribuição alvo. Este método funciona bem quando as distribuições condicionais forem “bem comportadas”, caso contrário, a obtenção de uma distribuição sobredispersa, quando as distribuições condicionais não tem uma forma conhecida, é um problema de difícil solução devido aos problemas numéricos geralmente envolvidos. Eles interpretam o fato que o fator de escala se aproxima de 1 quando a variância combinada dentro das cadeias domina a variância entre as médias das cadeias, ponto, a partir do qual, todas as cadeias “esqueceram” os valores iniciais e convergiram para distribuição alvo. Os autores ainda afirmam que não há como determinar que isto ocorreu em uma simples cadeia. Este método também está implementado no pacote CODA (Plummer et al., 2006).

Referências Bibliográficas

Andrade, D.F., Tavares, H.R. e Valle R.C. (2000). Teoria de Resposta ao Item: conceitos e aplicações. São Paulo: Associação Brasileira de Estatística.

Azevedo, C. L. N. (2008). Modelos longitudinais de grupos múltiplos multiníveis na teoria da resposta ao item: métodos de estimação e seleção estrutural sob uma perspectiva bayesiana. Tese de doutorado, IME-USP.

Azevedo, C.L.N., Bolfarine, H. e Andrade, D.F. (2011) Bayesian inference for a skew-normal IRT model under the centred parameterization, *Computational Statistics and Data Analysis*, 55, 353-365.

Bazán, J. L. (2005). Uma família de modelos de resposta ao ítem normal assimétrica. Tese de doutorado, IME-USP.

Beichl, I. and Sullivan, F. (2000), "The Metropolis algorithm", *Computing in Science & Engineering*, 2:1, 65-69.

Chen, M-H. and Shao, Q-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1), 69-92.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.*, 91, 883-904.

Croon, M. (2002), "Using Predicted Latent Scores in General Latent Structure Models", in G. A. Marcoulides and I. Moustaki (eds.) *Latent Variable and Latent Structure Models*, Mahwah, NJ: Lawrence Erlbaum Associates, 195-223.

du Toit, M. (2003). IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT.

Lincolnwood, IL: Scientific Software International.

Eddelbuettel D. and François, R. (2011). Rcpp: Seamless R and C++ Integration. R package version 0.9.9, URL <http://CRAN.R-Project.org/package=Rcpp>.

Fox, J.-P. (2010). Bayesian Item Response Modeling: Theory and Applications, Statistics for Social and Behavioral Sciences. Springer Science+Business Media, New York.

François, R., Eddelbuettel, D. and Bates, D. (2011). RcppArmadillo: Rcpp integration for Armadillo templated linear algebra library. R package version 0.2.34, URL <http://CRAN.R-Project.org/package=RcppArmadillo>.

Gamerman D. and Lopes, H. L. (2006). Markov Chain Monte Carlo: Stochastic Simulation and Statistical Inference. Second Edition, Taylor and Francis, New York.

Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A. and Rubin, DB. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Hambleton, R.K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publishers.

Heidelberger P. and Welch, P.D. (1983) "Simulation Run Length Control in the Presence of an Initial Transient" *Operations Research* 31:1109-1144.

JESUS, G. R. e LAROS, J. A. (2004) Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Aval. psicol.* [online]., vol.3, n.2, 93-106. ISSN 2175-3431 .

Liu, C., Liu, J., and Rubin, D.B. (1992). A Variational Control Variable for Assessing the Convergence of the Gibbs Sampler. In *Proceedings of the American Statistical Association, Statistical Computing Section*, 74-78.

-
- McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. New York: Wiley.
- Patz, R. J. and Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24, 146-178.
- Pereira, V. R. (2004). *Métodos Alternativos no Critério Brasil para Construção de Indicadores Sócio-Econômico: Teoria da Resposta ao Item*. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.
- Plummer, M., Best, N., Cowles, K. and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News*, vol 6, 7-11.
- R Development Core Team, (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org>.
- Raftery, A.E. and Lewis, S. (1992), "How Many Iterations in the Gibbs Sampler?" In *Bayesian Statistics 4* (eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith), Oxford: Oxford University Press, 763- 773.
- Roberts, G. O. (1995). Markov chain concepts related to sampling algorithms. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 45(57)). London: Chapman and Hall.
- Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, 72, 217-232.
- Sanderson, C. (2010). *Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments*. Technical Report, NICTA.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded responses. *Psychometrika Monograph Supplement*, No. 17.
- Samejima, F. (1997). Graded Response Model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. New York: Springer-Verlag, 85-100.

Skrondal, A. and P. Laake (2001). "Regression among Factor Scores", *Psychometrika*, 66(4), 563-576.

Tanner, M. A. (1996). *Tools for statistical inference* (3rd ed.). New York: Springer-Verlag.

Tierney. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22, 1701-1786.

Silva, C. C., Madruga, M. R. e Tavares, H. R. (2010). Construção de fatores associados à proficiência média das escolas públicas do município de Belém na prova brasil 2007. Congresso da Teoria da Resposta ao Item.

Smith, B. J. (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference *Journal of Statistical Software*, 21(11), 1-37

Soares, T. M., Mendonça, M.C.M. (2003). Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000. *Pesquisa Operacional*, Rio de Janeiro, v. 23, n. n.3, p. 421-441.