

Bárbara da Costa Campos Dias

**MODELO BAYESIANO DA TEORIA DE RESPOSTA AO
ITEM: UMA ABORDAGEM GENERALIZADA PARA O
TRAÇO LATENTE VIA MISTURAS**

Belo Horizonte

2015

UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Bárbara da Costa Campos Dias

**MODELO BAYESIANO DA TEORIA DE RESPOSTA AO
ITEM: UMA ABORDAGEM GENERALIZADA PARA O
TRAÇO LATENTE VIA MISTURAS**

Orientador: Flávio Bamberra Gonçalves

Belo Horizonte

2015

Aos meus pais, Alice e ao Felipe.

"Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar.

Mas o mar seria menor se lhe faltasse uma gota."

Madre Teresa de Calcuta.

Agradecimentos

Primeiramente agradeço a Deus por me dar força e me proporcionar uma vida tão abençoada.

Ao meu orientador Flávio Bambirra, por exercer de forma impecável o seu papel. Este trabalho é resultado de seus ensinamentos e dedicação. Serei eternamente grata.

Aos meus pais Fernando e Neila, por serem os responsáveis pelo meu caráter e por me mostrar a importância do estudo na minha vida. Minha irmã Alice, que ao nascer trouxe mais amor e responsabilidade para minha vida. Amo vocês incondicionalmente.

Ao meu grande amor Felipe, pela paciência com meus intermináveis estudos, pelo incentivo e pela paz que me proporciona. Sua companhia diária fez tudo ficar mais fácil. Obrigada por tudo. A família do Felipe, por tanto amor e carinho.

A minha vózinha Maria da Penha, por transbordar amor e me fazer sentir sempre especial. Aos meus amados tios, padrinhos e primos, em especial a Thayene, por ser um exemplo de determinação e luta para mim, você é minha irmã de coração.

Aos meus amigos desde o colégio, Felipe, Humberto, Bruna, Camila e Jéssica. A Roberta por ser minha irmã por escolha, nossa amizade é realmente especial. Obrigada por me proporcionarem uma amizade eterna e especial.

Aos amigos da graduação, Bethania, Daniel, Marcel, Vinícius, Douglas, Gabi e Isa. A Ju, pela amor e amizade até os dias de hoje. O desespero se tornou diversão com a presença de vocês, obrigada por tudo.

Ao Departamento de Estatística da UFJF por me fornecerem uma base sólida na minha caminhada, principalmente ao professor Clécio Ferreira.

Aos novos amigos da UFMG, que compartilharam comigo esta caminhada. A Gabi, pelo sorriso diário, companheirismo, amizade e por ser tão generosa. Ao Estevão e Bráulio, pela imensa ajuda nos problemas computacionais. A Zaida, pelos conselhos valiosos nessa reta final. Agradeço a minha

turma inteira, todos contribuíram e me ajudaram de alguma forma.

Aos professores da UFMG, que durante este tempo contribuíram para o meu conhecimento. Em especial a professora Rosângela Loschi, por me apresentar de forma incontestável a Estatística Bayesiana, suas aulas foram essenciais para esta dissertação.

Agradeço a presença da banca. Ao professor Marcos Prates, pelas valiosas sugestões e comentários em avaliações anteriores, contribuindo satisfatoriamente para a finalização deste trabalho. Ao professor Tufi Soares, pela grande experiência compartilhada e pelo incentivo na finalização deste trabalho.

Resumo

A Teoria da Resposta ao Item é uma teoria psicométrica que visa construir escalas e estimar traços latentes baseada em respostas dadas a itens que são diretamente influenciadas por estes traços. Uma suposição usual dos modelos da TRI é assumir que os traços latentes são variáveis aleatórias que seguem uma distribuição normal. Embora distribuições normais sejam frequentemente observadas e computacionalmente convenientes, é improvável que a distribuição do traço latente seja bem aproximada pela normal em todos os casos. O objetivo deste trabalho é, através de mistura de distribuições normais, propor um novo modelo da TRI que flexibiliza a hipótese de normalidade dos traços latentes. Em particular, esta abordagem permite uma solução para modelar distribuições com caudas pesadas ou assimetria sem o uso de distribuições com caudas pesadas ou assimétricas (ex: Skew-Normal ou t-Student). Esta dissertação também apresenta uma parametrização particular do modelo Probit de 3 parâmetros, com o objetivo de melhorar o algoritmo MCMC usado para se fazer inferência no modelo proposto, utilizando uma abordagem Bayesiana. E por último é apresentado estudos de simulações e dados reais para avaliar a eficiência e a aplicabilidade do modelo proposto.

Abstract

The Item Response Theory (IRT) is a psychometric theory which aims to construct scales and estimate latent traits based on answers given to items that are directly influenced by these traits. A common assumption of IRT models is to assume that the latent traits are random variables that follow a normal distribution. Although normal distributions are often observed and computationally convenient, it is unlikely that the latent traits are always well approximated by the normal distribution. The aim of this dissertation is to propose a new IRT model that relaxes the assumption of normality by using mixtures of normal distributions. In particular, this approach provides a solution to modeling heavy-tails or asymmetry without the use of heavy-tailed or asymmetric distributions (e.g. t-Student or skew-normal). This dissertation also introduces a particular parametrisation of the 3 parameter Probit model using auxiliary variables to improve the MCMC algorithm used to make inference in the proposed model under a Bayesian approach. Finally, simulations and real data studies are presented to assess the efficiency and applicability of the proposed model.

Conteúdo

Lista de Figuras	10
Lista de Tabelas	12
1 Introdução	14
1.1 Introdução	14
1.2 Modelo da TRI para respostas dicotômicas	16
1.3 Curva Característica do Item	18
1.4 Análise Bayesiana	20
1.4.1 Inferência Bayesiana	20
1.4.2 Método Monte Carlo via Cadeia de Markov (MCMC)	21
2 Modelo probito de 3 parâmetros	23
2.1 Novo modelo aumentado de 3 parâmetros	23
2.2 Inferência Bayesiana para o Modelo Aumentado	24
2.2.1 MCMC	25
2.3 Outra Alternativa para o Modelo aumentado de 3 parâmetros	28
2.4 Estudo de simulação para o novo modelo aumentado de 3 parâmetros	29
3 Modelo da TRI com mistura de distribuições normais para a proficiência	33
3.1 Mistura de densidades	33
3.2 Mistura de normais para a distribuição das proficiências	34
3.2.1 Identificabilidade do modelo	36
3.2.2 Inferência Bayesiana para o modelo com mistura	37

4	Estudo de simulação para o modelo da TRI com mistura	40
4.1	Estudo de simulação 1	40
4.2	Estudo de simulação 2	43
4.3	Estudo de simulação 3	45
4.4	Comparação entre os modelos	48
5	Análise dos dados do ENEM	54
5.1	Análise dos dados no ENEM utilizando o modelo Tradicional da TRI	57
6	Conclusões	59

Lista de Figuras

1.1	Modelo de dois parâmetros com $a = 0.5$ e $b = 0$	18
1.2	Modelo de dois parâmetros fixando a e variando b	19
1.3	Modelo de dois parâmetros fixando b e variando a	19
2.1	Trace Plots para cadeias de θ	30
2.2	Trace Plots para cadeias de a, b, c	31
2.3	Valor real <i>versus</i> valor estimado de a, b, c e θ	32
4.1	Valor real <i>versus</i> valor estimado de a, b, c e θ	41
4.2	CCI real e CCI estimada para os itens 2 e 40.	42
4.3	Densidade real, densidade estimada e histograma das médias <i>a posteriori</i> de θ	43
4.4	Valor real <i>versus</i> valor estimado de a, b, c e θ	44
4.5	CCI real e CCI estimada para os itens 2 e 20.	45
4.6	Densidade real, densidade estimada e histograma das médias <i>a posteriori</i> de θ	46
4.7	Valor real <i>versus</i> valor estimado de a, b, c e θ	47
4.8	CCI real e CCI estimada para os itens 2 e 20.	47
4.9	Densidade real, densidade estimada e histograma das médias <i>a posteriori</i> de θ	48
4.10	Mistura de normais bimodal ajustado pelo modelo simétrico.	50
4.11	Mistura de normais assimétrico ajustado pelo modelo simétrico.	51
4.12	Mistura de normais com caudas pesadas ajustado pelo modelo simétrico.	51
4.13	Distribuição normal ajustado pelo modelo com mistura.	52
5.1	Histograma dos Escores Brutos Padronizados de cada individuo.	55

5.2	Estimativa pontual (média <i>a posteriori</i>) e intervalar (95%) dos parâmetros dos itens. O ponto no gráfico representa a estimativa pontual de cada parâmetro e a linha horizontal dentro do intervalo representa a estimação intervalar.	56
5.3	Densidade estimada e histograma das médias <i>a posteriori</i> das proficiências dos alunos.	57
5.4	Densidade estimada do modelo de mistura, densidade fixa do modelo tradicional e histograma das médias <i>a posteriori</i> das proficiências dos alunos do modelo tradicional.	58
6.1	Trace Plots para cadeias de a , b e c	69
6.2	Trace Plots para cadeias de θ	70
6.3	Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2	70
6.4	Trace Plots para cadeias de a , b e c	71
6.5	Trace Plots para cadeias de θ	72
6.6	Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2	72
6.7	Trace Plots para cadeias de a , b e c	73
6.8	Trace Plots para cadeias de θ	74
6.9	Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2	74

Lista de Tabelas

4.1	Média <i>a posteriori</i> para os parâmetros da mistura 1	42
4.2	Média <i>a posteriori</i> para os parâmetros da mistura 2	45
4.3	Média <i>a posteriori</i> para os parâmetros da mistura 3	49
4.4	Média <i>a posteriori</i> para os parâmetros do modelo simples.	53
5.1	Média <i>a posteriori</i> para os parâmetros da mistura.	58

Capítulo 1

Introdução

1.1 Introdução

A Teoria de Resposta ao Item (TRI) foi formalizada por Lord em 1952 e por Rasch em 1960. A TRI é uma teoria psicométrica que visa construir escalas e estimar traços latentes baseada em respostas dadas a itens que são diretamente influenciadas por estes traços. A variável de interesse é não observável, ou seja, não pode ser medida diretamente, como por exemplo a proficiência em matemática ou ansiedade do indivíduo. Esta variável é conhecida como traço latente, proficiência ou habilidade do indivíduo.

A TRI modela os traços latentes, relacionando a probabilidade de um indivíduo responder corretamente ao item com seu traço latente e com as características dos itens. Algumas importantes características dos itens são: dificuldade, discriminação e acerto casual, este último ocorre quando o indivíduo possui uma baixa proficiência e por isso não sabe responder ao item. Dadas essas características dos itens é possível classificar indivíduos segundo alguma característica de interesse, por exemplo a proficiência em matemática ou o nível de estresse.

Uma usual suposição dos modelos da TRI é assumir que os traços latentes são variáveis aleatórias que seguem uma distribuição normal, tal como, considerado por Lord e Novik (1968, Chap. 16). No entanto, alguns trabalhos sugerem que este pressuposto não se aplica em muitos casos e que vários fatores podem influenciar na não normalidade desta distribuição, como por exemplo, a presença de populações distintas, assimetria e caudas pesadas (ver Micceri, T., 1989 e Samejima, F. 1997).

Além disso, quando esta hipótese não é plausível as estimativas dos parâmetros tendem a ser

tendenciosas, por isso é importante modelar adequadamente tal distribuição. Embora distribuições normais sejam frequentemente observadas e computacionalmente convenientes, é improvável que a distribuição do traço latente seja bem aproximada pela normal em todos os casos.

Segundo Micceri (1989), para muitos dados psicométricos as suposições de normalidade na distribuição dos traços latentes não se sustentam em dados reais. Em uma tentativa de fornecer uma base empírica Micceri (1989) realizou uma pesquisa sobre a robustez da suposição de normalidade, analisando simetria e peso de caudas de 440 conjuntos de dados, levando em consideração opiniões de psicométricos. No final do estudo conclui-se que a maior parte destes dados não seguiam distribuições normais. Entre as medidas psicométricas foi observado que 13,6% tinham caudas próximas da distribuição normal, 20,8% tinham ambas caudas menos pesadas que a normal e 65,6% tinham ao menos uma cauda pesada.

Woods & Thissen (2006) apresentam um método não-paramétrico para ajustes de modelo da TRI, a Ramsay-curve da Teoria de Resposta ao Item (RC-IRT). Este é um método que estima simultaneamente os parâmetros dos itens e a distribuição da variável latente, usando estimação de máxima verossimilhança marginal em um EM modificado. Para isso é utilizado a função B-Splines, que é uma combinação linear de funções polinomiais, formando uma base de espaços splines (de Boor, C. 1972). É desejável para identificar a escala para modelos RC-IRT, fixar a média e o desvio padrão da curva de Ramsay como 0 e 1, respectivamente; no entanto, fazer isto não é simples porque os momentos não são parâmetros, nem funções simples dos parâmetros.

Para Azevedo et al. (2011) uma solução seria o uso da distribuição normal assimétrica centrada, proposta por Azzalini (1985), para os traços latentes. Sua abordagem é comparada com outros métodos de estimação que usam a normalidade nos traços latentes. Os autores afirmam que a falta de normalidade da distribuição pode está relacionada com a presença de assimetria nos dados.

Mislevy (1984) utiliza misturas de normais para solucionar o problema da falta de normalidade da distribuição do traço latente. Porém, os parâmetros dos itens são tratados como conhecidos e a estimação dos parâmetros da distribuição latente é obtida maximizando a verossimilhança marginal dos dados.

O objetivo deste trabalho é, através de mistura de distribuições normais, propor um novo modelo da TRI que flexibiliza a hipótese de normalidade dos traços latentes.

Esta dissertação está organizada da seguinte forma: No restante do capítulo 1 serão apresen-

tados os conceitos de TRI e definições básicas de análise Bayesiana e MCMC. No capítulo 2 será apresentado uma versão aumentada do modelo Probit de 3 parâmetros da TRI, com o objetivo de facilitar o MCMC, no final do capítulo será apresentado um estudo de simulação utilizando o modelo proposto. No capítulo 3 é proposto um novo modelo da TRI que flexibiliza a hipótese de normalidade utilizando uma mistura de distribuições normais para modelar as proficiências dos indivíduos. No capítulo 4 serão realizados estudos de simulações para avaliar a eficiência do modelo da TRI com mistura de distribuições normais para a proficiência. No capítulo 5 é composto pelos estudos de aplicação a dados reais. Por fim, o capítulo 6 apresenta uma conclusão para este trabalho.

1.2 Modelo da TRI para respostas dicotômicas

A TRI considera modelos estatísticos para descrever como a probabilidade de responder corretamente ao item depende da proficiência do indivíduo e, também, de características dos itens. Existem modelos da TRI para diferentes tipos de itens, como por exemplo itens dicotômicos, politômicos ou com respostas contínuas. Neste trabalho iremos tratar de itens dicotômicos, ou seja, existem duas possíveis respostas no item.

Seja Y_{ij} uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i , com $i=1,\dots,I$ e $j=1,\dots,J$. Os modelos clássicos da TRI para itens dicotômicos tem a seguinte forma para a probabilidade do indivíduo j acertar o item i :

$$P(Y_{ij} = 1 | \theta_j, b_i, a_i, c_i) = c_i + (1 - c_i)F(a_i\theta_j - b_i), \quad (1.1)$$

onde a proficiência do indivíduo j é expressada por θ_j e os parâmetros de discriminação, dificuldade e acerto casual do item i são denotados por a_i , b_i e c_i , respectivamente.

F é uma função de distribuição, ou seja F é monótona não decrescente, isto implica que estamos assumindo na prática que quanto maior a proficiência do indivíduo, maior a probabilidade dele acertar ao item.

O parâmetro de discriminação mede a capacidade deste item diferenciar indivíduos com proficiências distintas. Já o parâmetro de dificuldade descreve os itens ao longo da escala de proficiência, quanto maior a dificuldade, maior é a proficiência exigida do indivíduo para responder corretamente

ao item. O parâmetro de acerto casual descreve a probabilidade mínima que todo indivíduo tem de responder de forma correta ao item, que na prática é a probabilidade de um indivíduo com uma proficiência extremamente baixa acertar ao item. Um ótimo teste é aquele que possui muitos itens, com níveis de dificuldade variados, todos com a discriminação muito alta.

As duas escolhas mais comuns na literatura de TRI são o modelo logístico, onde F é uma função logística, dado por:

$$P(Y_{ij} = 1|\theta_j, b_i, a_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \quad (1.2)$$

e o modelo probito, onde F é a acumulada da normal padrão, dado por:

$$P(Y_{ij} = 1|\theta_j, b_i, a_i, c_i) = c_i + (1 - c_i)\Phi(a_i\theta_j - b_i), \quad (1.3)$$

onde $\Phi(\cdot)$ é função de distribuição da $N(0,1)$.

Na prática os modelos logístico e probito são a mesma coisa, uma vez que $|L(1.702x) - \Phi(x)| < 0.01$, onde $L(x) = \frac{1}{1+e^{-x}}$ é a função logística no ponto x , com $-\infty < x < \infty$.

Feito uma dessas duas escolhas, temos um modelo da TRI de três parâmetros. Se fixarmos $c = 0$, obtemos o modelo de dois parâmetros e se fixarmos $c = 0$ e $a = 1$, obtemos o modelo de um parâmetro.

Pode-se notar que a verossimilhança não é identificável para ambos os modelos definidos em (1.2) e (1.3). Qualquer transformação do tipo $\theta_j^* = \theta_j + r$ e $b_i^* = b_i + r, r \in R$, com $j = 1, \dots, J, \forall i$ levam a uma mesma probabilidade de acerto no modelo (1.2) e qualquer transformação do tipo $\theta_j^* = \theta_j + r$ e $b_i^* = b_i + a_i r, r \in R$, com $j = 1, \dots, J, \forall i$ levam a uma mesma probabilidade de acerto no modelo (1.3). Uma maneira de aliviar este problema é fixar a distribuição para as proficiências, que funciona na maioria das situações. Neste caso, é usual adotar a distribuição normal com média e variância conhecidas, o que define uma escala para as proficiências.

No entanto, como já foi citado anteriormente, vários fatores podem influenciar em uma não normalidade desta distribuição, como por exemplo, a presença de populações distintas, assimetria e caudas pesadas (ver Micceri, T., 1989 e Samejima, F. 1997).

1.3 Curva Característica do Item

A curva característica do item (CCI) é uma ferramenta gráfica usada para descrever o comportamento dos parâmetros da TRI. Pelas características dada no modelo (1.1) já sabemos que quanto maior a proficiência do indivíduo, maior a probabilidade dele acertar ao item. Este comportamento é completamente descrito pela CCI, além disso é possível analisar a relação existente entre os parâmetros dos itens com a proficiência do indivíduo.

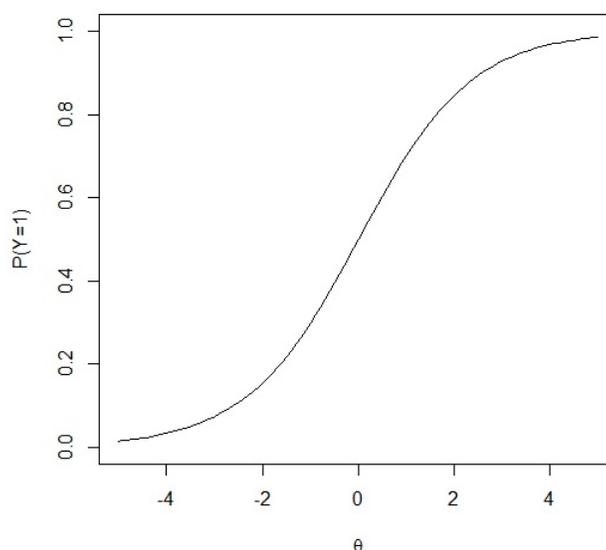


Figura 1.1: Modelo de dois parâmetros com $a = 0.5$ e $b = 0$.

Observe, pela Figura 1.1, que quando a proficiência do indivíduo é igual a dificuldade do item a probabilidade de se acertar este item é igual a 0.5, que será o ponto de maior inclinação da curva característica. Se a proficiência do indivíduo é maior que a dificuldade do item, ele tem mais chance de acertar do que errar o item e vice-versa.

Para a análise do gráfico na Figura 1.2, foi fixado o parâmetro de discriminação do item, verificando assim a relação entre a dificuldade do item com a proficiência do indivíduo. Note que a dificuldade é um parâmetro de posição. Itens com menor dificuldade implicam em probabilidade alta dos indivíduos acertarem ao item para níveis baixos de proficiência. E itens com maior dificuldade implicam em probabilidades baixas dos indivíduos acertarem ao item para boa parte da escala, exceto para níveis altos de proficiências.

