Juan Leonardo Padilla Gomez

Modelos da teoria de resposta ao item multidimensionais assimétricos de grupos múltiplos para respostas dicotômicas sob um enfoque bayesiano

CAMPINAS
2014

Universidade Estadual de Campinas

Instituto de Matemática, Estatística
e Computação Científica

Juan Leonardo Padilla Gomez

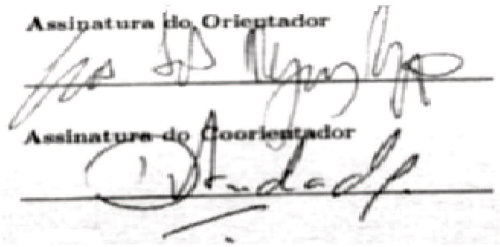
Modelos da teoria de resposta ao item multidimensionais assimétricos de grupos múltiplos para respostas dicotômicas sob um enfoque bayesiano

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em estatística.

Orientador: Prof. Dr. Caio Lucidius Naberezny Azevedo

Coorientador: Prof. Dr. Dalton Francisco de Andrade

Este exemplar corresponde a versão final da dissertação defendida pelo aluno Juan Leonardo Padilla Gomez, e orientada pelo Prof. Dr. Prof. Dr. Caio Lucidius Naberezny Azevedo.



Campinas
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

P134m

Padilla Gomez, Juan Leonardo, 1989-
Modelos da teoria de resposta ao item multidimensionais assimétricos de grupos múltiplos para respostas dicotômicas sob um enfoque bayesiano / Juan Leonardo Padilla Gomez. – Campinas, SP : [s.n.], 2014.

Orientador: Caio Lucidius Naberezny Azevedo.
Coorientador: Dalton Francisco de Andrade.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Teoria de resposta ao item. 2. Distribuição normal assimétrica. 3. Métodos MCMC. I. Azevedo, Caio Lucidius Naberezny,1979-. II. Andrade, Dalton Francisco de,1951-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Assimetric multidimensional item response theory models for multiple groups and dichotomic responses under a bayesian perspective
Palavras-chave em inglês:
Item response theory
Skew-normal distribuition
MCMC methods
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Caio Lucidius Naberezny Azevedo [Orientador]
Mariana Cúri
Victor Hugo Lachos Dávila
Data de defesa: 06-03-2014
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 06 de março de 2014 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO

Prof(a). Dr(a). VÍCTOR HUGO LACHOS DÁVILA

Prof(a). Dr(a). MARIANA CÚRI

Abstract

In this work it is proposed a new class of Multidimensional Item Response Theory (MIRT) models for dichotomic or dichotomized answers considering a multiple group structure. For the latent traits distribution of each group, it is proposed a new parametrization of the centered multivariate skew normal distribution, which combines the proposed by Lachos (2004) and the one proposed by Arellano-Valle et.al (2008), which not only ensures de identifiability of our proposed models, but also it makes simpler the interpretation and estimation of their parameters. Hence, our model stands as an important alternative, in order to solve the identifiability problems found for the one group multidimensional skewed models proposed by Matos (2010) and Nojosa (2008). Simulation studies, taking into account some situations of practical interest, were conducted in order to evaluate the potential of the triad: modeling, estimation methods and diagnostic tools. The results indicate that the models considering a skew component on the latent traits, in general, produced more accurate results than those ones obtained with the symmetric models. For model selection, it was used the deviance information criterion (DIC), the expected values of both the AkaikeŠs information criterion (EAIC) and bayesian information criteron (EBIC). Concerning assessment of model fit quality, it was explored posterior predictive checking methods, which allows for evaluating the quality of the measure instrument as well as the quality fit of the model from a global point of view and related to specific assumptions, as the test dimensionality. Concerning the estimation methods, it was adopted and implemented several MCMC algorithms proposed in the literature for other models, including the convergence accelerating propose algorithm by Gonzalez (2004), which were compared concerning some convergence quality aspects through the Sahu (2002) effective sample size. The analysis of a real data set, from the 2013 first stage of the UNICAMP admission exam was done as well.

Keywords: Multidimensional Item Response Theory, identifiability, multivariate skew normal, centered parametrization, bayesian estimation, MCMC algorithm, augmented data procedures.