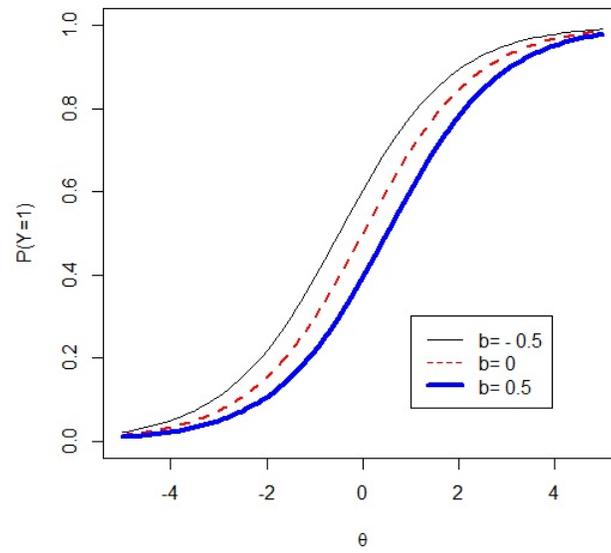


Figura 1.2: Modelo de dois parâmetros fixando a e variando b .

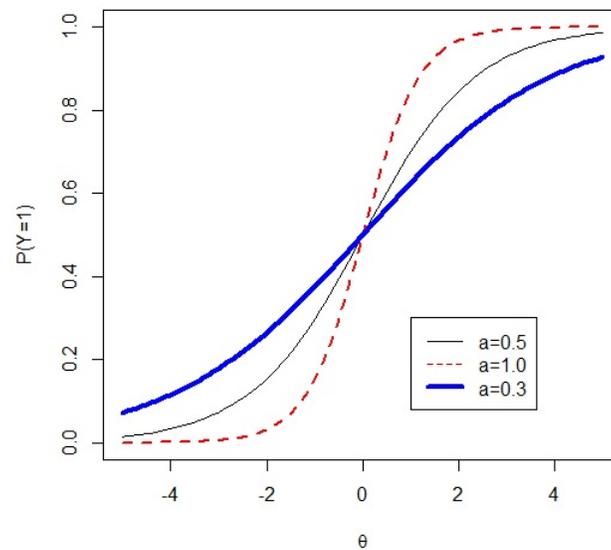


Figura 1.3: Modelo de dois parâmetros fixando b e variando a .

Para analisar a relação entre o parâmetro de discriminação do item com a proficiência do indivíduo, foi fixado o parâmetro de dificuldade no gráfico dado pela Figura 1.3. Através disto é possível

notar que a discriminação é um parâmetro de inclinação, quanto maior for o valor do parâmetro a , maior é a inclinação da curva, e vice-versa. Um item com discriminação perfeita seria aquele que para indivíduos com alta proficiência a probabilidade deles acertarem a este item seria 1, e caso contrário seria 0.

1.4 Análise Bayesiana

Em estatística Bayesiana a incerteza é descrita através de probabilidade. Seja Ψ o parâmetro de interesse, o qual pode ser escalar, matriz ou vetor. Além dos dados amostrais, para se realizar inferência Bayesiana é necessário a utilização de uma informação a priori sobre os parâmetros de interesse, dada pela densidade de probabilidade $\pi(\Psi)$, a qual representa, de alguma forma, o conhecimento do pesquisador sobre o parâmetro de interesse.

1.4.1 Inferência Bayesiana

A maneira utilizada para inferir sobre Ψ é construir uma medida de probabilidade que descreva a incerteza do mesmo. Isto é feito através da construção da distribuição *a posteriori*, denotada por $\pi(\Psi|y)$, onde y representa os dados amostrais. A distribuição $\pi(\Psi|y)$ contém toda a informação probabilística de interesse a respeito de Ψ . Assim, a inferência sobre o parâmetro é realizada por meio desta distribuição.

Usualmente, é utilizado o Teorema de Bayes como um mecanismo de atualização da distribuição *a priori* e como ferramenta para a construção da distribuição *a posteriori*, que é dada por

$$\pi(\Psi|y) = \frac{\pi(y|\Psi)\pi(\Psi)}{\pi(y)},$$

onde $\pi(y) = \int_{\Psi} \pi(y|\Psi)\pi(\Psi) d\Psi$, para o caso contínuo e $\pi(y) = \sum_{\Psi} \pi(y|\Psi)\pi(\Psi)$, para o caso discreto.

O Teorema de Bayes pode ser interpretado como uma regra de atualização na qual os dados permitem atualizar o conhecimento *a priori* de Ψ e a distribuição *a posteriori* é o resultado deste processo, combinando o que é conhecido (amostra) com o que é desconhecido (parâmetro).

1.4.2 Método Monte Carlo via Cadeia de Markov (MCMC)

Como foi citado na seção anterior, a inferência sobre o parâmetro de interesse é realizada por meio da distribuição *a posteriori*, mas na maioria das vezes, obter esta distribuição não é trivial, especialmente pela dificuldade em se obter a distribuição marginal dos dados, dada por $\pi(y)$.

Para resolver este problema, alguns métodos são utilizados para se obter uma aproximação da distribuição *a posteriori*. Um dos métodos mais utilizados é o de Monte Carlo via Cadeia de Markov (MCMC), apresentado nesta seção. A teoria de MCMC é extensa e complexa e, portanto, será apresentado apenas as ideias básicas de sua metodologia. Uma discussão mais geral sobre o tema pode ser encontrada por exemplo em Gamerman (1997).

No caso Bayesiano, a idéia básica deste método é obter uma amostra (aproximada) grande o suficiente da distribuição *a posteriori*. Os valores são gerados de forma iterativa, baseadas em Cadeias de Markov. A cadeia de Markov é um processo estocástico caracterizado por seu estado futuro depender apenas do seu estado atual, sendo que os estados passados não influenciam no estado futuro. Os métodos MCMC requerem ainda que a cadeia seja: Homogênea (as probabilidades de transição de um estado para outro são invariantes), irredutível (cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações) e aperiódica (não haja estados absorventes).

Será simulado um passeio aleatório no espaço do parâmetro Ψ , que converge para uma distribuição estacionária, que em inferência Bayesiana é a distribuição *a posteriori*.

Gibbs Sampling

Existem vários métodos MCMC, dentre eles o *Gibbs Sampling*, originado no contexto estatístico por Geman e Geman(1984) e Gelfand e Smith (1990). Este método é uma ferramenta extremamente útil na resolução de problemas que envolvem a estimação de mais de um parâmetro.

Através do *Gibbs Sampling* é possível gerar amostras aproximadas da distribuição conjunta dos parâmetros de interesse *a posteriori*, a partir das distribuições condicionais completas *a posteriori* de cada parâmetro, denotada por $\pi(\Psi_i | \Psi_-, y)$, onde $\Psi_- = (\Psi_1, \dots, \Psi_{i-1}, \Psi_{i+1}, \dots, \Psi_d)$, considerando que $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_d)$ são os parâmetros do modelo em estudo. Neste caso é necessário as distribuições condicionais completas tenham forma fechada para se efetuar o algoritmo. Para implementação de um algoritmo *Gibbs Sampling* são necessários os seguintes passos:

1. Inicialize Ψ atribuindo-lhe um valor inicial $\Psi^0 == (\Psi_1^0, \Psi_2^0, \dots, \Psi_d^0)$.
2. Inicie o contador igual a $t=1$ iterações.
3. Obtenha um novo valor Ψ^t a partir de Ψ^{t-1} :

$$\Psi_1^t = (\Psi_1 | \Psi_2^{t-1}, \Psi_3^{t-1}, \dots, \Psi_d^{t-1}, y)$$

$$\Psi_2^t = (\Psi_2 | \Psi_1^t, \Psi_3^{t-1}, \dots, \Psi_d^{t-1}, y)$$

$$\vdots$$

$$\Psi_d^t = (\Psi_d | \Psi_1^t, \Psi_2^t, \dots, \Psi_{d-1}^t, y)$$

4. Aumente o contador de t para $t+1$ e volte para o passo 3 até que a convergência seja atingida.

É gerada uma amostra grande o suficiente, para haver uma boa aproximação da distribuição exata, neste caso *a posteriori* conjunta. Dessa forma, é estipulado um número de iterações (burn-in) que serão descartadas para a convergência da cadeia.

Capítulo 2

Modelo probito de 3 parâmetros

Neste capítulo será apresentado uma versão aumentada do modelo probito de 3 parâmetros, com o objetivo de facilitar o MCMC usado para se fazer inferência no modelo (1.3), utilizando uma abordagem Bayesiana.

Uma solução para o modelo de dois parâmetros existe a mais tempo e foi proposta por Albert (1992), ele introduz uma variável auxiliar e desta forma é possível construir um *Gibbs Sampling* amostrando diretamente de todas as densidades das condicionais completas.

Quando introduzimos o parâmetro de acerto casual do item, ou seja, quando trabalhamos com o modelo probito de 3 parâmetros voltamos ao problema original, dessa forma a proposta do trabalho de Albert (1992) não é suficiente para facilitar o MCMC. Uma solução para isto é a introdução de mais uma variável latente. Uma proposta já existente na literatura foi dada por Beguin and Glas (2001). Nosso trabalho propõe uma solução parecida, porém com diferenças significativas que serão discutidas e comparadas mais a diante no texto (Seção 2.3).

2.1 Novo modelo aumentado de 3 parâmetros

Para aplicar o método MCMC no modelo (1.3) seria necessário o uso de um algoritmo *Gibbs Sampling* com passo *Metropolis-Hastings*. Com o objetivo de obter melhores taxas de convergência e amostrar diretamente das distribuições condicionais completas propomos um modelo aumentado, introduzindo um conjunto de variáveis auxiliares.

Nossa proposta de modelo aumentado introduz duas variáveis auxiliares. A primeira variável

auxiliar é denotada por Z_{ij} , $i=1,\dots,I$ e $j=1,\dots,J$, onde $Z_{ij} \sim \text{Bernoulli}(c_i)$, ou seja, $P(Z_{ij} = 1) = c_i$ e $P(Z_{ij} = 0) = 1 - c_i$. A segunda variável auxiliar é denotada por X_{ij} , $i=1,\dots,I$ e $j=1,\dots,J$, onde $(X_{ij}|Z_{ij} = 0) \sim N(a_i\theta_j - b_i, 1)$ e $(X_{ij}|Z_{ij} = 1) = \delta_0$, onde δ_0 é um ponto de massa em zero, ou seja $P(X_{ij} = 0|Z_{ij} = 1) = 1$.

Dessa forma, pode-se reescrever Y_{ij} da seguinte maneira:

$$Y_{ij} = \begin{cases} 1 & \text{se } (X_{ij} = 0, Z_{ij} = 1) \text{ ou } (X_{ij} \geq 0, Z_{ij} = 0) \\ 0 & \text{se } (X_{ij} < 0, Z_{ij} = 0) \end{cases} \quad (2.1)$$

Proposição 1 O modelo descrito em (2.1) é equivalente ao modelo descrito em (1.3), ou seja, $P(Y_{ij} = 1|\theta_j, b_i, a_i, c_i)$ é igual nos dois casos:

Demonstração:

$$\begin{aligned} P(Y_{ij} = 1|\theta_j, b_i, a_i, c_i) &= P(\{X_{ij} = 0, Z_{ij} = 1\} \cup \{X_{ij} \geq 0, Z_{ij} = 0\}) = \\ &= P(X_{ij} = 0, Z_{ij} = 1) + P(X_{ij} \geq 0, Z_{ij} = 0) = \\ &= P(Z_{ij} = 1)P(X_{ij} = 0|Z_{ij} = 1) + P(Z_{ij} = 0)P(X_{ij} \geq 0|Z_{ij} = 0) = \\ &= c_i + (1 - c_i)P(X_{ij} > 0) = c_i + (1 - c_i)\Phi(a_i\theta_j - b_i). \end{aligned}$$

Através deste modelo aumentado (2.1) é possível construir um *Gibbs Sampling*, onde podemos amostrar diretamente de todas as densidades das condicionais completas, não necessitando assim de passo *Metropolis-Hastings*.

2.2 Inferência Bayesiana para o Modelo Aumentado

Defina $\Psi = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \theta, \mathbf{X}, \mathbf{Z}\}$, onde $\mathbf{c} = (c_1, \dots, c_I)$, $\mathbf{a} = (a_1, \dots, a_I)$, $\mathbf{b} = (b_1, \dots, b_I)$ e $\theta = (\theta_1, \dots, \theta_J)$, com $i=1,\dots,I$ e $j=1,\dots,J$. Neste caso, a inferência sobre os parâmetros $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \theta)$ é feita através da densidade *a posteriori* conjunta destes componentes, denotada por:

$$\pi(\Psi|Y) = \frac{\pi(Y|\Psi) \pi(\Psi)}{\int \pi(Y|\Psi) \pi(\Psi) d\mathbf{a} d\mathbf{b} d\mathbf{c} d\mathbf{X} d\mathbf{Z} d\boldsymbol{\theta}} \quad (2.2)$$

A distribuição conjunta de Y com Ψ é dada por:

$$\pi(Y, \Psi) = \left[\prod_{i=1}^I \prod_{j=1}^J \pi(Y_{ij}|X_{ij}, Z_{ij}) \pi(Z_{ij}|c_i) \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j) \right] \left[\prod_{i=1}^I \pi(a_i) \pi(b_i) \pi(c_i) \right] \prod_{j=1}^J \pi(\theta_j), \quad (2.3)$$

onde:

$\pi(Y_{ij}|X_{ij}, Z_{ij}) = I_{(Y_{ij}=1)} \left[I_{(Z_{ij}=1)} I_{(X_{ij}=0)} + I_{(Z_{ij}=0)} I_{(X_{ij}>0)} \right] + I_{(Y_{ij}=0)} \left[I_{(Z_{ij}=0)} I_{(X_{ij}<0)} \right]$, sendo $I_{(.)}$ a função indicadora de $(.)$.

- **Especificação das distribuições *a priori***

Para a realização da inferência Bayesiana é utilizado o conhecimento que se tem sobre os parâmetros do modelo, e isso é feito através das distribuições *a priori*, cujos parâmetros devem ser especificados de acordo com este conhecimento e são denominados de hiperparâmetros.

Escolhas razoáveis para as distribuições *a priori* que facilitam a derivação do algoritmo MCMC são dadas por: $\theta_j \sim N(\mu_\theta, \sigma_\theta^2)$, $a_i \sim N_{(0,\infty)}(\mu_a, \sigma_a^2)$, $b_i \sim N(\mu_b, \sigma_b^2)$ e $c_i \sim Beta(\alpha_c, \beta_c)$, $\forall i = 1, \dots, I$ e $\forall j = 1, \dots, J$, onde $N_{(0,\infty)}$ representa a distribuição normal truncada nos reais positivos. Os parâmetros μ_θ e σ_θ^2 são usados para fixar a escala e identificar a verossimilhança, dado isto, as prioris de a_i e b_i precisam estar de acordo com esta escala.

2.2.1 MCMC

Devido à dificuldade para se explorar a distribuição dada em (2.1) de forma analítica, utiliza-se um algoritmo MCMC (Markov Chain Monte Carlo) para se obter uma amostra (aproximada) da mesma, mais especificamente um *Gibbs Sampling* como é definido na seção 2.2, onde adotamos o seguinte sistema de blocagem:

$$(\mathbf{X}, \mathbf{Z}) \quad \boldsymbol{\theta} \quad (\mathbf{a}, \mathbf{b}) \quad \mathbf{c}$$

Defina como $\pi(*|\cdot)$ a distribuição condicional completa de $*$. Por independência, temos o seguinte resultado:

$$\pi(\mathbf{X}, \mathbf{Z}|\cdot) = \prod_{i=1}^I \prod_{j=1}^J \pi(Z_{ij}, X_{ij}|\cdot)$$

$$\pi(\mathbf{c}|\cdot) = \prod_{i=1}^I \pi(c_i|\cdot)$$

$$\pi(\mathbf{a}, \mathbf{b}|\cdot) = \prod_{i=1}^I \pi(a_i, b_i|\cdot)$$

$$\pi(\boldsymbol{\theta}|\cdot) = \prod_{j=1}^J \pi(\theta_j|\cdot)$$

Como foi citado anteriormente, o *Gibbs Sampling* consiste em amostrar através das distribuições condicionais completas em cada iteração. Com o resultado de independência dado acima podemos concluir que para amostrar dos blocos (\mathbf{X}, \mathbf{Z}) , $\boldsymbol{\theta}$, (\mathbf{a}, \mathbf{b}) e \mathbf{c} é equivalente amostrar individualmente de cada distribuição completa de $(Z_{ij}, X_{ij}|\cdot)$, $(\theta_j|\cdot)$, $(a_i, b_i|\cdot)$ e $(c_i|\cdot)$. Todas as condicionais completas têm densidades proporcionais à expressão (2.3), o que resulta nas seguintes distribuições:

Lema 1: *As distribuições condicionais completas do modelo aumentado proibido de 3 parâmetros são dadas por:*

$$\pi(Z_{ij}, X_{ij}|\cdot) \begin{cases} \propto \phi(x_{ij} - m) I_{(Z_{ij}=0)} I_{(X_{ij}<0)}, & \text{se } Y_{ij} = 0 \\ w I_{(Z_{ij}=1)} I_{(X_{ij}=0)} + (1 - w) \frac{\phi(x_{ij} - m)}{\Phi(m)} I_{(Z_{ij}=0)} I_{(X_{ij}>0)}, & \text{se } Y_{ij} = 1, \end{cases} \quad (2.4)$$

onde $m = a_i \theta_j - b_i$ e $w = \frac{c_i}{c_i + (1 - c_i) \Phi(m)}$.

$$c_i|\cdot \sim \text{Beta} \left(\sum_{j=1}^J Z_{ij} + \alpha_c, J - \sum_{j=1}^J Z_{ij} + \beta_c \right) \quad (2.5)$$

$$\theta_j | \cdot \sim N(m_\theta, v_\theta), \quad (2.6)$$

$$\text{com } m_\theta = \frac{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i(x_{ij} + b_i) + \mu_\theta}{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1} \text{ e } v_\theta = \frac{\sigma_\theta^2}{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1}, \text{ onde } L_j\{i, z_{ij} \neq 0\}.$$

$$a_i, b_i | \cdot \sim N_2(\mu, \Sigma), \quad (2.7)$$

com $\mu = \begin{bmatrix} \mu_a^* \\ \mu_b^* \end{bmatrix}$ e $\Sigma = \begin{bmatrix} \sigma_a^{2*} & \gamma \\ \gamma & \sigma_b^{2*} \end{bmatrix}$ sendo o vetor de média e matriz de covariância de uma distribuição normal bivariada, respectivamente. Onde:

$$\sigma_a^{2*} = \frac{\sigma_a^2}{\left(\sigma_a^2 \sum_{j=1}^{L_i} \theta_j^2 + 1\right)(1-\gamma^2)}, \quad \sigma_b^{2*} = \frac{\sigma_b^2}{(\sigma_b^2 J + 1)(1-\gamma^2)}, \quad \gamma = \frac{\sigma_a \sigma_b \sum_{j=1}^{L_i} \theta_j}{\left[\left(\sigma_a^2 \sum_{j=1}^{L_i} \theta_j^2 + 1\right)(\sigma_b^2 J + 1)\right]^{\frac{1}{2}}},$$

$$\mu_a^* = \sigma_a^{2*} \left(\sum_{j=1}^J x_{ij} \theta_j + \mu_a \sigma_a^{-2} \right) - \sigma_a^* \sigma_b^* \gamma \left(\sum_{j=1}^{L_i} x_{ij} - \mu_b \sigma_b^{-2} \right) \text{ e}$$

$$\mu_b^* = \sigma_a^* \sigma_b^* \gamma \left(\sum_{j=1}^J x_{ij} \theta_j + \mu_a \sigma_a^{-2} \right) - \sigma_b^{2*} \left(\sum_{j=1}^{L_i} x_{ij} - \mu_b \sigma_b^{-2} \right).$$

Com $L_i\{j, z_{ij} \neq 0\}$.

Prova Lema 1: Ver apêndice A.

Para entendermos a distribuição (2.4) usamos a seguinte fatorização: $\pi(X_{ij}, Z_{ij} | \cdot) = \pi(X_{ij} | Z_{ij}, \cdot) \pi(Z_{ij} | \cdot)$. Quando $Y_{ij} = 0$, temos que $\pi(X_{ij} | Z_{ij}, \cdot)$ é uma densidade da distribuição normal truncada nos reais negativos e $\pi(Z_{ij} | \cdot)$ é um ponto de massa em zero.

Se $Y_{ij} = 1$, temos que $\pi(Z_{ij} | \cdot)$ é função de probabilidade da Bernoulli(w), onde $\pi(X_{ij} | Z_{ij}, \cdot)$ é um ponto de massa em zero (se $Z_{ij} = 1$) e $\pi(X_{ij} | Z_{ij}, \cdot)$ é uma densidade normal truncada nos reais positivos (se $Z_{ij} = 0$).

2.3 Outra Alternativa para o Modelo aumentado de 3 parâmetros

Uma abordagem similar ao nosso trabalho é feita por Beguin and Glas (2001), onde o modelo(1.3) é reescrito da seguinte forma:

$$\begin{aligned} P(Y_{ij} = 1|\theta_j, b_i, a_i, c_i) &= c_i + (1 - c_i)\Phi(a_i\theta_j - b_i) \\ &= \Phi(a_i\theta_j - b_i) + c_i(1 - \Phi(a_i\theta_j - b_i)) \end{aligned} \quad (2.8)$$

Baseada na representação em (2.8), os autores interpretam $(\Phi(a_i\theta_j - b_i))$ como sendo a probabilidade do indivíduo j saber a resposta correta do item i e, então, respondê-lo corretamente com probabilidade 1. Conseqüentemente $(1 - \Phi(a_i\theta_j - b_i))$ é interpretado como sendo a probabilidade do indivíduo não saber a resposta correta do item i e, então, respondê-lo corretamente com probabilidade c_i .

Assim, a probabilidade do indivíduo responder corretamente ao item i é a soma do termo $\Phi(a_i\theta_j - b_i)$ com o termo $c_i(1 - \Phi(a_i\theta_j - b_i))$.

Baseado nesta interpretação, é introduzido uma variável auxiliar binária, denotada por Z_{ij} , $i=1, \dots, I$ e $j=1, \dots, J$, tal que:

$$Z_{ij} \sim \text{Bernoulli}(\Phi(a_i\theta_j - b_i, 1)) \quad (2.9)$$

$$P(Y_{ij}|Z_{ij} = 1) \sim \text{Bernoulli}(1) \quad (2.10)$$

$$P(Y_{ij}|Z_{ij} = 0) \sim \text{Bernoulli}(c_i) \quad (2.11)$$

Além de introduzir a variável auxiliar Z_{ij} , para aplicar o método *Gibbs Sampling* foi necessário a introdução de uma segunda variável (proposta por Albert (1992)), denotada por X_{ij} , $i=1, \dots, I$ e $j=1, \dots, J$, tal que:

$$X_{ij}|\theta_j, a_i, b_i, c_i \sim N(a_i\theta_j - b_i, 1), \quad (2.12)$$

onde $X_{ij} > 0$ se $Z_{ij} = 1$ e $X_{ij} \leq 0$ se $Z_{ij} = 0$.

Existem algumas diferenças significativas em relação a nossa proposta e a de Beguin and Glas

(2001). Podemos notar que na abordagem de Beguin and Glas (2001) a variável auxiliar Z_{ij} depende da variável latente X_{ij} e dos parâmetros a_i, b_i e θ_j , pois $P(Z_{ij} = 1) = \Phi(a_i\theta_j - b_i)$. Já no nosso modelo, Z_{ij} depende somente do parâmetro c_i . Além disso, X e Z são amostrados separadamente no *Gibbs Sampling* de Beguin and Glas (2001) e conjuntamente no *Gibbs Sampling* proposto neste trabalho. Mesmo que a amostragem de X e Z fosse conjunta no *Gibbs Sampling* de Beguin and Glas (2001) (o que é possível), espera-se que nossa abordagem resulte em um *Gibbs Sampling* com menor correlação entre os blocos que, por sua vez, resulta em melhores propriedades de convergência, devido as diferentes especificações de Z_{ij} .

2.4 Estudo de simulação para o novo modelo aumentado de 3 parâmetros

Nesta seção será apresentado um estudo de simulação para avaliar a eficiência do novo modelo aumentado de 3 parâmetros da TRI proposta na Seção 2.1.

Neste estudo foram considerados 5000 indivíduos, respondendo cada um deles 50 itens. As proficiências dos indivíduos foram geradas a partir de uma distribuição Normal(0,1). A dificuldade, discriminação e acerto casual dos itens foram gerados a partir das ditribuições Uniforme(-3,3), Uniforme(0.5,2.5) e Uniforme(0.1,0.4), respectivamente.

Foi gerada uma amostra de tamanho 50000 para cada parâmetro com um *burn-in* de 30000, dessa forma, os parâmetros foram estimados a partir de uma amostra de tamanho 20000.

As distribuições *a priori* usadas foram as seguintes: $\theta_j \sim N(0, 1)$, $a_i \sim N_{(0,\infty)}(2, 3^2)$, $b_i \sim N(0, 3^2)$ e $c_i \sim Beta(4, 12)$, $\forall i = 1, \dots, 50$ e $\forall j = 1, \dots, 5000$.

A Figura 2.1 mostra os gráficos das cadeias dos parâmetros referentes aos indivíduos 2, 100, 500 e 1000, onde linha em cor cinza indica o valor real de cada parâmetro. Já a Figura 2.2 apresenta os gráficos das cadeias dos parâmetros referentes aos itens 2 e 10, onde linha em cor cinza indica o valor real de cada parâmetro. Note que existe forte evidência de que a cadeia convergiu.

Na Figura 2.3 é mostrado os gráficos dos valores reais *versus* os valores estimados (média *a posteriori*) de a , b , c e θ , respectivamente. Podemos observar que o modelo foi muito eficiente ao estimar os parâmetros de dificuldade e discriminação dos itens e de proficiência dos indivíduos. Note que a maior dificuldade é $b = 2.895$, ou seja, eu precisaria de itens mais difíceis para discriminar de forma mais eficiente os alunos com maior proficiência, por exemplo $\theta = 4.139$. Já o parâmetro

c é geralmente difícil de estimar, ver Baker and Kim (2004), mas mesmo assim os resultados são considerados bons, dado o tamanho do conjunto de dados.

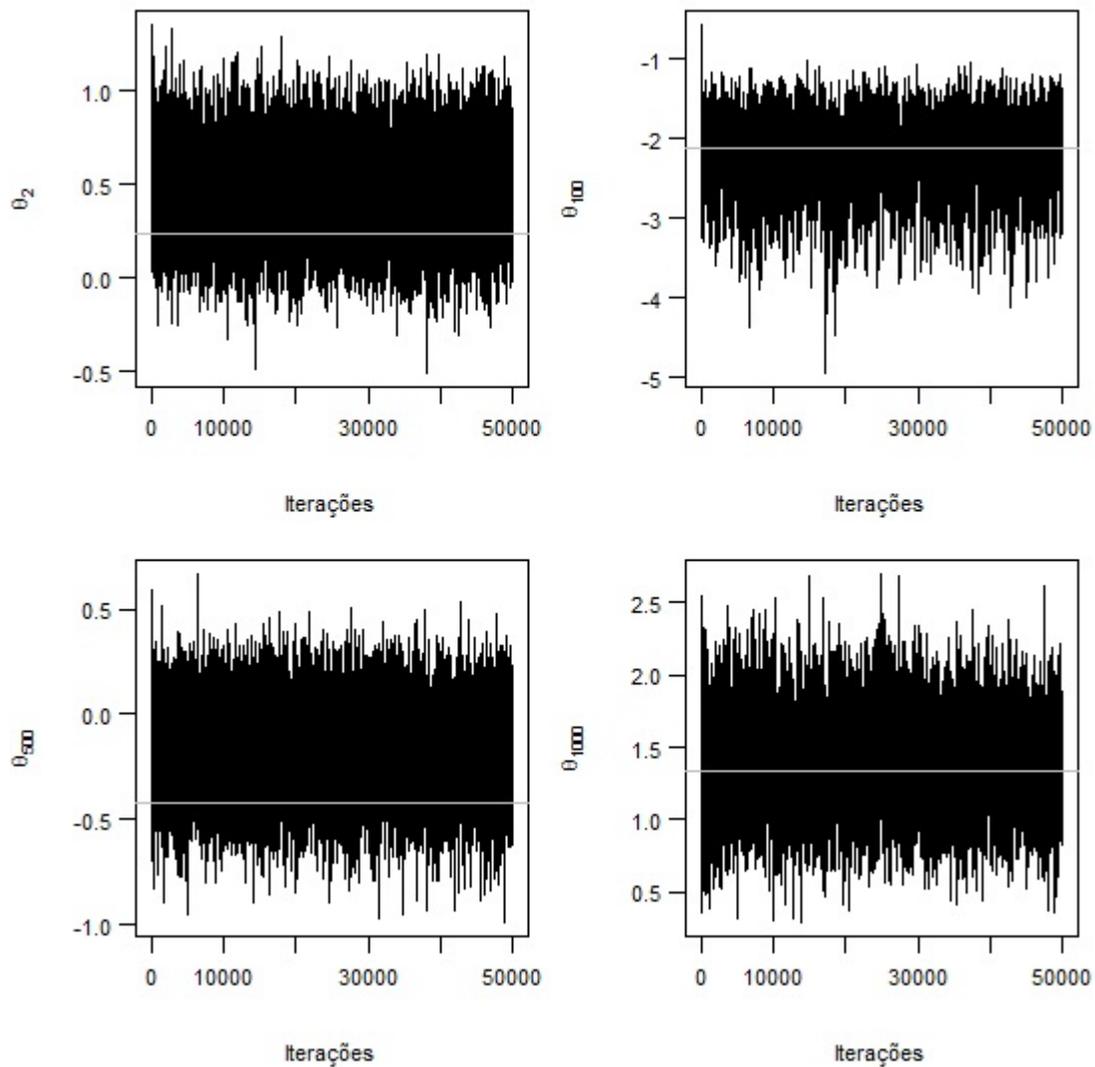


Figura 2.1: Trace Plots para cadeias de θ .

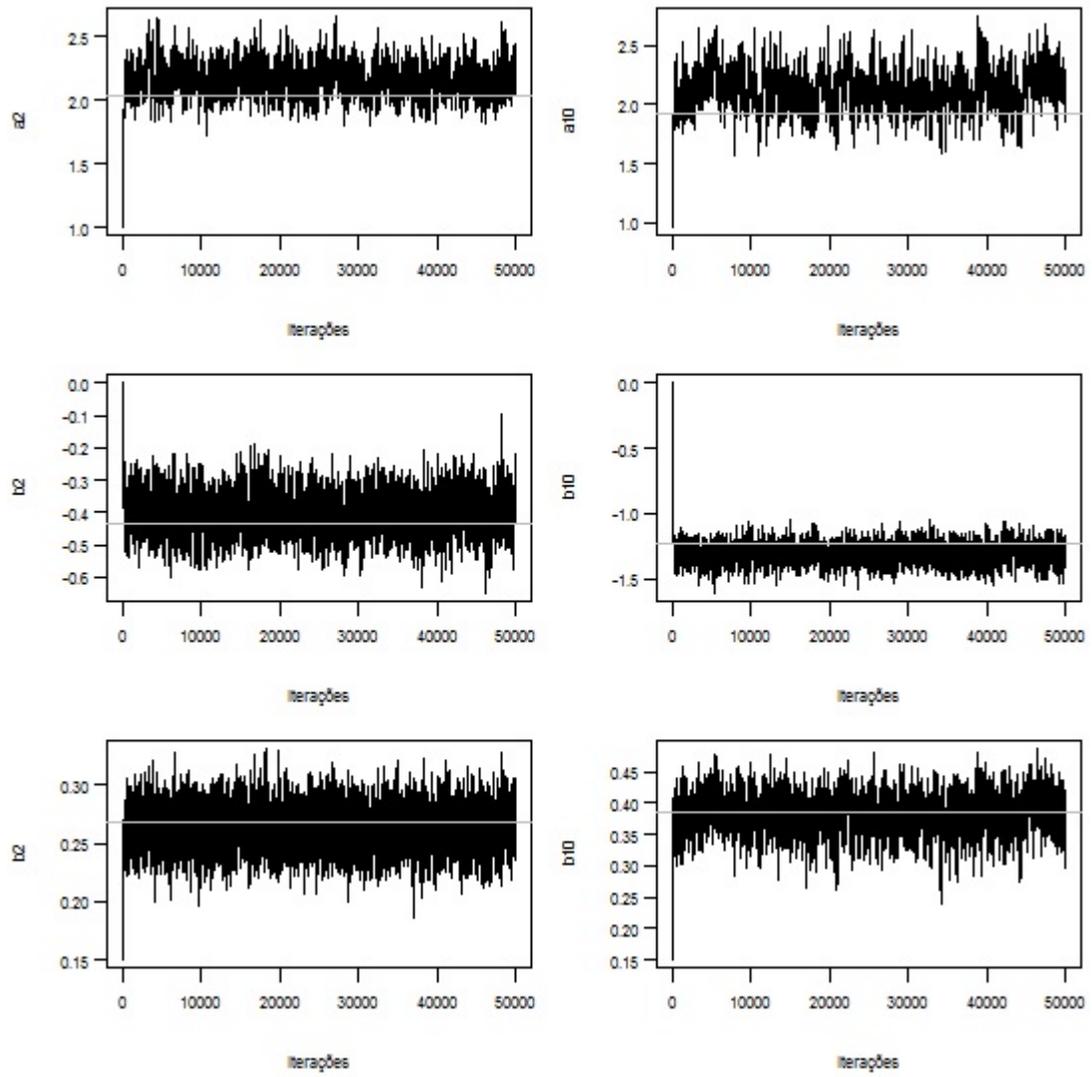


Figura 2.2: Trace Plots para cadeias de a , b , c .

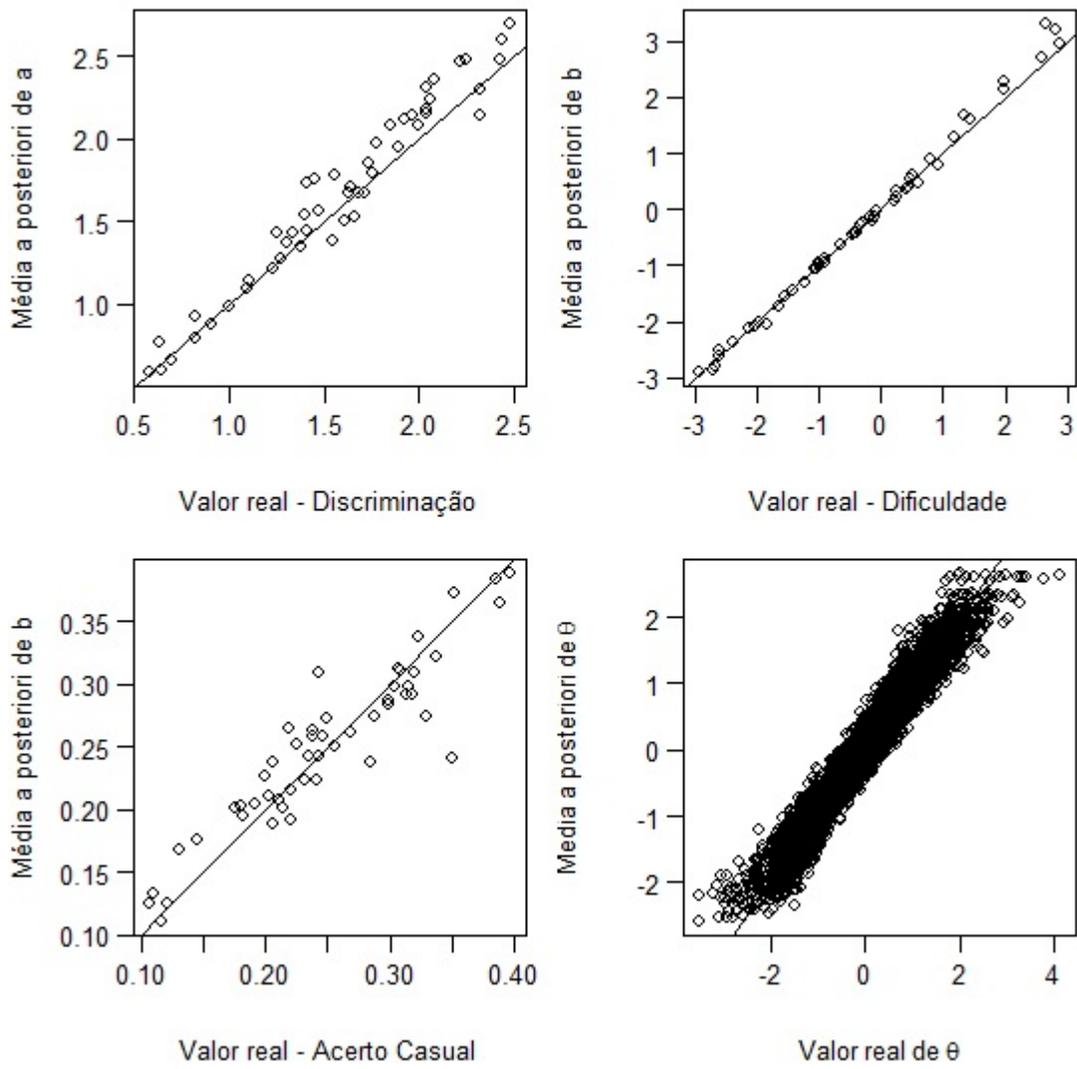


Figura 2.3: Valor real *versus* valor estimado de a , b , c e θ .

Capítulo 3

Modelo da TRI com mistura de distribuições normais para a proficiência

3.1 Mistura de densidades

Apesar da tradição no estudo de populações homogêneas em estatística, principalmente em distribuições normais, Pearson (1894) e Newcomb (1886) abriram a possibilidade de fornecer uma modelagem natural da heterogeneidade dos dados, com o primeiro fazendo decomposição de misturas de normais por meio do método de momentos e o segundo fazendo uma aplicação de misturas de normais com modelos para outliers.

Definição 3.1: Sejam π_k 's funções densidades de probabilidade e seja M uma variável aleatória contínua. Dizemos que M tem uma distribuição de mistura se sua densidade π pode ser expressa da seguinte maneira:

$$\pi(w) = \sum_{k=1}^K p_k \pi_k(m), \quad (3.1)$$

onde p_k 's são pesos, tal que $\sum_{k=1}^K p_k = 1$, com $p_k > 0$.

A seguir, podemos verificar que π é uma densidade de probabilidade:

$$\int_{\mathbb{R}} \pi(m) \, dm = \int_{\mathbb{R}} \sum_{k=1}^K p_k \pi_k(m) \, dm = \sum_{k=1}^K p_k \int_{\mathbb{R}} \pi_k(m) \, dm = \sum_{k=1}^K p_k = 1.$$

A esperança e variância de mistura de distribuições são dadas por:

$$E[M] = \sum_{k=1}^K p_k \mu_k$$

$$Var[M] = \sum_{k=1}^K p_k \left[(\mu_k - E[M])^2 + \sigma_k^2 \right]$$

onde μ_k e σ_k^2 são a média e a variância referentes as função densidade de probabilidade π_k , com $k = 1, \dots, K$.

3.2 Mistura de normais para a distribuição das proficiências

Como já foi citado anteriormente, no Capítulo 1, o objetivo deste trabalho é propor um novo modelo da TRI que flexibiliza a hipótese de normalidade dos traços latentes. Uma das motivações para isto é que uma distribuição normal não modela uma possível multimodalidade ou heterogeneidade dos dados. Desta forma, iremos utilizar uma mistura de distribuições normais para modelar as proficiências dos indivíduos. Em particular, esta abordagem permite uma solução para modelar distribuições com caudas pesadas ou assimetria sem o uso de distribuições assimétricas (ex: distribuição Skew-Normal).

Considere o modelo Probit de 3 parâmetros descrito em (2.1). No novo modelo proposto será utilizado a seguinte distribuição *a priori* para θ :

$$\pi(\theta_j | p_k, \mu_k, \sigma_k) = \sum_{k=1}^K p_k \left[\frac{1}{\sigma_k} \phi_k \left(\frac{\theta_j - \mu_k}{\sigma_k} \right) \right], \quad (3.2)$$

onde $\sum_{k=1}^K p_k = 1$, com $p_k > 0$, $\mu_k \in \mathbb{R}$ e $\sigma_k > 0$, $j = 1, \dots, J$, $k = 1, \dots, K$.

Na abordagem anterior (Seção 2.2), onde atribuíamos uma distribuição normal para a habilidade do indivíduo j , fixamos os parâmetros de locação e escala, tornando o modelo identificável e permitindo assim a estimação dos parâmetros dos itens. Nesta nova abordagem, a habilidade θ_j

será modelada por uma mistura e para evitar problema de identificabilidade também é necessário fazer restrições, que são discutidas na Seção 3.2.1.