Resumo

No presente trabalho propõe-se novos modelos da Teoria de Resposta ao Item Multidimensional (TRIM) para respostas dicotômicas ou dicotomizadas considerando uma estrutura de grupos múltiplos. Para as distribuições dos traços latentes de cada grupo, propõe-se uma nova parametrização da distribuição normal assimétrica multivariada centrada, que combina as propostas de Lachos (2004) e de Arellano-Valle et.al (2008), a qual não só garante a identificabilidade dos modelos aqui introduzidos, mas também facilita a interpretação e estimação dos seus parâmetros. Portanto, nosso modelo representa uma alternativa interessante, para solucionar os problemas de falta de identificabilidade encontrados por Matos (2010) e Nojosa (2008), nos modelos multidimensionais assimétricos de um único grupo por eles desenvolvidos. Estudos de simulação, considerando vários cenários de interesse prático, foram conduzidos a fim de avaliar o potencial da tríade: modelagem,

métodos de estimação e ferramentas de diagnósticos. Os resultados indicam que os modelos considerando a assimetria nos traços latentes, em geral, forneceram estimativas mais acuradas que os modelos tradicionais. Para a seleção de modelos, utilizou-se o critério de informação deviance (DIC), os valores esperados do critério de informação de Akaike (EAIC) e o critério de informação bayesiano (EBIC). Em relação à verificação da qualidade do ajuste de modelos, explorou-se alguns métodos de checagem preditiva a posteriori, os quais fornecem meios para avaliar a qualidade tanto do instrumento de medida, quanto o ajuste do modelo de um ponto de vista global e em relação à suposições específicas, entre elas a dimensão do teste. Com relação aos métodos de estimação, adaptou-se e implementou-se vários algoritmos MCMC propostos na literatura para outros modelos, inclusive a proposta de aceleração de convergência de González (2004), os quais foram comparados em relação aos aspectos de qualidade de convergência através do critério de tamanho efetivo da amostra de Sahu (2002). A análise de um conjunto de dados reais, referente à primeira fase do vestibular da UNICAMP de 2013 também foi realizada.

Palavras-chave: Teoria de Resposta ao Ítem Multidimensional, inferência bayesiana, identificabilidade, distribuição normal assimétrica multivariada, parametrização centrada, algoritmos MCMC, dados aumentados.

Sumário

Dedico este trabalho à minha mãe ...

Take the time to make some sense
Of what you want to say
And cast your words away upon the waves
Bring them back with Acquiesce
On a ship of hope today
And as they fall upon the shore
Tell them not to fear no more
Say it loud and sing it proud
And they...
Will dance if they want to dance
Please brother take a chance
You know they're gonna go
Which way they wanna go
All we know is that we don't know
What is gonna be
Please brother let it be
Life on the other hand won't let you understand
Why we're all part of the masterplan
I'm not saying right is wrong
It's up to us to make
The best of all things that come our way
And all the things that came have past
The answer's in the looking glass
There's four and twenty million doors
Down life's endless corridor
Say it loud and sing it proud
And they...
(...)
Noel Galagher
The Masterplan

Agradecimentos

Os esforços na construção deste trabalho teriam sido em vão se, ao final, não pudéssemos agradecer a aquelas pessoas que ajudaram e encorajaram a construí-lo.

Começarei agradecendo àquela pessoa que sempre esteve em todo momento, que me encorajou a continuar com meus estudos em um pais e uma cultura até então desconhecida, que com seus gestos e palavras cheias de carinho e amor fizeram que não me sentisse tao longe do meu lar. Para você mãe querida, meu respeito, meu eterno amor, e minha admiração.

Prossigo com meus irmãos, que sempre foram meus modelos a seguir, e para os quais sempre sere o seu irmãozinho. Principalmente ao meu irmão, Giovanni por me ter convidado e recebido nos períodos de ferias na sua casa.

À minha namorada Camila, pelo carinho e os incontáveis momentos compartilhados, não tenho palavras para agradecer o teu amor e companhia .

Ao meu orientador, Caio Azevedo pela inestimável experiência, conselhos, confiança e especialmente pela enorme paciência no tempo da construção do presente trabalho, especialmente por se ter arriscado em um projeto com um estudante cuja lingua materna não é o português. O conhecimento que ganhei trabalhando perto do senhor, realmente é imenso.

Ao meu coorientador, o Professor Doutor Dalton Francisco de Andrade, pela enorme experiência emprestada para a construção do presente trabalho. Ao Prof. Jayme Vaz Jr., Professor do Departamento de Matemática Aplicada do IMECC que no momento do desenvolvimento deste trabalho era o coordenador de pesquisa da COMVEST. Ao Renato Hirata, que no momento do desenvolvimento deste trabalho era Gerente de Informática por ser quem formatou e nos enviou todas as bases de dados relativas ao Vestibular 2013 da Unicamp.

Aos colegas dos cursos de Pos-graduação do IMECC com os quais dividi bons e difíceis momentos, particularmente a Abel, Julian e Ali.

Aos membros participantes da banca examinadora pelas sugestões.

A voce leitor, que depois de tudo, é a principal motivação deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro.