Os pesos p_k 's não são conhecidos, para $k = 1, \dots, K$, ou seja, teremos que estimá-los e isto no contexto da TRI, significa estimar a qual proporção de alunos se refere cada componente da mistura. Isto é feito atribuindo uma distribuição *a priori* para p , onde $p = (p_1, \dots, p_K)$, no caso $p \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$, para $\alpha_1, \dots, \alpha_K > 0$ conhecidos. Além disto, a média μ_k e variância σ_k^2 da k -ésima componente da mistura também serão estimados, atribuindo a distribuição *a priori* $(\mu_k, \sigma_k^2) \sim \text{NIG}\left(m, \frac{\sigma_k^2}{\beta}, d, e\right)$, com d, e, m e β conhecidos, onde *NIG* representa a distribuição Normal-Gama Inversa. Desta forma, considerando $\mu = (\mu_1, \dots, \mu_K)$ e $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)$, temos a seguinte distribuição *a priori* conjunta para (p, μ, σ^2) :

$$\pi(p, \mu, \sigma^2) = \prod_{k=1}^K p_k^{\alpha_k - 1} \pi(\mu_k, \sigma_k^2) I_{(p \in A_p)}, \quad (3.3)$$

onde $\pi(\mu, \sigma^2)$ é a densidade de uma distribuição Normal-Gama Inversa e A_p é uma região do espaço paramétrico onde a distribuições a priori será truncada (ver Seção 3.2.1).

Com o objetivo de amostrar diretamente das distribuições condicionais completas é necessário acrescentar mais uma variável auxiliar no modelo, uma solução já existente e comum em modelos de mistura. Esta variável auxiliar é denotada por W_j com $j = 1, \dots, J$, onde $W_j \sim \text{Multinomial}(1, p_1, \dots, p_K)$, ou seja, $W_{jk} \in \{0, 1\}$ e $\sum_{j=1}^J W_{jk} = 1$, tal que $W_{jk} = 1$ se $\theta_j \in k$, com $k = 1, \dots, K$.

Desta forma, considerando a variável W_j , podemos reescrever a distribuição *a priori* para θ da seguinte forma:

$$\pi(\theta_j | W_j, \mu_k, \sigma_k) = \prod_{k=1}^K \left[\frac{1}{\sigma_k} \phi_k \left(\frac{\theta_j - \mu_k}{\sigma_k} \right) \right]^{W_{jk}}, \quad (3.4)$$

onde $\sum_{j=1}^J W_{jk} = 1$, com $W_{jk} \in \{0, 1\}$, $\mu_k \in \mathbb{R}$ e $\sigma_k > 0$, $j = 1, \dots, J, k = 1, \dots, K$.

Neste trabalho iremos tratar somente de mistura de duas distribuições normais, a justificativa para isto é que uma mistura com $K = 2$ é suficiente para modelar uma possível presença de assimetria, cauda pesada ou bimodalidade na distribuição das proficiências dos indivíduos. Além disso, seria mais complexo identificar um modelo com mistura de $K > 2$, dificultando assim a estimação dos parâmetros.

3.2.1 Identificabilidade do modelo

A abordagem generalizada adotada neste trabalho requer um cuidado especial com a questão de identificabilidade do modelo. Discutimos aqui questões práticas de identificabilidade e algumas estratégias para evitar problemas desta natureza.

Primeiramente, problemas de identificabilidade podem ocorrer em dois níveis. O primeiro se refere à identificação da distribuição das proficiências e o segundo à identificação das componentes da mistura que definem esta distribuição.

A maneira mais intuitiva de identificar este modelo seria fixar a média e a variância da mistura, definindo assim uma escala para as proficiências, porém é extremamente complexo trabalhar com esta restrição. Seria razoável pensar em uma solução para fixar a média da mistura, no entanto o problema estaria na fixação da variância. Por isso, no nosso modelo, foram atribuídos outros tipos de restrições.

A identificação do modelo é feita fixando os parâmetros de locação e escala da primeira componente da mistura, por exemplo, locação zero e escala 1. Com isto, estaremos fixando a distribuição para uma parcela da população, quanto maior for esta parcela, maior a capacidade de identificação dos parâmetros dos itens e conseqüentemente será possível estimar as habilidades das demais parcelas para as quais as distribuições não foram fixadas.

Neste sentido, adota-se uma segunda restrição onde o peso da primeira componente é restrito a ser maior que todos os demais pesos, ou seja, esta será a componente da mistura referente ao maior grupo de indivíduos, se tornando assim um grupo de referência. Em particular, a proporção mínima de alunos na primeira componente da mistura sempre será $1/K$, ou seja, esta restrição é mais eficiente quando o número de componentes da mistura (K) não é muito grande. Como estamos tratando de casos com $K = 2$, restringimos $p_1 > 0.6$, a justificativa para isso é que não se espera uma distribuição muito diferente da normal que modele bem a distribuição das proficiências.

Outro ponto importante é que as componentes da mistura não podem estar completamente segregadas, pois se isso ocorre temos um problema de identificabilidade na segunda componente da mistura, por motivo análogo ao modelo original da TRI. Dessa forma, sabemos que as componentes devem se sobrepor, ou seja, entre elas devem existir uma massa de probabilidade significativa em uma região comum. Isso justifica as médias das componentes não estarem muito longe uma das outras. Dessa forma, restringimos $|\mu_1 - \mu_2| < 3\sigma_1$.

Note que a densidade conjunta de (Y, θ) é dada por:

$$\left[\prod_{i=1}^I \prod_{j=1}^J (\Phi(a_i \theta_j - b_i))^{y_{ij}} (1 - \Phi(a_i \theta_j - b_i))^{1-y_{ij}} \right] \prod_{j=1}^J \left[p_1 \phi(\theta_j) + (1 - p_1) \frac{1}{\sigma_2} \phi\left(\frac{\theta_j - \mu_2}{\sigma_2}\right) \right] \quad (3.5)$$

Pelo mesmo motivo do modelo convencional (que atribui normalidade para proficiências) teríamos problemas de não identificabilidade do modelo se não fosse estabelecido nenhum tipo de restrição. Pode-se notar que, sem restrições a densidade de $(Y|\theta, a, b, c)$ não é identificável, uma vez que qualquer transformação do tipo $\theta_j^* = \theta_j + r$ e $b_i^* = b_i + a_i r$, $r \in R$, com $j = 1, \dots, J$ e $i = 1, \dots, I$ levam a uma mesma probabilidade de acerto, pois $\Phi(a_i \theta_j - b_i) = \Phi(a_i \theta_j^* - b_i^*)$.

Porém, fixando $\mu_1 = 0$ e $\sigma_1 = 1$ temos que $\phi(\theta_j^*) \neq \phi(\theta_j)$ (a menos que $\theta_j = u_r$, onde $\phi(u_r|\mu_2, \sigma_2, p) = \phi(u_r + r|\mu_2, \sigma_2, p)$) e conseqüentemente:

$$p_1 \phi(\theta_j^*) + (1 - p_1) \frac{1}{\sigma_2} \phi\left(\frac{\theta_j^* - \mu_2}{\sigma_2}\right) \neq p_1 \phi(\theta_j) + (1 - p_1) \frac{1}{\sigma_2} \phi\left(\frac{\theta_j - \mu_2}{\sigma_2}\right)$$

Dessa forma, o único problema de identificabilidade que ainda pode ocorrer no modelo seria o *label switch* para (μ_2, σ_2, p_1) , que será evitado quando fixamos $p_1 > 0.6$.

3.2.2 Inferência Bayesiana para o modelo com mistura

Defina $S = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \theta, \mu_2, \sigma_2, p, \mathbf{X}, \mathbf{Z}, \mathbf{W}\}$, a distribuição conjunta de Y com S é dada por:

$$\begin{aligned} \pi(Y, S) &= \left[\prod_{i=1}^I \prod_{j=1}^J \pi(Y_{ij}|X_{ij}, Z_{ij}) \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j) \pi(Z_{ij}|c_i) \right] \left[\prod_{i=1}^I \pi(a_i) \pi(b_i) \pi(c_i) \right] \\ &\times \left[\prod_{j=1}^J \pi(\theta_j|W_j, \mu_k, \sigma_k) \pi(W_j|p) \right] \pi(p) [\pi(\mu_2|\sigma_2^2) \pi(\sigma_2^2)]. \end{aligned} \quad (3.6)$$

Como estamos utilizando uma mistura de distribuições normais para modelar as proficiências dos indivíduos, será necessário obter uma nova distribuição condicional completa para θ_j e além disso devemos obter também as condicionais completas de μ_2, σ_2, p e W_j .

Para o *Gibbs Sampling* deste modelo é adotado o seguinte sistema de blocagem:

$$(\mathbf{X}, \mathbf{Z}) \quad (\theta, \mathbf{W}) \quad (\mathbf{a}, \mathbf{b}) \quad \mathbf{c} \quad (\mu_2, \sigma_2^2) \quad p.$$

Todas as condicionais completas têm densidades proporcionais à expressão (3.6), resultando nas

seguintes distribuições:

Lema 2: Além das citadas no Lema 1, as distribuições condicionais completas do modelo proibito com mistura de distribuições normais para θ são dadas por:

$$\mu_2, \sigma_2^2 | \cdot \sim NIG \left(m^*; \frac{\sigma_2^2}{\beta^*}; d^*; e^* \right), \quad (3.7)$$

$$\text{onde } m^* = \frac{\sum_{j=1}^J W_{j2} \theta_j + m\beta}{\sum_{j=1}^J W_{j2} + \beta}, \quad \beta^* = \sum_{j=1}^J W_{j2} + \beta, \quad d^* = d + \left(\sum_{j=1}^J W_{j2} \right) / 2,$$

$$e^* = e + \frac{\sum_{j=1}^J W_{j2} \beta}{2 \left(\beta + \sum_{j=1}^J W_{j2} \right)} (m - \theta_{bar})^2 + s/2, \quad \text{onde } s = \sum_{j=1}^J W_{j2} \theta_j^2 - \sum_{j=1}^J W_{j2} \text{ e } \theta_{bar} = \frac{\sum_{j=1}^J W_{j2} \theta_j}{\sum_{j=1}^J W_{j2}}.$$

$$p_1 | \cdot \sim Beta \left(\sum_{j=1}^J W_{j1} + \alpha_1, \sum_{j=1}^J W_{j2} + \alpha_2 \right) I_{(p_1 > 0.6)} \quad (3.8)$$

$$\pi(\theta_j, W_j | \cdot) \propto \left[\frac{1}{\sigma_k^*} \phi \left(\frac{\theta_j - \mu_k^*}{\sigma_k^*} \right) \right]^{W_{jk}} \pi_{Mult}(W_j; 1, p_1^*, p_2^*), \quad k = 1, 2. \quad (3.9)$$

$$\text{onde } k \text{ é a componente que } \theta_j \text{ pertence, } p_k^* = \frac{\alpha_k^*}{K}, \quad \mu_k^* = \frac{\mu_k + \sigma_k^2 \sum_{i=1}^{L_j} a_i (x_{ij} - b_i)}{\sum_{k=1}^K \alpha_k^*},$$

$$\sigma_k^{2*} = \frac{\sigma_k^2}{L_j} \text{ e } \frac{1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2}{1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2}$$

$$\alpha_k^* = p_k \left(1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\mu_k^2}{\sigma_k^2} - \frac{\left(\mu_k + \sigma_k^2 \sum_{i=1}^{L_j} a_i (x_{ij} - b_i) \right)^2}{\sigma_k^2 \left(1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2 \right)} \right] \right\}.$$

Para entendermos a distribuição (3.9) usamos a seguinte fatorização:

$$\pi(\theta_j, W_j | \cdot) = \pi(\theta_j | W_j, \cdot) \pi(W_j | \cdot).$$

Quando $W_{jk} = 1$, temos que $\pi(\theta_j | W_{jk} = 1, \cdot)$ é uma densidade da distribuição normal com média μ_k^* e variância σ_k^{2*} , com probabilidade p_k , para $k = 1, 2$.

Para amostrar da distribuição (3.8) será utilizado o método da CDF inversa.

Prova Lema 2: Ver apêndice B.

Capítulo 4

Estudo de simulação para o modelo da TRI com mistura

Neste capítulo serão apresentados estudos de simulações para avaliar a eficiência do modelo da TRI com mistura de distribuições normais para a proficiência. Serão feitos 3 estudos, o primeiro utiliza uma mistura de normais que tornam a distribuição assimétrica, o segundo utiliza uma mistura de normais bimodal e por ultimo é utilizado uma mistura de normais com caudas pesadas. Todas elas são usadas para modelar a distribuição do traço latente dos indivíduos que respondem aos itens. No final desta seção será feita uma comparação entre o modelo da TRI com mistura e o modelo da TRI tradicional.

4.1 Estudo de simulação 1

Nesta seção será apresentado um estudo de simulação para avaliar a eficiência do modelo da TRI com mistura bimodal de duas distribuições normais para a proficiência. Foram considerados 5000 indivíduos, respondendo cada um deles 50 itens. As proficiências dos indivíduos foram geradas a partir da mistura $0.8N(0, 1) + 0.2N(2.5, 0.5^2)$.

A dificuldade, discriminação e acerto casual dos itens foram gerados a partir das distribuições Uniforme(-4,4.5), Uniforme(0.5,2.5) e Uniforme(0.1,0.4), respectivamente.

Foi gerada uma amostra de tamanho 200000 para cada parâmetro com um *burn-in* de 100000, dessa forma, os parâmetros foram estimados a partir de uma amostra de tamanho 100000. As

distribuições *a priori* usadas foram as seguintes: $\theta_j \sim p_1 N(0, 1) + p_2 N(\mu_2, \sigma_2^2)$, $p = (p_1, p_2) \sim Dirichlet(2, 1)_{I(p_1 > p_2)}$, $(\mu_2, \sigma_2^2) \sim NIG(0, 100\sigma_k^2, 0.001, 0.001)$, $a_i \sim N_{(0, \infty)}(2, 3^2)$, $b_i \sim N(0, 2^2)$ e $c_i \sim Beta(4, 12)$, $\forall i = 1, \dots, 50$ e $\forall j = 1, \dots, 5000$.

Os gráficos das cadeias dos parâmetros da mistura, dos Itens 2 e 4 e dos indivíduos 200 e 300 se encontram no Apêndice C, onde é possível ter evidência de que a cadeia convergiu.

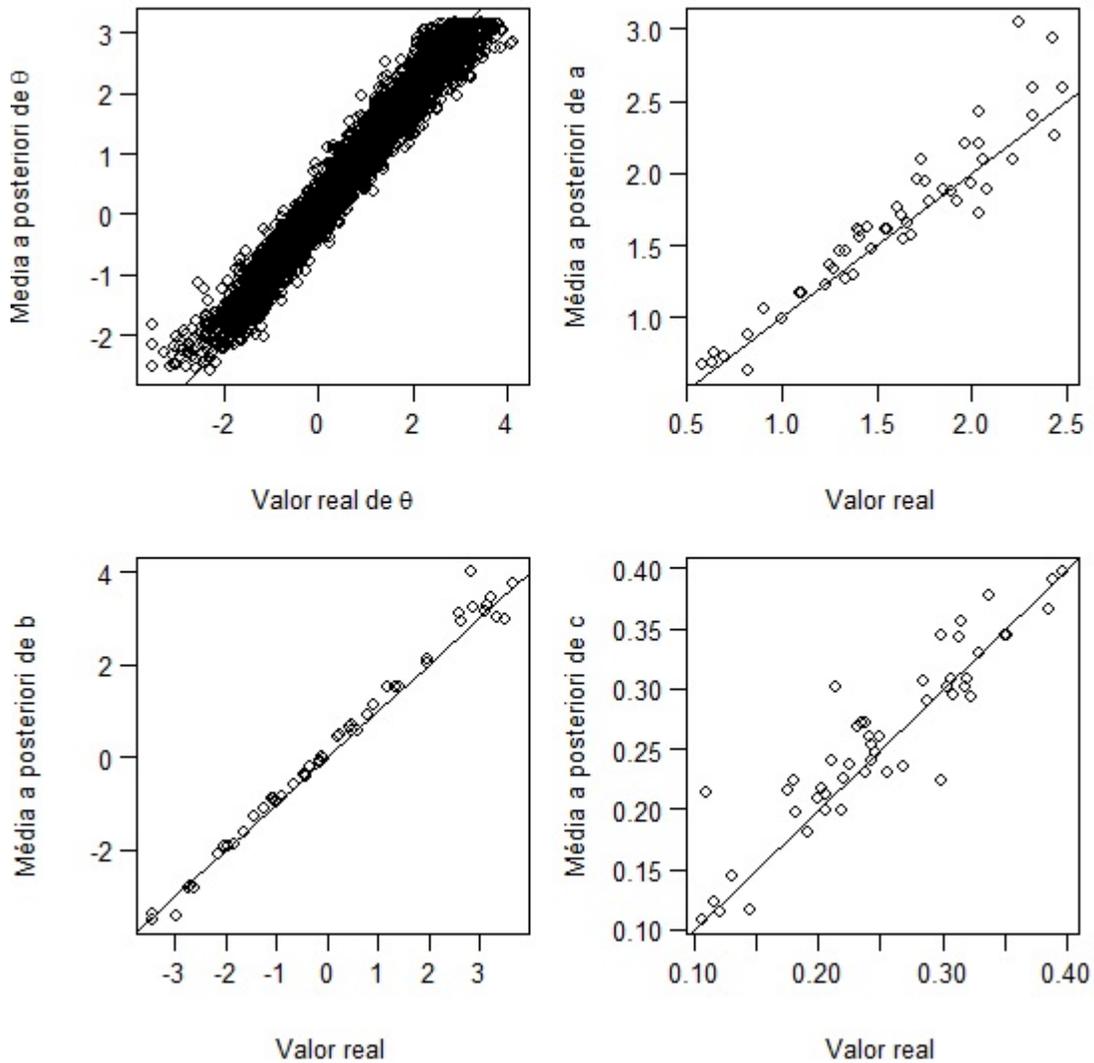


Figura 4.1: Valor real *versus* valor estimado de a , b , c e θ .

Na Figura 4.1 é mostrado os gráficos dos valores reais *versus* os valores estimados (média *a posteriori*) de θ , a , b e c , respectivamente. Podemos observar que estes parâmetros são satisfatoriamente

	Valor Real	Valor Estimado
μ_2	2.5	2.35
σ_2^2	0.25	0.36
p_1	0.80	0.78
p_2	0.20	0.22

Tabela 4.1: Média *a posteriori* para os parâmetros da mistura 1

estimados.

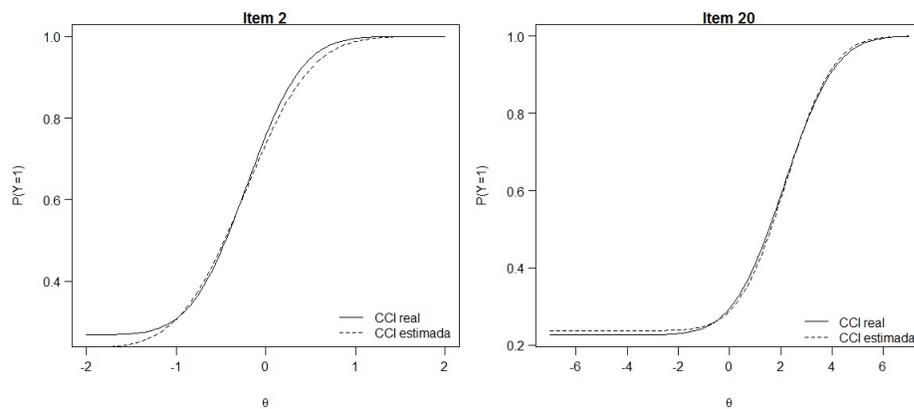


Figura 4.2: CCI real e CCI estimada para os itens 2 e 40.

As Curvas Características do Item estimada e real referentes aos itens 2 e 40 são apresentadas na Figura 4.2. Pode-se notar que a CCI estimada se aproxima muito da CCI real.

Através da Figura 4.3 pode-se perceber que a densidade real da proficiência é muito bem estimada. Note que na mesma figura é feito um histograma dos valores estimados para as proficiência, que também se ajusta muito bem a densidade real. As médias *a posteriori* para μ_k, σ_2^2, p_1 e p_2 são apresentados na Tabela 4.1.

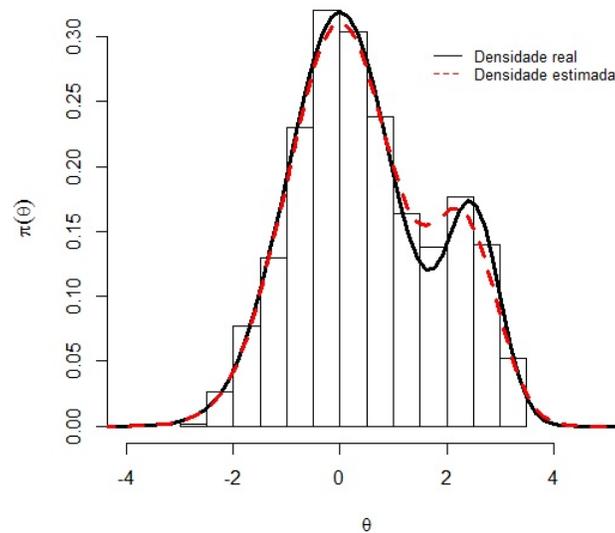


Figura 4.3: Densidade real, densidade estimada e histograma das médias *a posteriori* de θ .