Lista de Ilustrações

Lista de Tabelas

Capítulo 1

Introdução

1.1 Aspectos gerais

1.1.1 Modelos da Teoria de Resposta ao Item

Em muitas áreas do conhecimento é bastante comum encontrar conjuntos de dados que estão relacionados a variáveis não observáveis diretamente, denominadas variáveis latentes. Um exemplo onde tais variáveis são de interesse, esta relacionado a estudantes, sobre os quais se tem interesse em analisar características de aprendizagem, tais como: facilidades de aprendizagem de habilidades novas, facilidade de obtenção de boas notas e de relacionar diferentes fontes de informação. Cada uma destas habilidades podem ser chamadas de traços latentes. Outros exemplos podem ser encontrados em Hambleton, Swaminathan e Rogers (1991). A medição (obtenção dos valores) dos traços latentes pode ser muito complexa. Isto vem do fato de que "habilidades", como as descritas anteriormente, não podem ser medidas diretamente como outras características, tais como a altura ou o peso. Em casos onde é requerido medir alguma característica, é indispensável ter à mão alguma ferramenta para poder não só atribuir um número àquilo que se está medindo, mas também poder realizar comparações de interesse. Isto é, precisa-se de uma escala de medição ou em outras palavras, uma régua com uma métrica bem definida. Em Baker (2001) é mencionado como em casos de pesquisa psicológica ou educacional, muitas vezes os pesquisadores só tem à mão os resultados de uma pesquisa ou de um teste, e tentam de alguma forma, ligá-los com aquele traço latente que é de interesse. A modelagem estatística é então usada para representar esta relação.

A Teoria de Resposta ao Item (TRI) (Hambleton e Swaminathan (1985)), é um conjunto de modelos, que representam o relacionamento entre características não observáveis de indivíduos (ou alguma outra unidade amostral de interesse), também chamadas de traços latentes, algumas características dos instrumentos de medida (teste, questionários, etc) e as respostas dadas a esses testes. Estes modelos foram desenvolvidos como resposta a problemas inerentes à metodologia predecessora, a denominada Teoria Clássica do Item (TCI).

A TCI foi durante décadas a ferramenta dominante nas pesquisas de avaliação nas áreas educacionais e da psicologia, ver Gulliksen (1950). A TRI, assim como a TCI, é uma tentativa de explicar o erro de medição. Na TCI, a modelagem do erro de medição é baseada no coeficiente de correlação. Desenvolvido por Charles Spearman, o coeficiente de correlação, tenta explicar o

erro mediante duas componentes: a verdadeira correlação e uma correlação observada (vide Crocker e Algina (1986), Traub (1997) e Courville (2004)). Porém, A TCI sofre de varias limitações amplamente referenciadas na literatura (por exemplo Embretson e Reise (2000) e Hambleton e Swaminathan (1985)). Em Courville (2004) se resume o problema com os estimadores da TCI no sentido que eles envolvem dependência circular, isto é, as estatísticas da TCI dependem da amostra no sentido de as estimativas pode mudar significativamente para diferentes amostras, além de que podem ser fortemente influenciadas pelo teste. Portanto, os estimadores da TCI não podem ser generalizados entre diferentes populações ou grupos. A TRI é, para muitos pesquisadores, a resposta à este tipo de limitações da TCI.

1.1.2 Motivação e revisão da literatura

Os modelos da TRI estão baseados no pressuposto de que os relacionamentos entre os indivíduos e os itens podem ser representados por um ou vários modelos. Os modelos da TRI tem sua origem no final da década dos 30. Particularmente em Richardson (1936) derivou-se o relacionamento entre os parâmetros do modelo da TRI e os parâmetros dos itens do modelo da TCI, o que forneceu um caminho inicial para a obtenção das estimativas dos parâmetros dos modelos na TRI. Em Muthen (1978) abordou-se o problema da estimação dos parâmetros dos modelos dando uma solução baseada em mínimos quadrados generalizados, enquanto que em Darrell Bock e Lieberman (1970) foi introduzido a estimação via Maxima Verossimilhança Marginal (MVM). Posteriormente Bock e Aitkin (1981) desenvolveram o algoritmo Esperança-Maximização (EM) para a estimação via MVM. O forte incremento do poder de processamento dos computadores nos últimos anos, tornou mais factível a implementação de métodos computacionalmente intensivos, o que por sua vez significou um ponto de inflexão no avanço da inferência estatística bayesiana em diversas áreas.

A inferência bayesiana de modelos complexos tem-se tornado atraente e muito foi desenvolvido para modelos referentes à Teoria da Resposta ao Item (vide Patz e Junker (1999a), Patz e Junker (1999b), Beguin e Glas (2001), Sahu (2002)). A abordagem via dados aumentados (Tanner e Wong (1987); Albert (1992)) também trouxe facilidades outras para a implementação de métodos computacionalmente intensivos, em modelos da Teoria da Resposta ao Item, através da inferência bayesiana com o uso de métodos de Monte Carlo via Cadeias de Markov.