4.2 Estudo de simulação 2

Nesta seção será apresentado um estudo de simulação para avaliar a eficiência do modelo da TRI com mistura assimétrica de duas distribuições normais para a proficiência. Foram considerados 5000 indivíduos, respondendo cada um deles 50 itens. As proficiências dos indivíduos foram geradas a partir da mistura $0.7N(0, 1) + 0.3N(1.5, 1.8^2)$.

A dificuldade, discriminação e acerto casual dos itens foram gerados a partir das distribuições Uniforme(-4,4.5), Uniforme(0.3,2.5) e Uniforme(0.1,0.4), respectivamente.

Foi gerada uma amostra de tamanho 200000 para cada parâmetro com um *burn-in* de 100000, dessa forma, os parâmetros foram estimados a partir de uma amostra de tamanho 100000.

As distribuições *a priori* usadas foram as seguintes: $\theta_j \sim p_1N(0, 1) + p_2N(\mu_2, \sigma_2^2)$, $(\mu_2, \sigma_2^2) \sim NIG(0, 100\sigma_k^2, 0.001, 0.001)$, $p = (p_1, p_2) \sim Dirichilet(2, 1)_{I(p_1 > p_2)}$, $a_i \sim N_{(0, \infty)}(1, 3^2)$, $b_i \sim N(0, 4^2)$ e $c_i \sim Beta(4, 12)$, $\forall i = 1, \dots, 50$ e $\forall j = 1, \dots, 5000$.

Os gráficos das cadeias dos parâmetros da mistura, dos Itens 2 e 4 e dos indivíduos 200 e 300 se encontram no Apêndice C, onde é possível ter evidência de que a cadeia convergiu.

Na Figura 4.4 é mostrado os gráficos dos valores reais *versus* os valores estimados (média *a poste-*

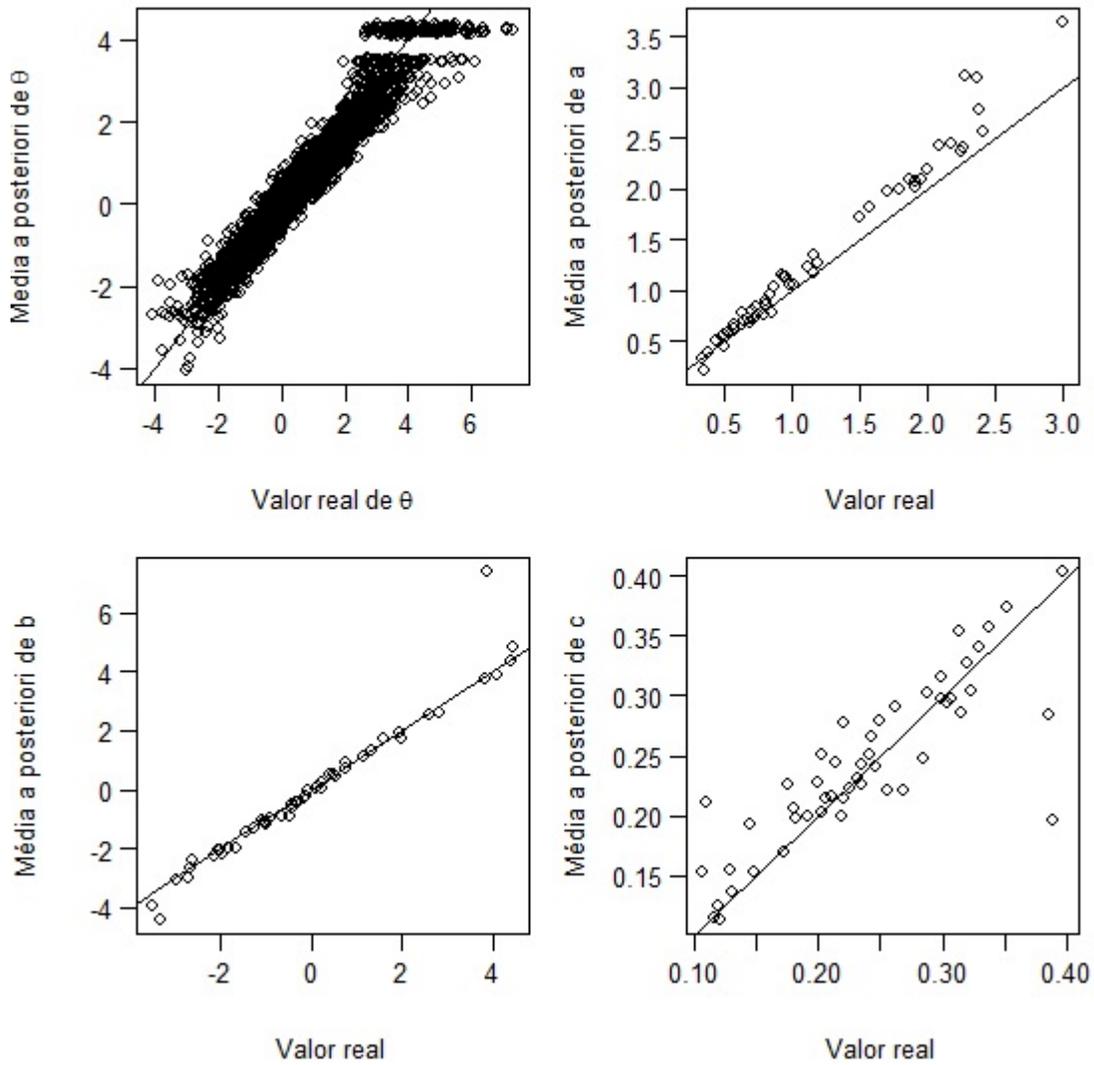


Figura 4.4: Valor real *versus* valor estimado de a , b , c e θ .

riori) de θ , a , b e c , respectivamente. Podemos observar que estes parâmetros são satisfatoriamente estimados.

As Curvas Características do Item estimada e real referentes aos itens 2 e 20 são apresentadas na Figura 4.5. Pode-se notar que a CCI estimada se aproxima muito da CCI real. Através da Figura 4.6 pode-se perceber que a densidade real é muito bem estimada. Note que na mesma figura é feito um histograma dos valores estimados para as proficiências, que também se ajusta muito bem a densidade real. As médias *a posteriori* para μ_k , σ_2^2 , p_1 e p_2 são apresentadas na Tabela 4.2.

	Valor Real	Valor Estimado
μ_2	1.5	1.44
σ_2^2	3.24	4.01
p_1	0.70	0.78
p_2	0.30	0.22

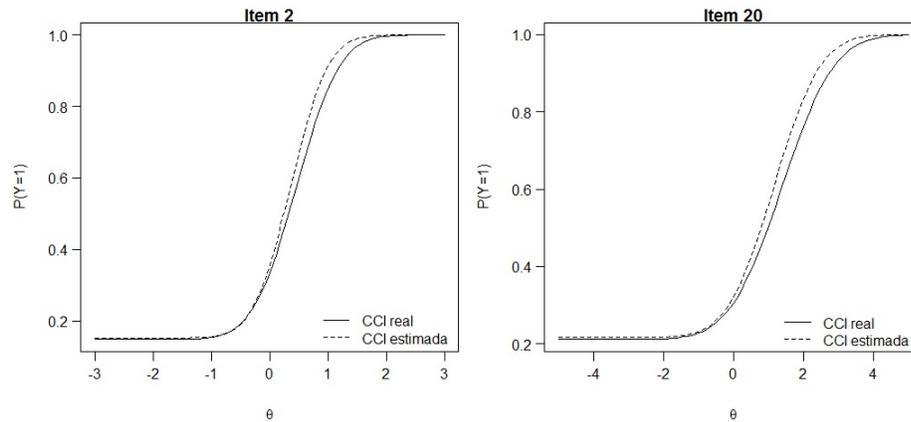
Tabela 4.2: Média *a posteriori* para os parâmetros da mistura 2

Figura 4.5: CCI real e CCI estimada para os itens 2 e 20.

4.3 Estudo de simulação 3

Nesta seção será apresentado um estudo de simulação para avaliar a eficiência do modelo da TRI com mistura de duas distribuições normais com caudas pesadas para modelar a proficiência dos indivíduos. Foram considerados 5000 indivíduos, respondendo cada um deles 50 itens. As proficiências dos indivíduos foram geradas a partir da mistura $0.7N(0, 1) + 0.3N(0.5, 12^2)$.

A dificuldade, discriminação e acerto casual dos itens foram gerados a partir das distribuições Uniforme(-4,4.5), Uniforme(0.3,3) e Uniforme(0.1,0.4), respectivamente. Foi gerada uma amostra de tamanho 200000 para cada parâmetro com um *burn-in* de 100000, dessa forma, os parâmetros foram estimados a partir de uma amostra de tamanho 100000.

As distribuições *a priori* usadas foram as seguintes: $\theta_j \sim p_1N(0, 1) + p_2N(\mu_2, \sigma_2^2)$, $(\mu_2, \sigma_2^2) \sim NIG(0, 100\sigma_k^2, 0.001, 0.001)$, $p = (p_1, p_2) \sim Dirichlet(2, 1)_{I(p_1 > p_2)}$, $a_i \sim N_{(0, \infty)}(1, 3^2)$, $b_i \sim N(0, 5^2)$ e $c_i \sim Beta(4, 12)$, $\forall i = 1, \dots, 50$ e $\forall j = 1, \dots, 5000$.

Os gráficos das cadeias dos parâmetros da mistura, dos Itens 2 e 4 e dos indivíduos 200 e 300

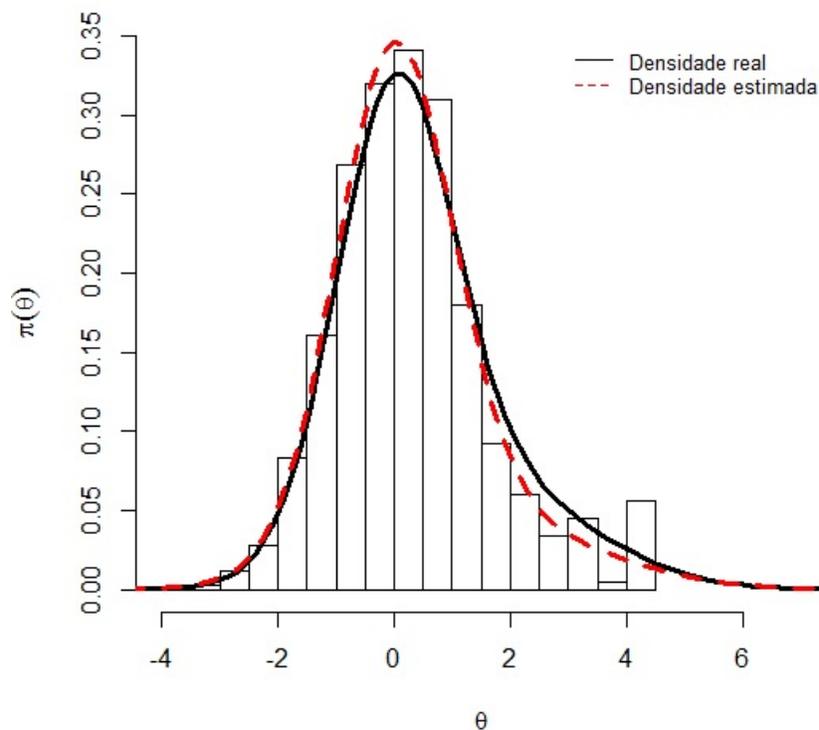


Figura 4.6: Densidade real, densidade estimada e histograma das médias *a posteriori* de θ .

se encontram no Apêndice C, onde é possível ter evidência de que a cadeia convergiu.

Na Figura 4.7 é mostrado os gráficos dos valores reais *versus* os valores estimados (média *a posteriori*) de θ , a e b , respectivamente. Podemos observar que estes parâmetros são satisfatoriamente estimados.

A Figura 4.8 mostra as Curvas Características do Item estimada e real. Pode-se notar que a CCI estimada se aproxima muito da CCI real.

Através da Figura 4.9 pode-se perceber que a densidade real da proficiência é muito bem estimada. Note que na mesma Figura 4.9 é feito um histograma dos valores estimados para as proficiência, que também se ajusta muito bem a densidade real. As médias *a posteriori* para μ_k, σ_2^2, p_1 e p_2 são apresentadas na Tabela 4.3.

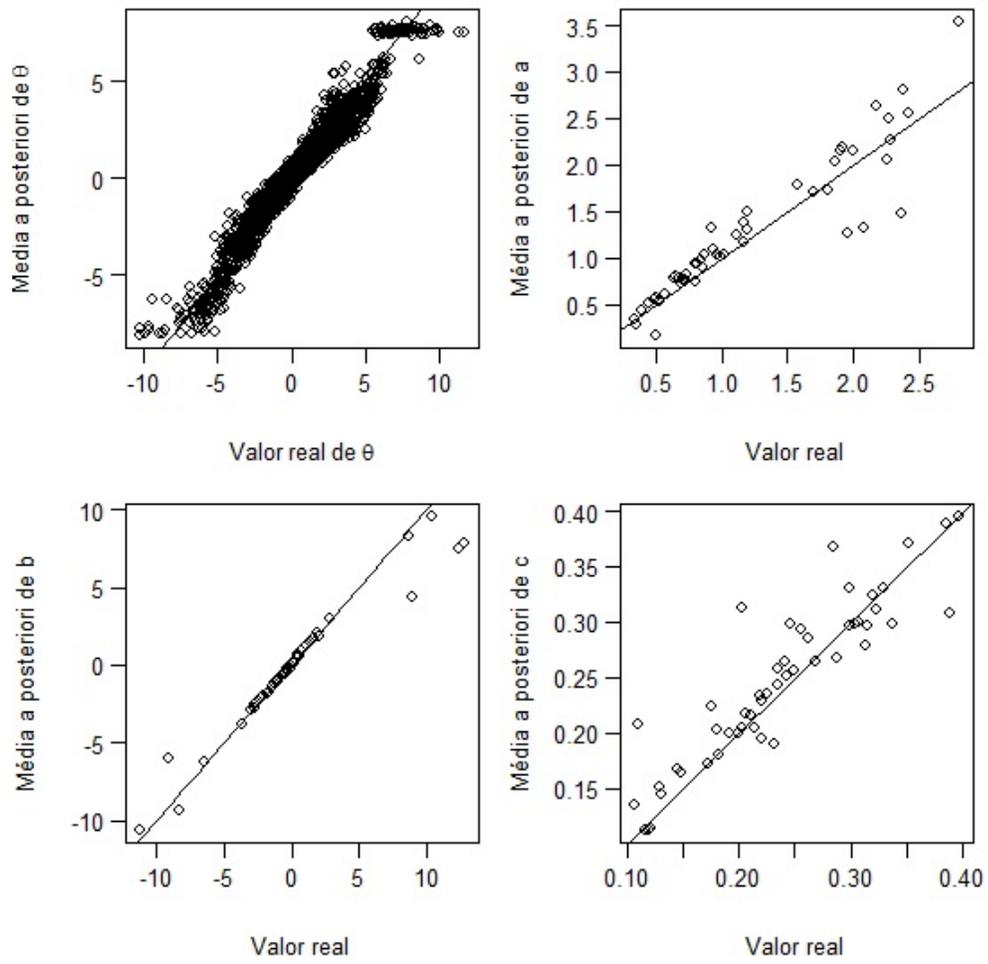


Figura 4.7: Valor real *versus* valor estimado de a , b , c e θ .

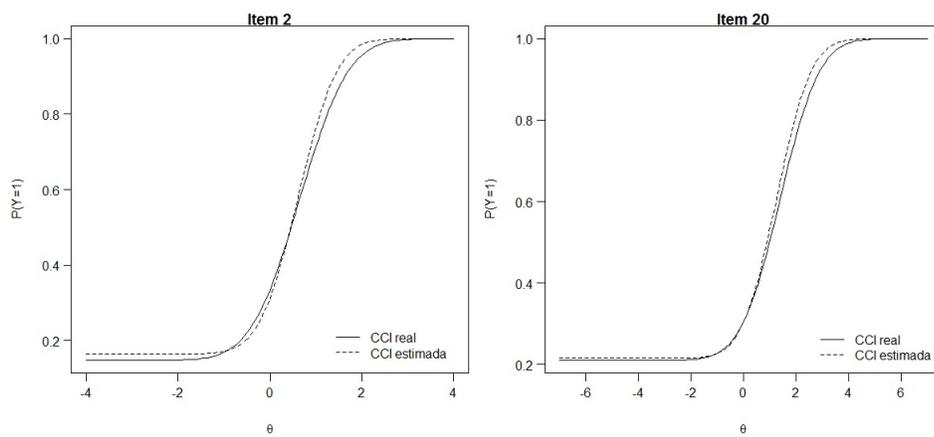


Figura 4.8: CCI real e CCI estimada para os itens 2 e 20.

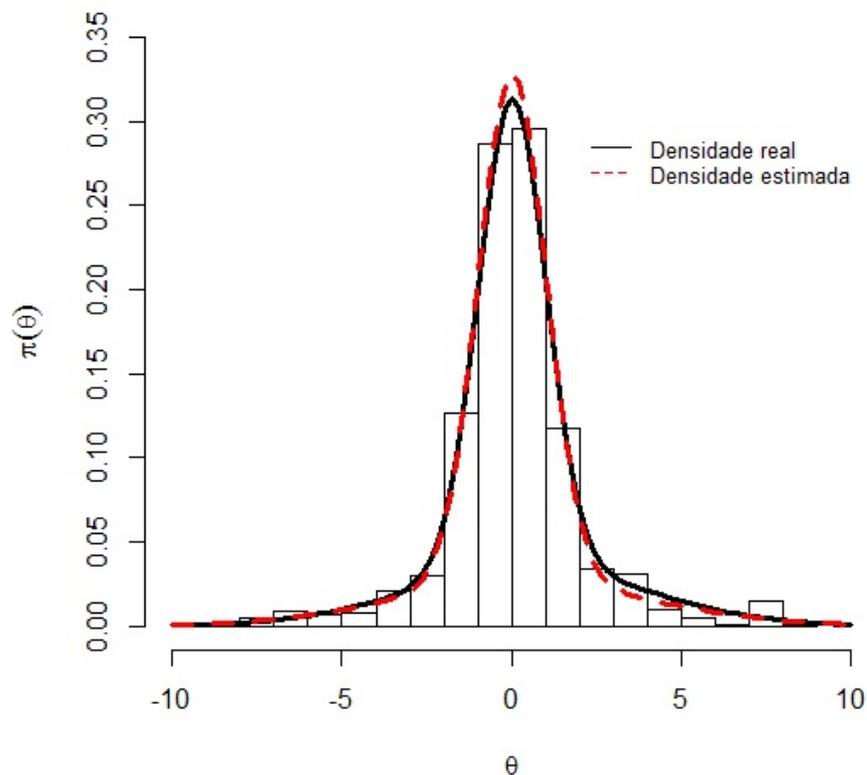


Figura 4.9: Densidade real, densidade estimada e histograma das médias *a posteriori* de θ .

4.4 Comparação entre os modelos

Nesta seção será feita uma comparação entre o modelo da TRI com mistura e o modelo da TRI tradicional.

Primeiro, será utilizado o modelo Normal (apresentado na Seção 2.1) para ajustar os dados simulados nas seções anteriores (Seções 4.1, 4.2 e 4.3). O objetivo desta análise será mostrar a importância do novo modelo proposto em casos que a distribuição das proficiências não apresentem um comportamento normal.

Em seguida será utilizado o modelo da TRI com mistura para ajustar os dados simulados na Seção 2.4, onde as proficiências foram geradas a partir de uma distribuição $N(0,1)$. O objetivo desta análise será verificar se o modelo com mistura para as proficiências recupera bem proficiências

	Valor Real	Valor Estimado
μ_2	0.5	0.29
σ_2^2	12	14.43
p_1	0.70	0.76
p_2	0.30	0.24

Tabela 4.3: Média *a posteriori* para os parâmetros da mistura 3

simétricas.

- **Ajuste dos dados referentes ao estudo de Simulação 1 utilizando o modelo normal da TRI**

Neste estudo foi utilizado o modelo Normal da TRI para modelar os dados considerados no estudo de Simulação 1 (Seção 4.1), onde as proficiências dos indivíduos foram geradas a partir da mistura $0.8N(0, 1) + 0.2N(2.5, 0.5^2)$. Fixamos os parâmetros da distribuição de θ igual a média e variância da mistura gerada, baseado na definição 3.1 fixamos $\theta \sim N(0.5, 1.85)$. Isto é feito para que os dois modelos sejam comparados em uma mesma escala.

A Figura 4.10 mostra o gráfico dos valores reais *versus* os valores estimados (média *a posteriori*) de θ e o histograma dos θ estimados.

O coeficiente de correlação entre as proficiências reais e as estimadas foi igual a 0.97, já no modelo de mistura este valor foi igual a 0.98.

Este estudo indica que, para este cenário, os dois modelos (Mistura e Tradicional) tiveram comportamento bem semelhantes em relação à proficiência dos indivíduos, dado que os coeficientes de correlação linear são praticamente iguais. Porém, através da Figura 4.10 é possível perceber que existe uma grande diferença entre a densidade real e a estimada, dessa forma é possível verificar que o histograma dos valores estimados para as proficiência não se ajusta bem a densidade real.

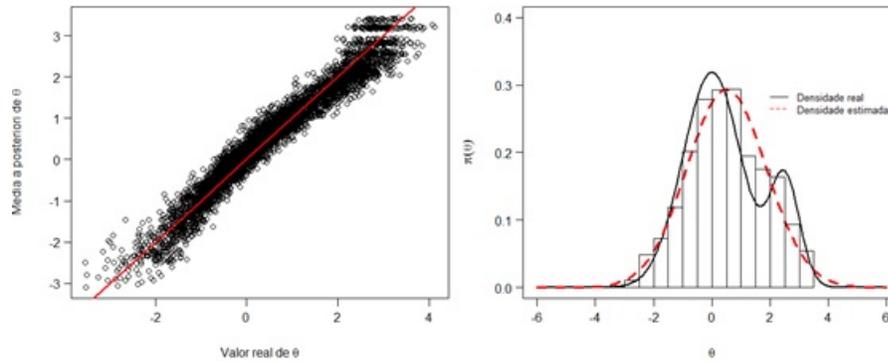


Figura 4.10: Mistura de normais bimodal ajustado pelo modelo simétrico.

- **Ajuste dos dados referentes ao estudo de Simulação 2 utilizando o modelo normal da TRI**

Neste estudo foi utilizado o modelo Normal da TRI para modelar os dados considerados no estudo de Simulação 2 (Seção 4.2), onde as proficiências dos indivíduos foram geradas a partir da mistura $0.7N(0, 1) + 0.3N(1.5, 1.8^2)$. Fixamos os parâmetros da distribuição de θ igual a média e variância da mistura gerada, baseado na definição 3.1 fixamos $\theta \sim N(0.45, 2.14)$. Isto é feito para que os dois modelos sejam comparados em uma mesma escala.

A Figura 4.11 mostra o gráfico dos valores reais *versus* os valores estimados (média *a posteriori*) de θ e o histograma dos θ estimados.

O coeficiente de correlação entre as proficiências reais e as estimadas foi igual a 0.96, já no modelo de mistura este valor foi igual a 0.97.

Este estudo indica que, para este cenário, os dois modelos (Mistura e Tradicional) tiveram comportamento bem semelhantes em relação à proficiência dos indivíduos, dado que os coeficientes de correlação linear são praticamente iguais. Porém o modelo sub-estima alguns valores, não sendo capaz de estimar bem as proficiências de maior valor (cauda direita). Além disso, existe uma grande diferença entre a densidade real e a estimada, por isso é possível verificar que o histograma dos valores estimados para as proficiência não se ajusta bem a densidade real.

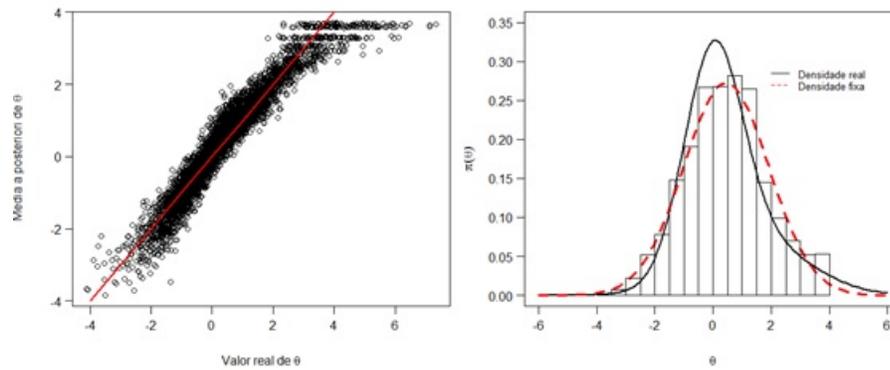


Figura 4.11: Mistura de normais assimétrico ajustado pelo modelo simétrico.

- **Ajuste dos dados referentes ao estudo de Simulação 3 utilizando o modelo normal da TRI**

Neste estudo foi utilizado o modelo Normal da TRI para modelar os dados considerados no estudo de Simulação 3 (Seção 4.3), onde as proficiências dos indivíduos foram geradas a partir da mistura $0.7N(0, 1) + 0.3N(0.5, 12)$. Fixamos os parâmetros da distribuição de θ igual a média e variância da mistura gerada, baseado na definição 3.1 fixamos $\theta \sim N(0.15, 4.35)$. Isto é feito para que os dois modelos sejam comparados em uma mesma escala.

A Figura 4.12 mostra o gráfico dos valores reais *versus* os valores estimados (*média a posteriori*) de θ e o histograma dos θ estimados.

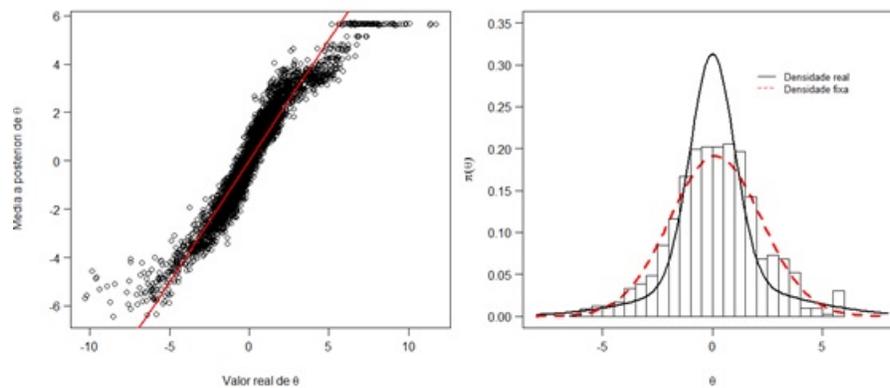


Figura 4.12: Mistura de normais com caudas pesadas ajustado pelo modelo simétrico.

O coeficiente de correlação entre as proficiências reais e as estimadas foi igual a 0.94, já no modelo de mistura este valor foi igual a 0.97.

Este estudo de simulação indica que o modelo normal não recupera muito bem os valores reais. Note na Figura 4.12 que o modelo sub-estima as maiores proficiências e super-estima as menores, ou seja, o modelo não foi capaz de estimar bem as proficiências localizadas nas caudas. Por isso, é possível verificar que o histograma dos valores estimados para as proficiência não se ajusta bem a densidade real.

- **Ajuste dos dados simétricos utilizando o modelo de mistura da TRI**

Neste estudo foi utilizado o modelo da TRI com mistura de distribuições normais para modelar os dados considerados no estudo de Simulação do capítulo 2 (Seção 2.4), onde as proficiências dos indivíduos foram geradas a partir da distribuição $N(0, 1)$.

A Figura 4.13 mostra o histograma dos θ estimados. Pode-se perceber que a densidade real da proficiência é muito bem estimada. Este estudo indica que o modelo o modelo da TRI com mistura

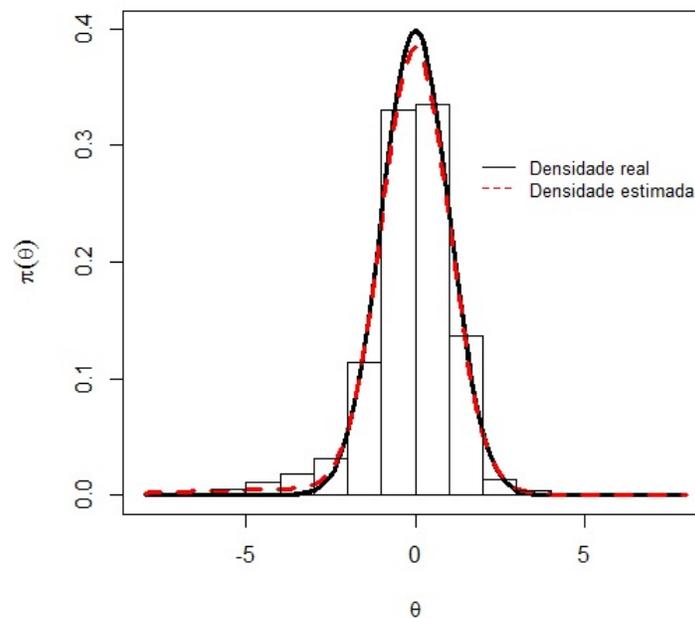


Figura 4.13: Distribuição normal ajustado pelo modelo com mistura.

de distribuições normais recupera bem os valores reais. As médias *a posteriori* para μ_k, σ_2^2, p_1 e p_2 são apresentados na Tabela 4.4. Como podemos ver, p_2 foi estimado bem próximo de zero (0.04),

	Valor Estimado
μ_2	-3.13
σ_2^2	12.93
p_1	0.96
p_2	0.04

Tabela 4.4: Média *a posteriori* para os parâmetros do modelo simples.

indicando que as proficiências seguem uma distribuição Normal, evidenciando assim que os dados são ajustados pelo modelo simétrico.

Capítulo 5

Análise dos dados do ENEM

Neste capítulo o modelo da TRI com mistura de distribuições normais para as proficiências dos indivíduos (proposto no capítulo 3) será aplicado ao conjunto de dados do Exame Nacional do Ensino Médio (ENEM) do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) do Ministério da Educação (MEC). O ENEM é uma prova elaborada pelo Ministério da Educação para verificar o domínio de competências e habilidades dos estudantes que estão concluindo ou que já concluíram o ensino médio em anos anteriores, onde na maioria dos casos os resultados são usados para o ingresso em diversas universidades no país. Para evitar fraude, a prova é realizada em 4 versões identificadas por cores (amarela, branca, rosa e azul). O que difere uma prova da outra é a ordem das questões e alternativas. O exame é composto por quatro provas. A prova de Ciências Humanas e suas Tecnologias traz questões sobre as disciplinas de história, geografia, filosofia e sociologia. A prova de Ciências da Natureza e suas Tecnologias cobra conhecimentos de química, física e biologia. Linguagens, Códigos e suas Tecnologias envolve questões de língua portuguesa, literatura, língua Estrangeira (Inglês ou Espanhol), artes, educação física e tecnologias da informação e comunicação. Já a prova de Matemática e suas Tecnologias tem questões de matemática (Geometria e Álgebra).

Os dados analisados nesta dissertação se referem ao teste azul de Matemática e suas Tecnologias, aplicado ao estado de São Paulo no ano de 2010. Foram analisadas provas de 52210 alunos, respondendo cada um deles 45 itens de matemática.

Uma análise dos escores brutos padronizados (Figura 5.1) indica uma provável assimetria positiva das proficiências. O escore bruto padronizado é calculado como segue.

$$\theta_j^{sbp} = \frac{s_j - E(s)}{Var(s)},$$

onde $s_j = \frac{\text{Número de itens corretos}}{\text{Número de itens respondidos}}$, $E(s) = \frac{\sum_{j=1}^J s_j}{J}$ e $Var(s) = \frac{\sum_{j=1}^J (s_j - E(s))^2}{J-1}$.

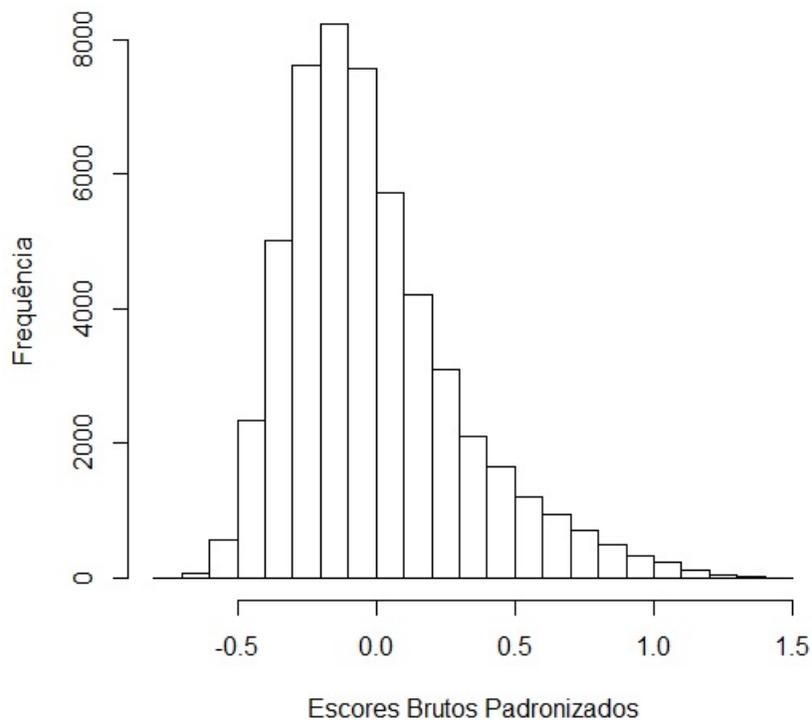


Figura 5.1: Histograma dos Escores Brutos Padronizados de cada indivíduo.

Para a execução do MCMC, foram adotados os seguintes valores iniciais para os parâmetros: $a_i^{(0)} = 1$, $b_i^{(0)} = 0$, $c_i^{(0)} = 0.1$, $\mu_2^{(0)} = 2$, $\sigma_2^{2(0)} = 1$ e $p_1^{(0)} = 0.8$, $\forall i = 1, \dots, 50$. Os valores iniciais para as proficiências foram os escores brutos padronizados de cada indivíduo j , $\forall j = 1, \dots, 52210$.

Foi gerada uma amostra de tamanho 70000 para cada parâmetro, porém nas 10000 primeiras iterações cada parâmetro c_i , $\forall i = 1, \dots, 45$ foi fixado em 0.10 e após estas 10000 iterações eles passaram a ser estimados juntamente com os outros parâmetros. A justificativa para isto é que o parâmetro

c é geralmente difícil de estimar, ver Baker and Kim (2004), afetando assim a estimação dos outros parâmetros. Com isso, as 10000 primeiras iterações foram descartadas da análise juntamente com um burn-in de 20000, dessa forma, os parâmetros foram estimados a partir de uma amostra de tamanho 40000.

Como se espera que a maioria dos itens tenham um parâmetro de acerto casual com um baixo valor, truncamos a distribuição a priori de c ao intervalo $(0.05, 0.15)$.

As distribuições *a priori* usadas foram as seguintes: $\theta_j \sim p_1 N(0, 1) + p_2 N(\mu_2, \sigma_2^2)$, $(\mu_2, \sigma_2^2) \sim NIG(0, 100\sigma_k^2, 1, 0)$, $p = (p_1, p_2) \sim Dirichlet(2, 1)_{I(p_1 > p_2)}$, $a_i \sim N_{(0, \infty)}(1, 3^2)$, $b_i \sim N(0, 3^2)$ e $c_i \sim Beta(1, 9)$, $\forall i = 1, \dots, 45$ e $\forall j = 1, \dots, 52210$.

As médias *a posteriori* para μ_2 , σ_2^2 , p_1 e p_2 são apresentados na Tabela 5.1.

O histograma das médias *a posteriori* das proficiências dos alunos e sua densidade estimada estão na Figura 5.6, sendo possível detectar uma assimetria positiva presente na distribuição das proficiências.

A Figura 5.4 mostra as estimativas intervalares (intervalo de credibilidade de 95%) dos parâmetros dos itens a , b e c , respectivamente.

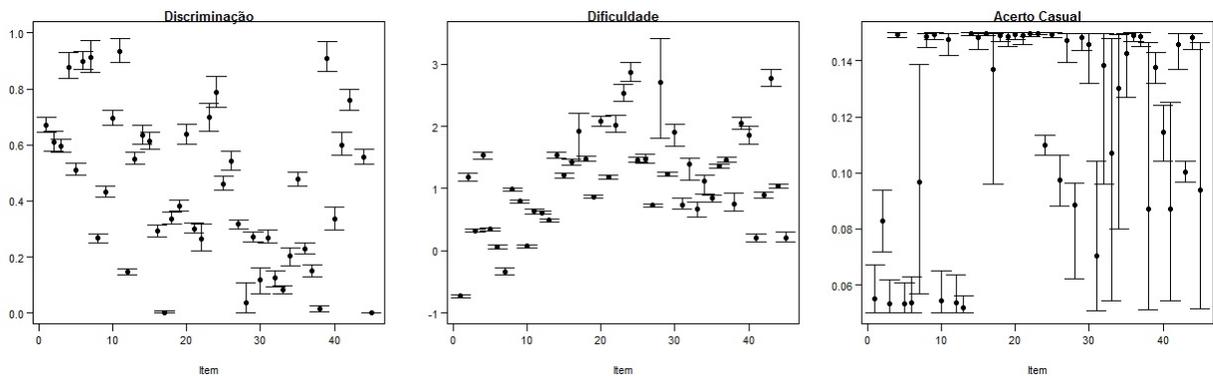


Figura 5.2: Estimativa pontual (média *a posteriori*) e intervalar (95%) dos parâmetros dos itens. O ponto no gráfico representa a estimativa pontual de cada parâmetro e a linha horizontal dentro do intervalo representa a estimação intervalar.

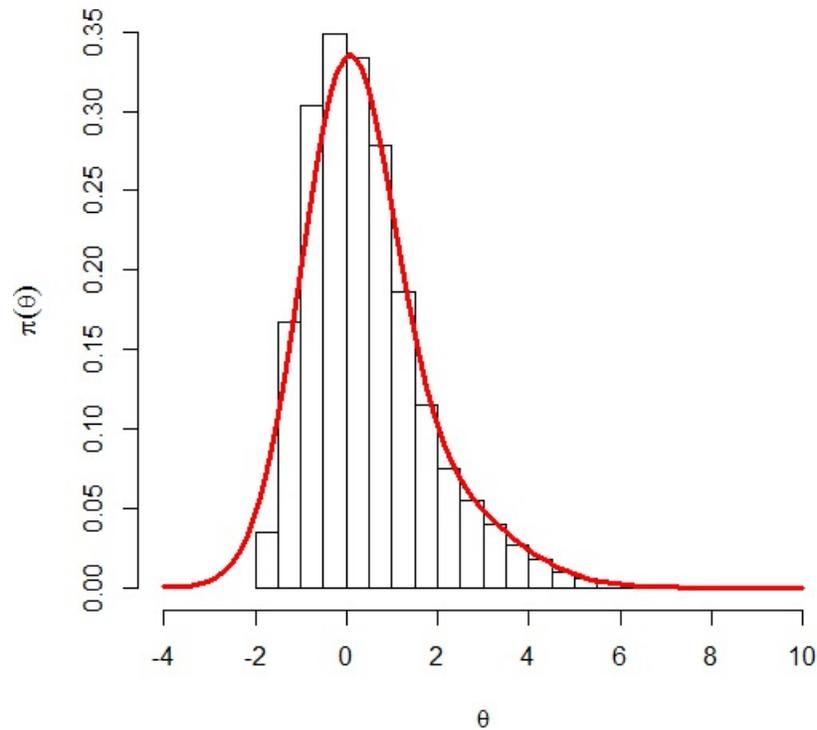


Figura 5.3: Densidade estimada e histograma das médias *a posteriori* das proficiências dos alunos.

5.1 Análise dos dados no ENEM utilizando o modelo Tradicional da TRI

Nesta seção será feita uma comparação entre o modelo da TRI com mistura e o modelo da TRI tradicional no ajuste de dados reais. Será utilizado o modelo Normal (apresentado na Seção 2.1) para ajustar o conjunto de dados do Exame Nacional do Ensino Médio (ENEM) utilizado anteriormente. O objetivo desta análise será mostrar a importância do novo modelo proposto em casos que a distribuição das proficiências não apresentam um comportamento normal.

Fixamos os parâmetros da distribuição de θ igual a média e variância da mistura estimada pelo modelo de mistura, que foi igual a $\theta \sim 0.76N(0, 1) + 0.24N(1.78, 2.62)$. Baseado na definição 3.1, fixamos $\theta \sim N(0.43, 1.97)$. Isto é feito para que os dois modelos sejam comparados em uma mesma escala.

	Valor Estimado
μ_2	1.78
σ_2^2	2.62
p_1	0.76
p_2	0.24

Tabela 5.1: Média *a posteriori* para os parâmetros da mistura.