Vários trabalhos têm desenvolvido algoritmos MCMC para estimação de MRI's. O primeiro deles é devido a Albert (1992), que desenvolve o amostrador de Gibbs, veja Gelfand e Smith (1990), sob a metodologia de dados aumentados Tanner e Wong (1987), para o modelo probito de 2 parâmetros considerando um único grupo de respondentes. Em Beguin e Glas (2001) foi desenvolvido um algoritmo baseado no amostrador de Gibbs com a estrutura de dados aumentados para o modelo probito de 3 parâmetros, para os caso uni e multidimensional. Outros trabalhos que utilizam MCMC na TRI são os de Bazan, Branco e Bolfarine (2006), Azevedo, Bolfarine e Andrade (2011), Azevedo, Bolfarine e Andrade (2012), Santos (2012) e Santos, Azevedo e Bolfarine (2013), dentre outros.

Continuando com os modelos da TRI, em Andrade, Tavares e Valle (2000), identificou-se como os modelos propostos na literatura da TRI dependem fundamentalmente de três fatores:

- da natureza do item - dicotômicos ou não dicotômicos;

- do número de populações envolvidas - apenas uma ou mais de uma;
- e da quantidade de traços latentes que está sendo medida - apenas um ou mais de um.

No decorrer do presente trabalho consider-se-ão somente modelos onde a natureza do item é dicotômica (ou dicotomizavel). Modelos que consideram outros tipos de itens; tanto de de escolha livre quanto de múltipla escolha, que são avaliados de forma graduada, podem ser encontrados em Darrell Bock (1972), F. Samejima (1969), Andrich (1978), Masters (1982), dentre outros.

Com respeito ao número de grupos ou de populações de respondents envolvidas nos estudos, Andrade, Tavares e Valle (2000) comenta como estes dois conceitos estão diretamente ligados entre se e o termo grupo, por sua vez, está diretamente ligado ao processo de amostragem. Portanto quando falarmos em um único grupo de respondents, nós estamos nos referindo à uma amostra de indivíduos retirada de uma mesma população. Assim, dois ou mais grupos de respondentes correspondem à dois ou mais conjuntos de indivíduos que foram amostrados de duas ou mais populações. Pode ser comum encontrar diferenças nas populações de interesse por características próprias dos traços latentes dos indivíduos que as compõem, como a média, variância ou assimetria, e dependendo dos objetivos do estudo pode ou não ser relevante levar em consideração estas diferenças. Por exemplo; o grau de escolaridade, sexo, período (diário ou noturno), ano, tipo de escola. Nestes casos, levar em consideração as possíveis diferenças existentes entre os grupos é mais apropriado para a obtenção de resultados mais confiáveis. Nesse tipo de estudo, o primeiro passo para que os resultados relativos aos vários grupos possam ser comparáveis, através da TRI, é a exigência de itens comuns nos testes aplicados a estes grupos, criando uma estrutura de ligação entre os mesmos. Um procedimento usual neste tipo de situação, é fazer a estimação em cada um dos grupos em separado e, após isso, utilizar uma técnica de equalização, veja Kolen e Brennan (2004) por exemplo.

Uma abordagem alternativa é o modelo para grupos múltiplos introduzido por Bock e Zimowski (1987). Este modelo supõe que as diferenças entre os grupos podem ser traduzidas por diferentes distribuições de probabilidade. Tais distribuições estão associadas ao processo de amostragem dos indivíduos, mas também, podem ser vistas como distribuições à priori para os traços latentes (no contexto bayesiano).

O terceiro fator, que segundo Andrade, Tavares e Valle (2000) caracteriza os modelos da TRI, esta relacionado ao número de traços latentes que o teste esta medindo. No caso, em que se estiver medindo mais de um traço latente, temos os modelos da Teoria de Resposta ao Item Multidimensional (TRIM). Nos modelos da TRIM, para cada individuo em particular é assumido um valor para cada uma das variáveis latentes de interesse. Na literatura existem vários exemplos onde a suposição de unidimensionalidade não se verifica e para os quais sua utilização simplifica, por demais, a natureza dos instrumentos de medida, ver por exemplo: Ackerman (1994), Reckase, Ackerman e Carlson (1988) e mais recentemente M. Reckase (2009) e Montenegro ("Multidimensional item response models"). Isto ocorre pelo fato, da propria estrutura dos itens, a qual permite medir mais de um traço latente. Por exemplo, em testes de Matemática, em geral, os itens exigem uma certa capacidade de compreensão e interpretação de textos (e possivelmente outras competências), além do próprio conhecimento em Matemática. Outro exemplo pode ser a prova de vestibular da Universidade Estadual de Campinas (UNICAMP), exame que objetiva a seleção de estudantes para o ingresso nessa universidade, e que portanto avalia conhecimentos gerais do ensino médio nas áreas

de Matemática, Português, Biologia, Física, Química, Historia e Geografia. Tais aspectos motivam o estudo dos modelos da TRIM.