Através da Figura 5.4 é possível verificar que o modelo tradicional detectou uma assimetria positiva bem suave, não sendo capaz de detectar a mesma assimetria positiva dada pelo modelo de mistura. Ao fixarmos as proficiências em uma distribuição normal, não permitimos que os dados tenham um comportamento muito diferente do comportamento normal, por isso a importância em utilizar o nosso modelo de mistura proposto em ajustes de dados reais.

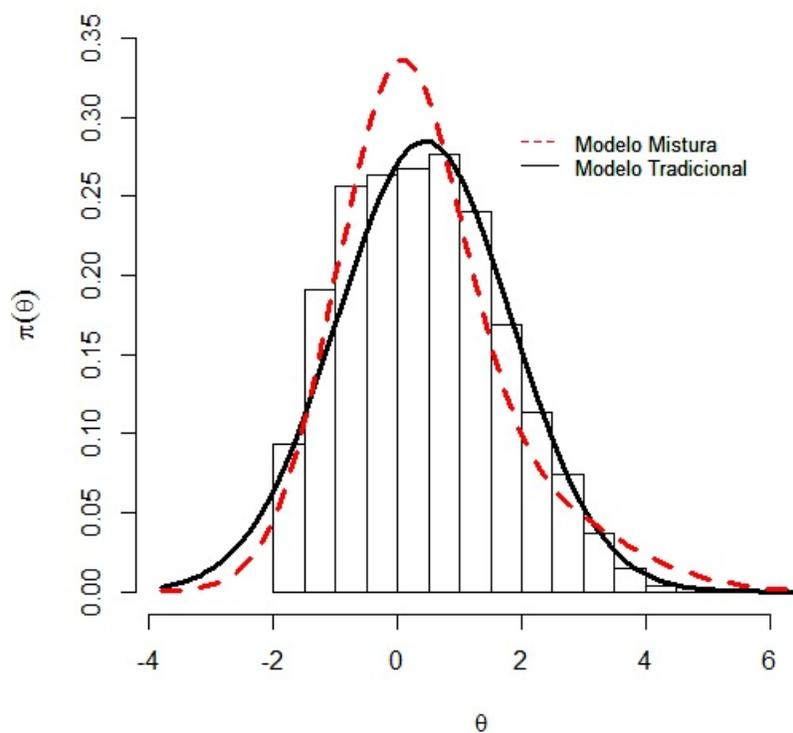


Figura 5.4: Densidade estimada do modelo de mistura, densidade fixa do modelo tradicional e histograma das médias *a posteriori* das proficiências dos alunos do modelo tradicional.

Capítulo 6

Conclusões

Com o objetivo de flexibilizar a suposição de normalidade das proficiências e generalizar os modelos tradicionais da TRI para resposta dicotômica convencional, propomos um novo modelo da TRI que utiliza uma mistura de duas distribuições normais para modelar as proficiências do indivíduos.

Foram discutidas e propostas soluções para importantes problemas de identificabilidade associadas ao modelo proposto. A inferência é feita sob a abordagem Bayesiana através de um algoritmo MCMC que amostra da distribuição a posteriori conjunta das quantidades desconhecidas do modelo. Este mesmo algoritmo é uma nova abordagem também para a estimação nos modelos tradicionais e possui boas propriedades de convergência ao ser comparado com propostas anteriores.

Através de estudos simulados, mostrou-se que esta abordagem permite uma solução eficiente para modelar distribuições com caudas pesadas, assimetria e bimodalidade. De tal forma, foi possível verificar a eficiência do modelo proposto, recuperando de forma satisfatória os valores reais em estudo. Além disso, tal modelo pode ser considerado uma generalização do modelo TRI convencional, pois ele também modela de forma eficiente proficiências que seguem uma distribuição normal.

Por meio de um estudo de comparação foi possível verificar a importância em modelar de forma adequada tal distribuição. Através de resultados obtidos, é possível concluir que os modelos tradicionais não conseguem estimar bem a real distribuição das proficiências.

O modelo também foi aplicado a um conjunto de dados reais do programa Exame Nacional do Ensino Médio (ENEM), analisando o teste de Matemática, aplicado ao estado de São Paulo no ano de 2010. Tal análise gerou resultados importantes do ponto de vista prático, pois ao usarmos o

modelo proposto, detectamos uma assimetria a direita presente nas proficiências.

Apêndice A

Neste apêndice estão os cálculos para se obter as distribuições condicionais completas necessárias para Método de simulação Monte Carlo baseados em Cadeia de Markov (MCMC) referentes ao Lema 1.

- **Condicional Completa para (Z_{ij}, X_{ij})**

A distribuição condicional completa conjunta para (Z_{ij}, X_{ij}) , com $i = 1, \dots, I$ e $j = 1, \dots, J$ é proporcional ao produto da distribuição de Z_{ij} pela distribuição de Y_{ij} e de X_{ij} , onde $Z_{ij} \sim \text{Bernoulli}(c_i)$, $(X_{ij}|Z_{ij} = 0) \sim N(a_i\theta_j - b_i, 1)$ e $(X_{ij}|Z_{ij} = 1) = \delta_0$, onde δ_0 é um ponto de massa em zero. A distribuição de Y_{ij} é dada em (2.1).

$$\pi(Z_{ij}, X_{ij}|Y_{ij}, a_i, b_i, c_i) \propto \pi(Y_{ij}|X_{ij}, Z_{ij}) \pi(Z_{ij}|c_i) \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j) \quad (6.1)$$

Se $Y_{ij} = 0$ ($Z_{ij} = 0, X_{ij} < 0$):

$$\pi(Z_{ij}, X_{ij}|Y_{ij} = 0, a_i, b_i, c_i) \propto I_{(Z_{ij}=0)}I_{(X_{ij}<0)} \pi(X_{ij}|Z_{ij} = 0, a_i, b_i, \theta_j) \quad (6.2)$$

Dessa forma:

$$Z_{ij}, X_{ij}|Y_{ij} = 0, a_i, b_i, c_i \propto \phi(x_{ij} - m)I_{(Z_{ij}=0)}I_{(X_{ij}>0)}. \quad (6.3)$$

Se $Y_{ij} = 1$ ($Z_{ij} = 1, X_{ij} = 0$ ou $Z_{ij} = 0, X_{ij} \geq 0$):

$$\begin{aligned}
& \pi(Z_{ij}, X_{ij} | Y_{ij} = 1, a_i, b_i, c_i) \propto \\
& \propto [I_{(Z_{ij}=1)}I_{(X_{ij}=0)} + I_{(Z_{ij}=0)}I_{(X_{ij}>0)}] \pi(Z_{ij}|c_i) \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j) \\
& \propto \pi(Z_{ij}|c_i) \left[\pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j)I_{(Z_{ij}=1)}I_{(X_{ij}=0)} + \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j)I_{(Z_{ij}=0)}I_{(X_{ij}>0)} \right] \\
& \propto \pi(Z_{ij}|c_i) \left[1I_{(Z_{ij}=1)}I_{(X_{ij}=0)} + f_N(X_{ij}, a_i\theta_j - b_i, 1)I_{(Z_{ij}=0)}I_{(X_{ij}>0)} \right] \\
& \propto c_i I_{(Z_{ij}=1)}I_{(X_{ij}=0)} + (1 - c_i) \Phi(a_i\theta_j - b_i) \frac{\phi(x_{ij} - m)}{\Phi(a_i\theta_j - b_i)} I_{(Z_{ij}=0)}I_{(X_{ij}>0)} \quad (6.4)
\end{aligned}$$

Com $w = \frac{c_i}{c_i + (1 - c_i) \Phi(a_i\theta_j - b_i)}$, temos:

$$\pi(Z_{ij}, X_{ij} | Y_{ij} = 1, a_i, b_i, c_i) = w I_{(Z_{ij}=1)}I_{(X_{ij}=0)} + (1 - w) \frac{\phi(x_{ij} - m)}{\Phi(m)} I_{(Z_{ij}=0)}I_{(X_{ij}>0)}$$

- **Condicional Completa para c_i**

A distribuição condicional completa para c_i , com $i = 1, \dots, I$ é proporcional ao produto da distribuição de Z_{ij} pela distribuição *a priori* de c_i , onde $Z_{ij} \sim \text{Bernoulli}(c_i)$ e $c_i \sim \text{Beta}(\alpha_c, \beta_c)$.

$$\pi(c_i | Z_{ij}, X_{ij}, a_i, b_i, \theta) \propto \prod_{j=1}^J \pi(Z_{ij} | c_i) \pi(c_i) \quad (6.5)$$

$$\begin{aligned}
\pi(c_i | Z_{ij}, X_{ij}, a_i, b_i) & \propto \prod_{j=1}^J \pi(Z_{ij} | c_i) \pi(c_i) \\
& \propto \prod_{j=1}^J \left[c_i^{Z_{ij}} (1 - c_i)^{1 - Z_{ij}} \right] c_i^{\alpha_c - 1} (1 - c_i)^{\beta_c - 1} \\
& \propto c_i^{\sum_{j=1}^J Z_{ij} + \alpha_c - 1} (1 - c_i)^{J - \sum_{j=1}^J Z_{ij} + \beta_c - 1} \quad (6.6)
\end{aligned}$$

Dessa forma:

$$c_i | Z_{ij}, X_{ij}, a_i, b_i \sim \text{Beta} \left(\sum_{j=1}^J Z_{ij} + \alpha_c, J - \sum_{j=1}^J Z_{ij} + \beta_c \right) \quad (6.7)$$

- **Condicional Completa para θ_j**

A distribuição condicional completa para θ_j , com $j = 1, \dots, J$ é proporcional ao produto da distribuição de X_{ij} pela distribuição *a priori* de θ_j , onde $(X_{ij} | Z_{ij} = 0) \sim N(a_i\theta_j - b_i, 1)$ e

$(X_{ij}|Z_{ij} = 1) = \delta_0$, com δ_0 sendo um ponto de massa em zero. E $\theta_j \sim N(\mu_\theta, \sigma_\theta^2)$.

$$\pi(\theta_j|\mathbf{Z}, \mathbf{X}, \mathbf{a}, \mathbf{b}) \propto \prod_{i=1}^{L_j} \pi(X_{ij}|Z_{ij}, a_i, b_i, \theta_j) \pi(\theta_i) \quad (6.8)$$

Onde $L_j\{i, z_{ij} \neq 0\}$.

$$\begin{aligned} \pi(\theta_j|Z_{ij} = 0, X_{ij}, a_i, b_i) &\propto \prod_{i=1}^{L_j} \pi(X_{ij}|Z_{ij} = 0, a_i, b_i, \theta_j) \pi(\theta_i) \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{L_j} (x_{ij} - a_i \theta_j + b_i)^2 \right\} \exp \left\{ \frac{1}{2\sigma_\theta^2} (\theta_j - \mu_\theta)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(-2\theta_j \sum_{i=1}^{L_j} a_i x_{ij} + \theta_j^2 \sum_{i=1}^{L_j} a_i^2 - 2\theta_j \sum_{i=1}^{L_j} a_i b_i + \frac{\theta_j^2}{\sigma_\theta^2} - 2\theta_j \frac{\mu_\theta}{\sigma_\theta^2} \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\theta_j^2 \left(\sum_{i=1}^{L_j} a_i^2 + \frac{1}{\sigma_\theta^2} \right) - 2\theta_j \sum_{i=1}^{L_j} a_i x_{ij} + \sum_{i=1}^{L_j} a_i b_i + \frac{\mu_\theta}{\sigma_\theta^2} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\frac{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1}{\sigma_\theta^2} \right) \left[\theta_j^2 - \frac{\left(\sigma_\theta^2 \sum_{i=1}^{L_j} a_i (x_{ij} + b_i) + \mu_\theta \right)}{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1} \right]^2 \right\} \end{aligned} \quad (6.9)$$

Dessa forma:

$$\theta_j|Z_{ij} = 0, X_{ij}, a_i, b_i \sim N \left(\frac{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i (x_{ij} + b_i) + \mu_\theta}{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1}, \frac{\sigma_\theta^2}{\sigma_\theta^2 \sum_{i=1}^{L_j} a_i^2 + 1} \right) \quad (6.10)$$

- **Condicional Completa para (a_i, b_i)**

A distribuição condicional completa conjunta para (a_i, b_i) , com $i = 1, \dots, I$ é proporcional ao produto da distribuição de X_{ij} pela distribuição *a priori* de a_i e pela distribuição *a priori* de b_i , onde $(X_{ij}|Z_{ij} = 0) \sim N(a_i \theta_j - b_i, 1)$ e $(X_{ij}|Z_{ij} = 1) = \delta_0$, onde δ_0 é um ponto de massa em zero. Onde $a_i \sim N_{(0, \infty)}(\mu_a, \sigma_a^2)$ e $b_i \sim N(\mu_b, \sigma_b^2)$.

$$\pi(a_i, b_i | Z_{ij}, X_{ij}, \theta_j) \propto \prod_{j=1}^{L_i} \pi(X_{ij} | Z_{ij}, a_i, b_i, \theta_j) \pi(a_i) \pi(b_i) \quad (6.11)$$

Com $L_i\{j, z_{ij} \neq 0\}$.

$$\begin{aligned} \pi(a_i, b_i | Z_{ij} = 0, X_{ij}, \theta_j) &\propto \prod_{j=1}^{L_i} \pi(X_{ij} | Z_{ij} = 0, a_i, b_i, \theta_j) \pi(a_i) \pi(b_i) \\ &\propto \exp \left\{ \frac{1}{2} \sum_{j=1}^{L_i} (x_{ij} - a_i \theta_j + b_i)^2 \right\} \exp \left\{ -\frac{1}{2} \left(\frac{-a_i^2}{\sigma_a^2} + \frac{-b_i^2}{\sigma_b^2} \right) \right\} \end{aligned} \quad (6.12)$$

Após algumas álgebras, temos que:

$$(a_i, b_i | Z_{ij}, X, c, \theta) \sim N_2(\mu, \Sigma), \quad (6.13)$$

com $\mu = \begin{bmatrix} \mu_a^* \\ \mu_b^* \end{bmatrix}$ e $\Sigma = \begin{bmatrix} \sigma_a^{2*} & \gamma \\ \gamma & \sigma_b^{2*} \end{bmatrix}$ sendo o vetor de média e matriz de covariância de uma distribuição normal bivariada, respectivamente. Onde:

$$\sigma_a^{2*} = \frac{\sigma_a^2}{\left(\sigma_a^2 \sum_{j=1}^{L_i} \theta_j^2 + 1 \right) (1-\gamma^2)}, \quad \sigma_b^{2*} = \frac{\sigma_b^2}{(\sigma_b^2 J + 1)(1-\gamma^2)}, \quad \gamma = \frac{\sigma_a \sigma_b \sum_{j=1}^{L_i} \theta_j}{\left[\left(\sigma_a^2 \sum_{j=1}^{L_i} \theta_j^2 + 1 \right) (\sigma_b^2 J + 1) \right]^{\frac{1}{2}}},$$

$$\mu_a^* = \sigma_a^{2*} \left(\sum_{j=1}^{L_i} x_{ij} \theta_j + \mu_a \sigma_a^{-2} \right) - \sigma_a^* \sigma_b^* \gamma \left(\sum_{j=1}^{L_i} x_{ij} - \mu_b \sigma_b^{-2} \right) \text{ e}$$

$$\mu_b^* = \sigma_a^* \sigma_b^* \gamma \left(\sum_{j=1}^{L_i} x_{ij} \theta_j + \mu_a \sigma_a^{-2} \right) - \sigma_b^{2*} \left(\sum_{j=1}^{L_i} x_{ij} - \mu_b \sigma_b^{-2} \right).$$

Com $L_i\{j, z_{ij} \neq 0\}$.

Apêndice B

Neste apêndice estão os cálculos para se obter as distribuições condicionais completas necessárias para Método de simulação Monte Carlo baseados em Cadeia de Markov (MCMC) referentes ao Lema 2.

- **Condicional Completa para μ_2, σ_2 :**

A distribuição condicional completa para μ_2, σ_2 , com é proporcional ao produto da distribuição de θ_j pela distribuição *a priori* de (μ_2, σ_2^2) , onde θ_j segue uma mistura de normais dado em (3.4) e $(\mu_2, \sigma_2^2) \sim NIG(m, \frac{\sigma_2^2}{\beta}, d, e)$.

$$\pi(\mu_2, \sigma_2^2 | W, \theta) \propto \left[\prod_{j=1}^J \pi(\theta_j | \mu_2, \sigma_2^2, W) \right] \pi(\mu_2, \sigma_2^2) \quad (6.14)$$

$$\begin{aligned} \pi(\mu_2 | \sigma_2^2, W, \theta) &\propto \left\{ \prod_{j=1}^J \left[\frac{1}{\sigma_2} \phi_2 \left(\frac{\theta_j - \mu_2}{\sigma_2} \right) \right]^{W_{j2}} \right\} \frac{\beta^{\frac{1}{2}}}{\sigma_2} \phi \left(\frac{\mu_2 - m}{\sigma_2} \right) (\sigma_2^2)^{-(d+1)} \exp \left\{ -\frac{e}{\sigma_2^2} \right\} \\ &\propto \exp \left\{ -\frac{\sum_{j=1}^J W_{j2} + \beta}{2\sigma_2^2} \left[\mu_2^2 - 2\mu_2 \left(\frac{\sum_{j=1}^J W_{j2}\theta_j + m\beta}{\sum_{j=1}^J W_{j2} + \beta} \right) \right] \right\} \end{aligned}$$

Dessa forma:

$$\mu_2, \sigma_2^2 | W, \theta \sim NIG \left(m^*; \frac{\sigma_2^2}{\beta^*}, d^*, e^* \right), \quad (6.15)$$

$$\text{onde } m^* = \frac{\sum_{j=1}^J W_{j2} \theta_j + m\beta}{\sum_{j=1}^J W_{j2} + \beta}, \quad \beta^* = \sum_{j=1}^J W_{j2} + \beta, \quad d^* = d + \left(\sum_{j=1}^J W_{j2} \right) / 2,$$

$$e^* = e + \frac{\sum_{j=1}^J W_{j2} \beta}{2 \left(\beta + \sum_{j=1}^J W_{j2} \right)} (m - \theta_{\text{bar}})^2 + s/2, \text{ onde } s = \sum_{j=1}^J W_{j2} \theta_j^2 - \sum_{j=1}^J W_{j2} \text{ e}$$

$$\theta_{\text{bar}} = \frac{\sum_{j=1}^J W_{j2} \theta_j}{\sum_{j=1}^J W_{j2}}.$$

• **Condicional Completa para $p = (p_1, p_2)$**

A distribuição condicional completa para $p = (p_1, \dots, p_k)$ é proporcional ao produto da distribuição de W pela distribuição *a priori* de $p = (p_1, p_2)$, onde $W \sim \text{Mult}(1, p_1, p_2)$ e $p \sim \text{Dir}(\alpha_1, \alpha_2) I_{(p_1 > 0.6)}$.

$$\pi(p|W) \propto \left[\prod_{j=1}^J \pi(W_j|p) \right] \pi(p) \quad (6.16)$$

$$\begin{aligned} \pi(p|W) &\propto \left[\prod_{j=1}^J p_1^{W_{j1}} p_2^{W_{j2}} \right] p_1^{\alpha_1+1} p_2^{\alpha_2+1} I_{(p_1 > 0.6)} \\ &\propto \frac{\sum_{j=1}^J W_{j1} + \alpha_1 - 1}{p_1^{\sum_{j=1}^J W_{j1}}} \frac{\sum_{j=1}^J W_{j2} + \alpha_2 - 1}{p_2^{\sum_{j=1}^J W_{j2}}} I_{(p_1 > 0.6)} \end{aligned}$$

Dessa forma:

$$p|W \sim \text{Dir} \left(\sum_{j=1}^J W_{j1} + \alpha_1, \sum_{j=1}^J W_{j2} + \alpha_2 \right) I_{(p_1 > 0.6)}, \quad (6.17)$$

que é equivalente a:

$$p_1 | W \sim \text{Beta} \left(\sum_{j=1}^J W_{j1} + \alpha_1, \sum_{j=1}^J W_{j2} + \alpha_2 \right) I_{(p_1 > 0.6)} \quad (6.18)$$

• **Condicional Completa para (θ_j, W_j)**

A distribuição condicional completa conjunta para (θ_j, W_j) , com $j = 1, \dots, J$ é proporcional ao produto da distribuição de X_{ij} pela distribuição *a priori* de θ_j , onde $(X_{ij} | Z_{ij} = 0) \sim N(a_i \theta_j - b_i, 1)$ e $(X_{ij} | Z_{ij} = 1) = \delta_0$, onde δ_0 é um ponto de massa em zero e θ_j segue uma mistura de normais dado em na expressão (3.4).

$$\pi(\theta_j, W_j | Z_{ij}, X_{ij}, a, b, \mu_k, \sigma_k, p_k) \propto \prod_{k=1}^2 \pi(\theta_j | \mu_k, \sigma_k^2, W) \left[\prod_{i=1}^{L_j} \pi(X_{ij} | Z_{ij}, a_i, b_i, \theta_j) \right] \pi(W | p) \quad (6.19)$$

$$\begin{aligned} \pi(\theta_j, W_j | Z_{ij}, X_{ij}, a, b, \mu_k, \sigma_k, p_k) &\propto \prod_{k=1}^2 \left[\frac{1}{\sigma_k} \phi_k \left(\frac{\theta_j - \mu_k}{\sigma_k} \right) \prod_{i=1}^{L_j} \phi(X_{ij} - (a_i \theta_j - b_i)) \right]^{W_{jk}} p_k^{W_{jk}} \\ &\propto \prod_{k=1}^2 \left[\frac{1}{\sigma_k^*} \phi_k \left(\frac{\theta_j - \mu_k^*}{\sigma_k^*} \right) \right]^{W_{jk}} (\alpha_k^*)^{W_{jk}} \end{aligned} \quad (6.20)$$

$$\text{Com } L_j \{i, z_{ij} \neq 0\}, \mu_k^* = \frac{\mu_k + \sigma_k^2 \sum_{i=1}^{L_j} a_i (x_{ij} - b_i)}{1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2}, \sigma_k^{2*} = \frac{\sigma_k^2}{1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2} \text{ e}$$

$$\alpha_k^* = p_k \left(1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\mu_k^2}{\sigma_k^2} - \frac{\left(\mu_k + \sigma_k^2 \sum_{i=1}^{L_j} a_i (x_{ij} - b_i) \right)^2}{\sigma_k^2 \left(1 + \sigma_k^2 \sum_{i=1}^{L_j} a_i^2 \right)} \right] \right\}$$

Dessa forma:

$$\begin{aligned} \theta_j &| W_j, Z_{ij}, X_{ij}, a, b, \mu_k, \sigma_k, p_k \sim N(\mu_k^*, \sigma_k^{2*}) \\ W_j &| Z_{ij}, X_{ij}, a, b, \mu_k, \sigma_k, p_k \sim Mult(1, p_1^*, \dots, p_K^*), \end{aligned} \quad (6.21)$$

onde k é a componente que θ_j pertence e $p_k^* = \frac{\alpha_k^*}{\sum_{k=1}^K \alpha_k^*}$, com $k = 1, 2$.

Apêndice C

Apresenta-se aqui as cadeias geradas no estudo de simulação apresentado no Capítulo 4.

- Cadeias referentes ao estudo de simulação 1

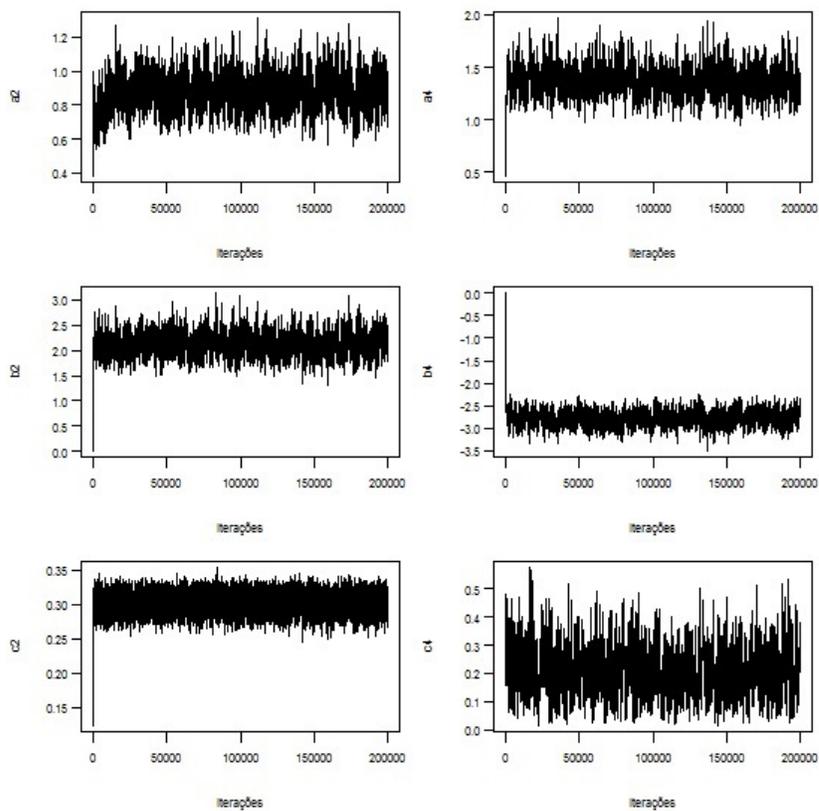


Figura 6.1: Trace Plots para cadeias de a , b e c .

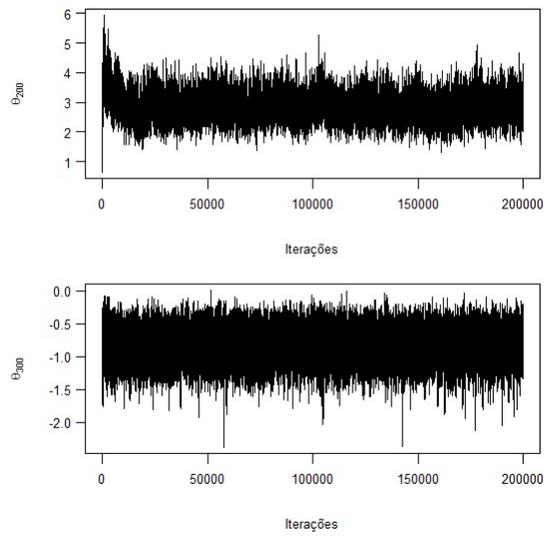


Figura 6.2: Trace Plots para cadeias de θ .

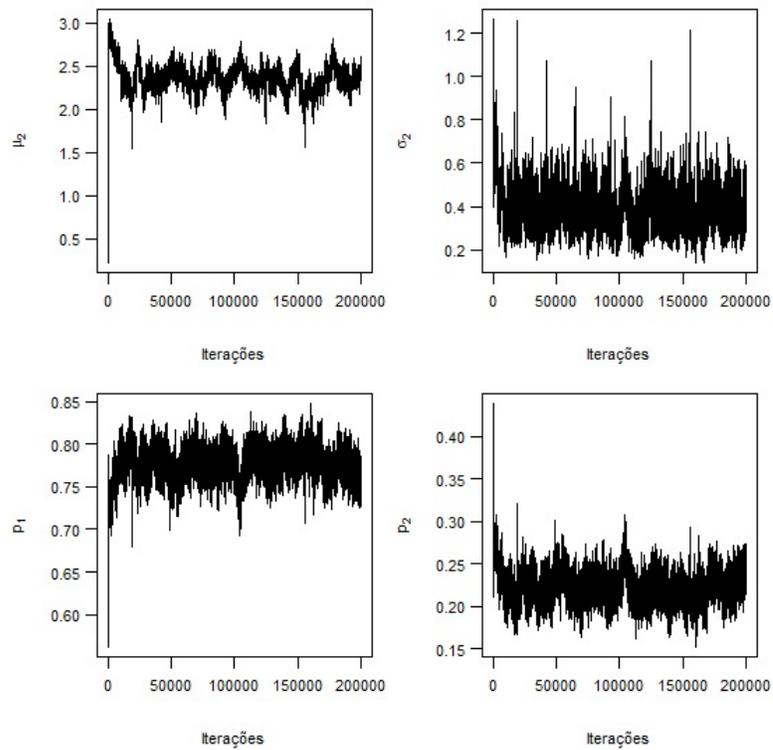


Figura 6.3: Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2 .

- Cadeias referentes ao estudo de simulação 2

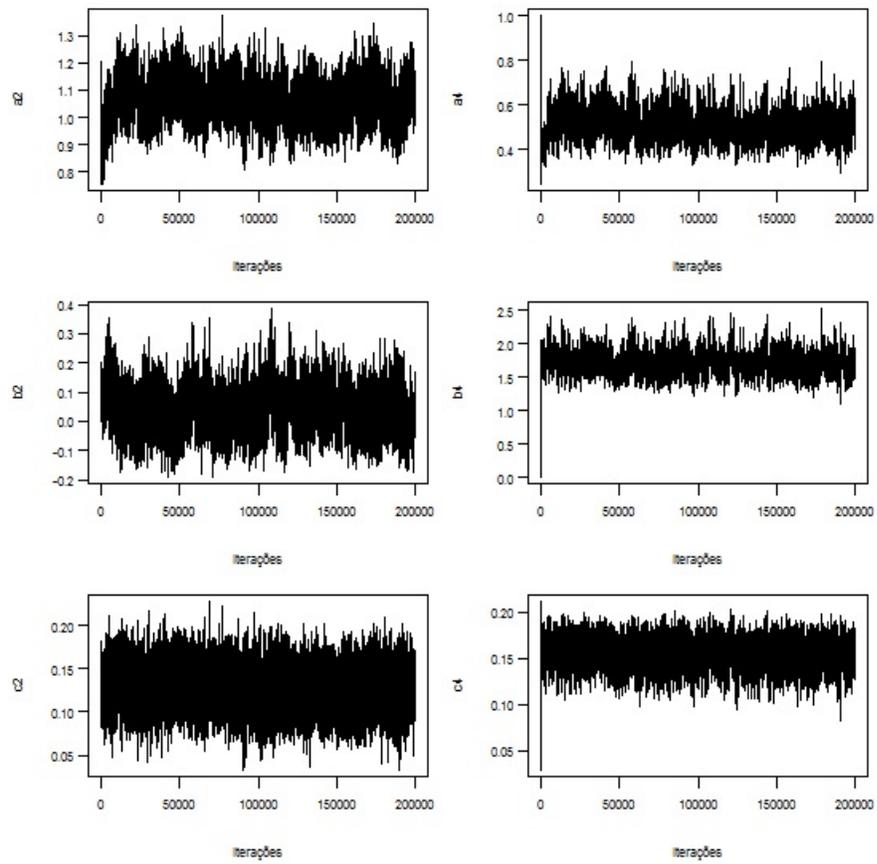


Figura 6.4: Trace Plots para cadeias de a , b e c .

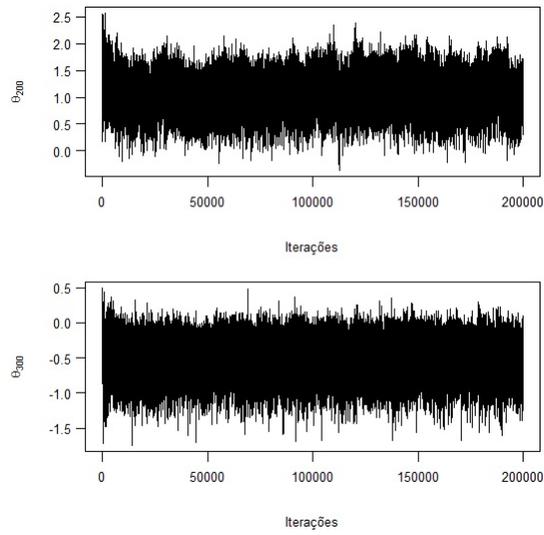


Figura 6.5: Trace Plots para cadeias de θ .

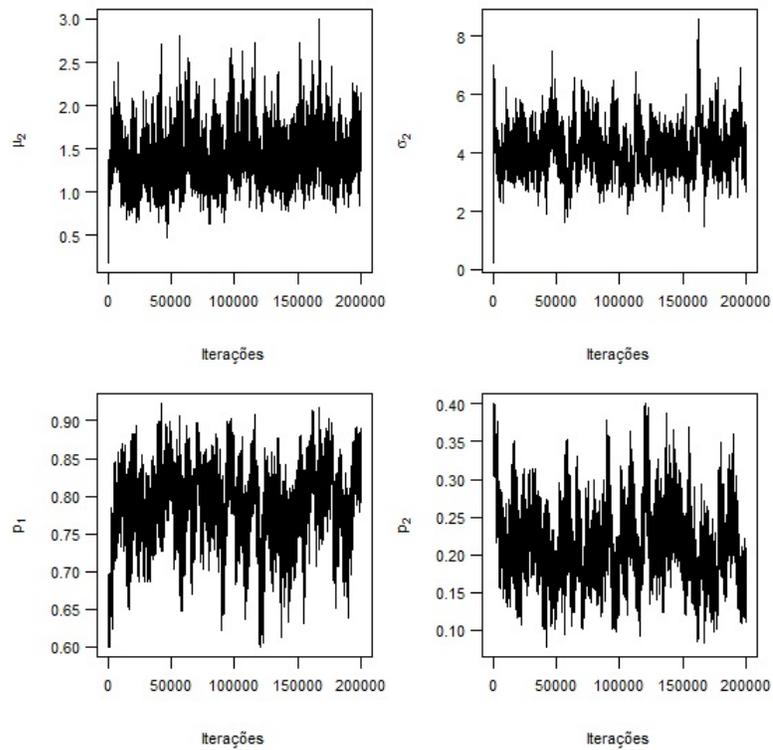


Figura 6.6: Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2 .

- Cadeias referentes ao estudo de simulação 3

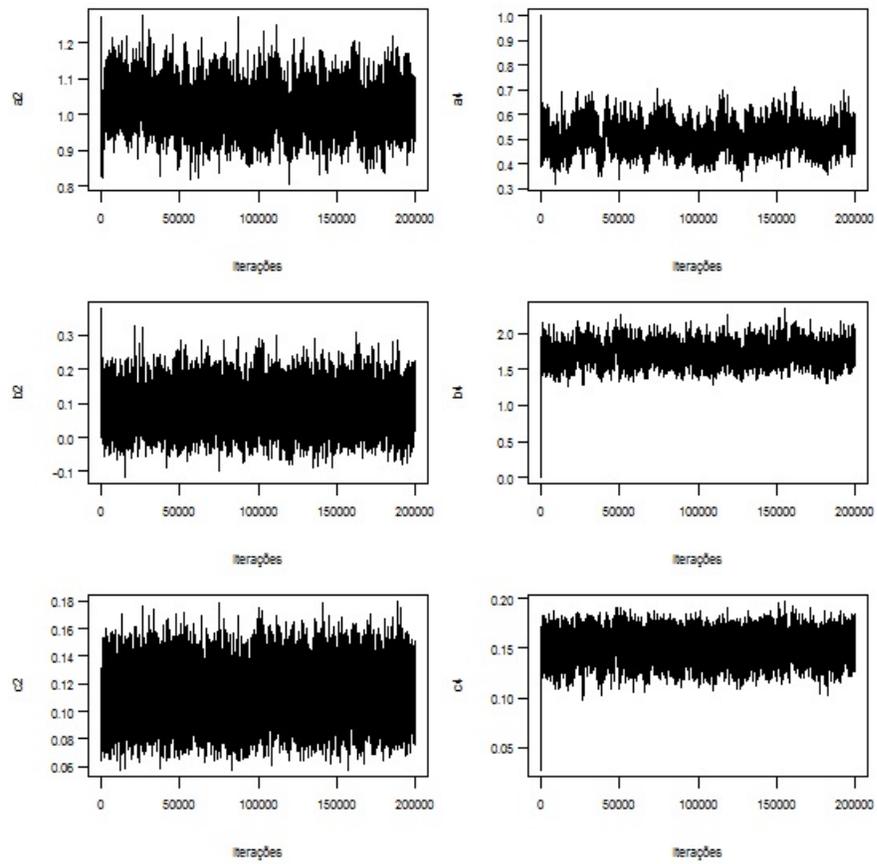


Figura 6.7: Trace Plots para cadeias de a , b e c .

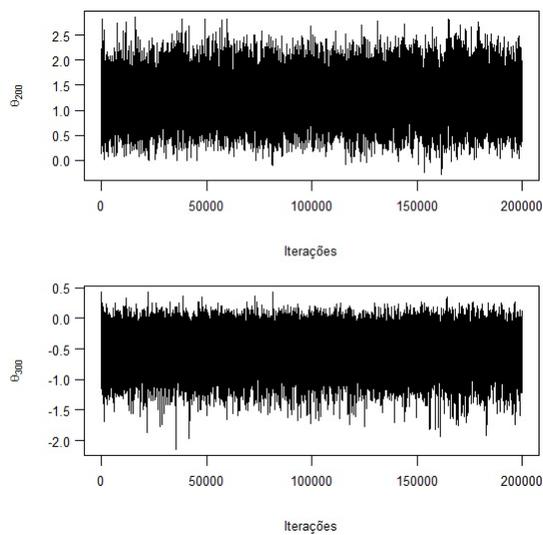


Figura 6.8: Trace Plots para cadeias de θ .

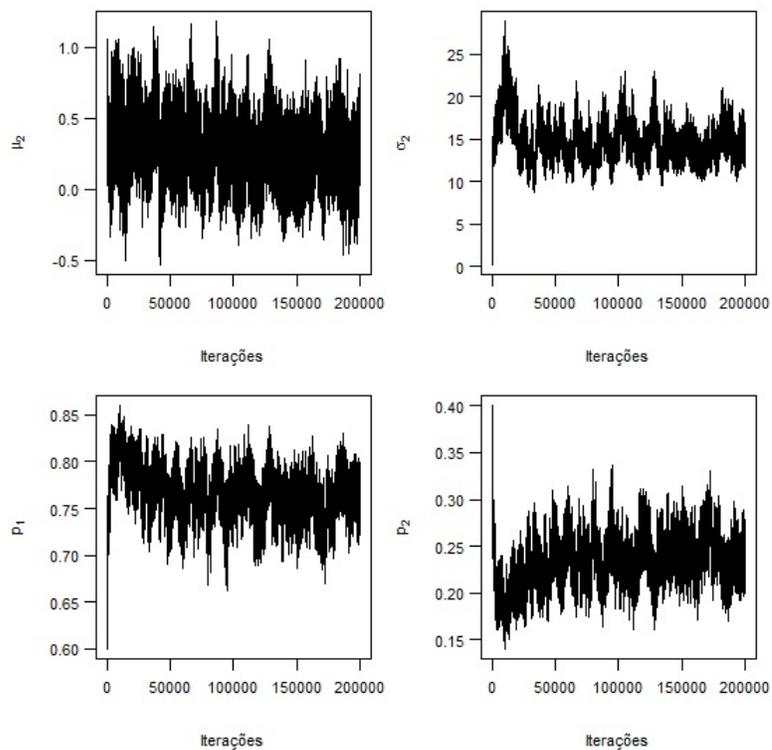


Figura 6.9: Trace Plots para cadeias de μ_2 , σ_2^2 , p_1 e p_2 .

Bibliografia

- [1] Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using gibbs sampling. *Journal of Educational and Behavioral Statistics*, 17 (3), 251-269.
- [2] Azevedo, C. L., Bolfarine, H., & Andrade, D. F. (2011). Bayesian inference for a skew-normal irt model under the centred parameterization. *Computational Statistics & Data Analysis*, 55 (1), 353-365.
- [3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, (pp. 171-178).
- [4] Béguin A. A., Glas C. A. W. (2001). MCMC estimation and some model-fit analysis of multi-dimensional IRT models, *Psychometrika*, 66, 541-562.
- [5] Baker, F.B., & Kim, S.H. (2004). Item response theory: Parameter estimation techniques. 2nd Ed. *New York: Marcel Dekker*.
- [6] Birnbaum, A. (1968) Some Latent Traits Models and Their Use in Inferring Examinee?s Ability. *Statistical Theories of Mental Test Scores*, edited by F. M. Lord and M. R. Novick, Reading, Ma. Addison-Wesley, 397-472.
- [7] de Boor, C. (1972). On calculating with B-splines. *Journal of Approximation Theory*, 6, 50-62.
- [8] Brooks, S. e Robert, G (1998). Assessment convergence of Markov Chain Monte Carlo algorithms. *Statistics and Computing*, 8, 319-335.
- [9] Gamerman, D. (1997). Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference. *Texts in Statistical Sciences. Chapman and Hall, London*.

- [10] Gelfand, A. E.; Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- [11] Geman, S., Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- [12] Lord, F. (1952). A theory of test scores. *Psychometric monographs*.
- [13] MCLACHLAN, G.J. PEEL, D. Finite mixture models. *New York: Wiley*, 413 p. 2000.
- [14] Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105 (1), 156.
- [15] Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, 49 (3), 359-381.
- [16] Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8, 343-366.
- [17] Pearson, K. (1894). Contributions to the mathematical theory of evolution, *Phil.Trans.A*, 185, 71-110.
- [18] Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Studies in mathematical psychology. Danmarks Paedagogiske Institut.
- [19] Samejima, F. (1997). Departure from normal assumptions: a promise for future psychometrics with substantive mathematical modeling. *Psychometrika*, 62 (4), 471-493.
- [20] Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, 71, 281-301.