Desde seus inícios a TRI tem estado fortemente relacionada com a psicometria, área, cujo principal objetivo, como é citado em Fumiko Samejima, 1997, pode ser postulado como "o modelamento matemático do comportamento humano". A modelagem do comportamento humano, pode implicar que muitas das técnicas e ferramentas estatísticas usualmente aplicadas em outras áreas apresentem dificuldades na sua implementação na TRI. Como Bazan (2005) expõe, suposições estatísticas na modelagem de variáveis associadas com a conduta humana estão baseadas na normalidade, tanto dos traços latentes quanto da função de ligação dos modelos. Na estatística o uso da distribuição normal tem sido fundamental no modelamento e análises de dados. Contudo,já foi verificado a necessidade de questionar a validade de alguns destes modelos no âmbito da TRI e portanto, verificou-se a necessidade do desenvolvimento de modelos de resposta ao item que fossem mais flexíveis com respeito aos pressupostos de normalidade. Outras alternativas paramétricas para a distribuição dos traços latentes estão sustentadas nos problemas descritos na literatura, por exemplo, Micceri (1989), o qual enumera várias situações onde as distribuições dos traços latentes parecem violar o pressuposto de normalidade, Outros exemplos podem ser encontrados em Bazan, Branco e Bolfarine, 2006. Outra suposição que pode ser questionada nos modelos da TRIM, diz respeito à função de ligação, que deve ser entendida como a função que explicita a relação entre os traços latentes e os parâmetros dos itens, com as respostas obtidas no teste. Nos modelos mais freqüentes da TRIM, é comum o uso das ligações probito ou logito. Outras alternativas podem ser encontradas na literatura, por exemplo Bazan, Branco e Bolfarine (2006) uma familia de modelos onde a função de enlace é a função de distribuição cumulada da normal assimétrica univariada.

Muitos trabalhos têm proposto distribuições alternativas para os traços latentes. Por exemplo, Mislevy (1986) considera uma estimação de máxima verossimilhança não-paramétrica da distribuição dos traços latentes baseada em métodos de integração por quadratura gaussiana; veja Baker e Kim (2004). Tais métodos foram desenvolvidos sob MVM e estimação bayesiana marginal, para estimação dos parâmetros dos items. Bazan, Branco e Bolfarine (2006) propuseram um MRI com distribuição normal assimétrica para os traços latentes baseada na parametrização direta (PD). Recentemente Santos (2012) e Santos, Azevedo e Bolfarine (2013) propuseram um modelo unidimensional para estrutura de grupos múltiplos, modelando os traços latentes segundo uma distribuição normal assimétrica na parametrização centrada, veja Azzalini (1985).

1.2 Objetivos e Organização da Dissertação

O objetivo principal desta dissertação é o de propor e discutir com detalhes um novo modelo da Teoria de Resposta ao Item Multidimensional para o esquema de um grupo e de grupos múltiplos. Ele pode ser entendido como uma extensão multidimensional dos modelos apresentados em Santos (2012) e Santos, Azevedo e Bolfarine (2013) dentre outros. Os modelos propostos são obtidos pelo uso da distribuição normal assimétrica multivariada centrada para modelar os traços latentes dos indivíduos. A metodologia de estimação bayesiana é desenvolvida para os modelos desenvolvidos. Além disto ilustra-se o uso deste modelo com uma aplicação em um banco de dados reais proveniente de um estudo de avaliação educacional.

No Capítulo 2, é feita uma revisão da distribuição normal multivariada assimétrica, e define-se a parametrização que será usada para o posterior desenvolvimento dos modelos. Também são discutidos alguns aspectos relacionados à estimação.

No Capítulo 3, é feita uma revisão dos modelos da TRIM, para um único grupo e grupos múltiplos. Se introduzem os modelos probito multidimensionais de 3 parâmetros para grupos múltiplos com distribuições normais multivariadas assimétricas centradas para modelar os traços latentes dos diferentes grupos. Também é feita uma discussão sobre o problema de identificabilidade dos modelos.

No Capítulo 4, se apresentam vários algoritmos de dados aumentados para o caso dos modelos da TRIM, incluindo a aplicação da proposta de aceleração de convergência dada por Gonzalez (2004). Um estudo de simulação é realizado para comparar estes algoritmos segundo o critério de comparação introduzido por Sahu (2002). Apresenta-se, também, o algoritmo MCMC para o modelo multidimensional normal assimétrico centrado proposto no Capítulo 3.

No Capítulo 5, se apresentam os resultados de um estudo de simulação, visando avaliar o desempenho dos modelos introduzidos no capítulo 4 com relação a certas características de interesse, como: vício, variância e erro quadrático médio, seguindo Azevedo, Bolfarine e Andrade (2011) e Azevedo, Bolfarine e Andrade (2012). As simulações foram estruturadas considerando-se diversas situações de interesse prático, formadas pela escolha de diferentes distribuições dos traços latentes, número de itens, dimensionalidade do teste e número de itens em comum no caso de grupos múltiplos.

No Capítulo 6, se faz a análise de um conjunto de dados reais, referente à primeira fase do vestibular da UNICAMP do ano 2013.

No Capítulo 7, são apresentadas diversas conclusões obtidas deste trabalho de pesquisa e sugestões para pesquisas futuras.

Todo o trabalho computacional feito nesta dissertação, foi desenvolvido sobre o programa livre ***R***, R Core Team (2012a).

Capítulo 2

Distribuição Normal Multivariada Assimétrica Centrada

2.1 Introdução

Neste capítulo apresentamos uma revisão sobre a distribuição normal assimétrica multivariada e definimos uma nova parametrização da distribuição normal multivariada assimétrica centrada (NMAC). As vantagens da parametrização aqui proposta no campo da TRI referem-se à identificabilidade dos modelos e interpretação dos parâmetros, quando esta distribuição é usada para modelar os traços latentes dos indivíduos. Propriedades importantes acerca referida parametrização sao apresentadas.

2.2 Distribuição normal univariada assimétrica centrada

Nas últimas décadas tem aumentado o interesse nas chamadas distribuições assimétricas elípticas cujas principais características estão resumidas em Genton (2004). Um dos membros mais importantes e utilizados dessa família, é a distribuição normal assimétrica. Embora a idéia de modelar assimetria mediante o desenvolvimento de famílias que contivessem a distribuição normal já tivesse sido proposta por outros autores, foi Azzalini, no seu pioneiro artigo de 1985, quem forneceu a definição formal da distribuição normal assimétrica para o caso univariado. No referido artigo, é definido que uma variável aleatória Z possui uma distribuição Normal Assimétrica (NA) com parâmetro de assimetria λ, empregando a notação $Z \sim NA(0, 1, \lambda)$, se sua densidade é dada por

$$f(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \quad \lambda \in \mathbb{R}, \tag{2.2.1}$$

onde como é usual, $\phi(\cdot)$ e $\Phi(\cdot)$ denotam respectivamente a função de densidade de probabilidade (fdp) e a função de distribuição cumulativa (fdc) da normal padrão univariada simétrica. Azzalini (1985) e Henze (1986) também estudaram algumas das propriedades desta classe de distribuições, obtendo assim uma representação estocástica baseada em variáveis aleatórias independentes, dada

por

$$Z \stackrel{d}{=} \frac{\lambda}{\sqrt{1+\lambda^2}}|X| + \frac{1}{\sqrt{1+\lambda^2}}Y, \tag{2.2.2}$$

onde o símbolo "$\stackrel{d}{=}$" deve ser lido como "distribui-se como". As variáveis X e Y são variáveis aleatórias independentes e identicamente distribuídas $N(0,1)$. Usando uma reparametrização, a forma vista em (2.2.3) pode ser reescrita como

$$Z \stackrel{d}{=} \delta|X| + (1-\delta^2)^{\frac{1}{2}}Y, \text{ onde } \delta = \frac{\lambda}{(1+\lambda^2)^{\frac{1}{2}}}. \tag{2.2.3}$$

Algumas propriedades desta variável aleatória, se resumem no seguinte resultado.

Proposição 2.2.1. Seja $Z \sim NA(0,1,\lambda)$. Então,

- $Z^2 \sim \chi^2_1$
- $E(Z) = \sqrt{\frac{2}{\pi}}\delta$
- $Var(Z) = 1 - \frac{2}{\pi}\delta^2$.

Para as demonstrações destas e de outras propriedades da NA univariada, sugerimos a leitura de Azzalini (1985) e Arellano-Valle e Genton (2005). Graças ao uso da propriedade da linearidade, a qual se apresenta na Proposição (2.2.2), é possível considerar parâmetros de localização e escala à NA.

Proposição 2.2.2. Se $Z \sim NA(0,1,\lambda)$ e sejam $\epsilon \in \mathbb{R}$ e $\omega \in \mathbb{R}^+$, então a densidade da variável aleatória Y definida mediante a transformação linear

$$Y = \xi + \omega Z, \tag{2.2.4}$$

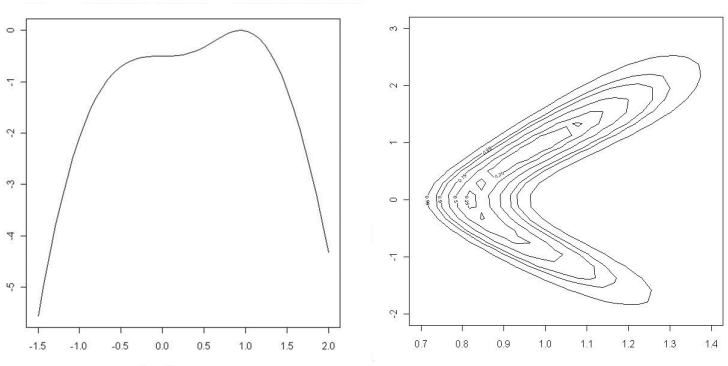
vai ser

$$f_Y(y;\xi,\omega^2,\lambda) = 2\omega^{-1}\phi\Big(\frac{y-\xi}{\omega}\Big)\Phi\Big[\lambda\Big(\frac{y-\xi}{\omega}\Big)\Big], \tag{2.2.5}$$

acrescentando, desta forma, na distribuição (2.2.5), parâmetros de localização ξ, com domínio nos reais, e um parâmetro de escala ω com domínio nos reais positivos.

A distribuição apresentada em (2.2.5), embora com algumas propriedades desejáveis, possui certas características que dificultam seu uso em termos inferenciais, interpretação dos parâmetros do modelo e estimação de parâmetros. Mais especificamente, como mencionado em Azzalini e Capitanio (1999), esta parametrização possui um ponto de singularidade na matriz de informação em $\lambda = 0$, fato que faz com que o modelo viole os pressupostos básicos da normalidade assintótica dos estimadores via maxima verossimilhança EMV.

A Figura 2.1a exibe uma forma não quadrática da log-verossimilhança gerada pela existência de um ponto de inflexão em $\lambda = 0$. Já na Figura 2.1b vemos uma forma bastante peculiar dos contornos. Isso indica que pode existir um ponto de sela na superfície de log-verossimilhança de



(a) Duas vezes a LVPR de λ (b) Contornos de duas vezes a LVPR de (ψ, λ)

Figura 2.1: Contorno Normal Assimétrica diferentes valores de α

$(\omega; \lambda)$. Dado que os parâmetros da equação (2.2.5) são identificáveis, os problemas inferenciais no uso desta densidade são devidos à parametrização e particularmente, ao fato de esta não ser adequada para problemas de estimação, o que motivou à proposta de uma parametrização alterna, ver Azzalini (1985).

Azzalini (1985) introduziu a parametrização centrada definida a partir da seguinte identidade

$$Y_c = \xi + \omega Z = Z = \mu + \sigma Z_0, \tag{2.2.6}$$

onde $Z \sim NA(0, 1, \lambda)$, $Z_0 = \sigma_z^{-1}(Z - \mu_z)$ e

$$\mu_z = E(Z) = b\delta, \qquad \sigma_z^2 = Var(Z) = 1 - b^2\delta^2,$$

em que $b = \sqrt{\frac{2}{\pi}}$; com μ e σ definidas de forma tal que se satisfaça (2.2.6). A parametrização alternativa, ou parametrização centrada (PC), é então formada por (μ, σ^2, γ) já que estes são construídos via a variável centrada Z_0, e por sua vez, os parâmetros (ξ, ω^2, λ) são conhecidos como "parâmetros diretos" (PD). A relação entre estas duas parametrizações é explicitada na equação (2.2.7),

$$\begin{aligned} \mu = E(Y_c) &= \xi + \omega\mu_z, \\ \sigma^2 = Var(Y_c) &= \omega^2(1 - \mu_z^2), \\ \gamma = \frac{E[(Y_c - E(Y_c))^3]}{Var(Y_c)^{3/2}} &= \frac{4-\pi}{2}\frac{\mu_z^3}{(1-\mu_z^2)^{3/2}}, \end{aligned} \tag{2.2.7}$$

em que γ é o coeficiente de assimetria de Pearson. Também é verdade, ver Arellano-Valle e Azzalini (2008), que

$$\mu_z = \frac{c}{\sqrt{1+c^2}}, \text{ onde } c = \left(\frac{2\gamma}{4-\pi}\right)^{1/3}. \tag{2.2.8}$$

A parametrização centrada apresenta vários benefícios. Como mencionado em Azzalini e Capitanio (1999), a PC apresenta as seguintes características: remoção da singularidade da matriz de informação em $\lambda = 0$, as estimativas dos parâmetros na PC são menos correlacionadas que as da PD e a forma da verossimilhança é mais apropriada em teremos inferenciaos. Em Azzalini (1985), Azzalini e Capitanio (1999), Arellano-Valle e Azzalini (2008), Santos, Azevedo e Bolfarine (2013), Azevedo, Bolfarine e Andrade (2011), se podem encontrar exemplos do porque a parametrização centrada representa uma alternativa mas conveniente para escrever a função de densidade.

Na Figura 2.2 apresentam-se algumas densidades para a PC, considernado diferentes valores de λ. Se pode notar que a transformação que conduz à PC preserva o comportamento da densidade NA. O parâmetro γ tem uma interpretação imediata, no sentido de que quanto mais próximo for a -1 ou 1, maior será a assimetria (positiva ou negativa) da variável. Seguindo a mesma linha, é considerado geralmente, que se $\gamma \in (-0,13; 0,13)$ a assimetria não é significante.

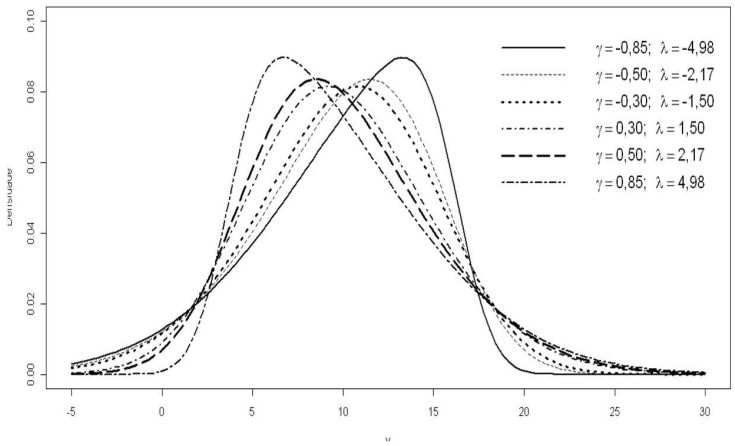


Figura 2.2: Densidades da NA para diferentes valores de λ

2.3 Distribuição Normal Multivariada Assimétrica

Nesta seção primeiramente fornecemos as ferramentas necessárias para a introdução da distribuição NMAC, para posteriormente fazer sua apresentação formal. Comentam-se as diferenças entre a NMAC proposta neste trabalho com aquela proposta em Arellano-Valle e Azzalini (2008). Também, apresentam-se vantagens do uso da distribuição NMAC no âmbito da TRIM.

2.3.1 Parametrização direta

A importância da densidade (2.2.1) pode ser explicada por dois fatores a saber: primeiro, ela possui certas propriedades semelhantes àquelas presentes na distribuição normal simétrica, acrescentado um parâmetro para modelar a assimetria dos dados e desta forma permitem justificar seu nome de normal-assimétrica; segundo, de um ponto de vista prático, permitem modelar apropriadamente dados com características de unimodalidade e de assimetria, situação que como

hill1982robustness comenta são bastantes comuns na prática. Assim a extensão da densidade (2.2.1) ao caso multivariado é importante para analisar dados multivariados que apresentem estas características.

Versões multivariadas da densidade (2.2.1) foram brevemente discutidas em Azzalini (1985), e posteriormente formalizadas em Azzalini e Valle (1996) e Azzalini e Capitanio (1999). Assim uma das varias formas da distribuição normal multivariada assimétrica é a seguinte: uma variável aleatória d-dimensional tem uma distribuição normal assimétrica multivariada se sua seguinte função de densidade é dada por

$$2\phi_d(\boldsymbol{z};\boldsymbol{\Omega})\Phi(\boldsymbol{\lambda}^t z), \quad z \in \mathbb{R}^d, \tag{2.3.1}$$

onde $\phi_d(\boldsymbol{z};\boldsymbol{\Omega})$ denota a função de densidade de uma variável aleatória normal d-dimensional com vetor de medias $\mathbf{0}$ e matriz de correlações $\boldsymbol{\Omega}$. Usar-se-á a notação $\boldsymbol{Z} \sim NA_{d,1}(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ para indicar que o vetor aleatório $\boldsymbol{Z}$ distribui-se conforme uma normal d-variada assimétrica com função de densidade dada por (2.3.1).

De forma análoga ao caso simétrico, quando $\boldsymbol{\lambda} = \mathbf{0}$, a densidade (2.3.1) se reduz à normal simétrica d-variada de parâmetros $(\mathbf{0}, \boldsymbol{\Omega})$. Azzalini e Capitanio (1999) se referem ao vetor $\boldsymbol{\lambda}$ como "parâmetro de forma", embora o verdadeiro parâmetro de forma tenha uma estrutura mais complexa. Esta distribuição permite a introdução de parâmetros de localização e escala, escrevendo

$$\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{Z}, \tag{2.3.2}$$

onde $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)^t$ é um vetor de parâmetros de localização e $\boldsymbol{\omega} = \text{diag}(\omega_1, \ldots, \omega_d)$ é uma matriz diagonal de parâmetros de escala com suas componentes necessariamente positivas. Assim, se obtém o seguinte resultado,

Proposição 2.3.1. Seja $\boldsymbol{Z} \sim NA_d(\mathbf{0}, \boldsymbol{\Omega}; \boldsymbol{\lambda})$ então a função de densidade do vetor aleatório $\boldsymbol{Y}$ definido segundo a transformação linear

$$\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{Z},$$

é dada por

$$2\phi_k(\boldsymbol{Y} - \boldsymbol{\xi}; \boldsymbol{\Omega}_y)\Phi\{\boldsymbol{\lambda}^t\omega^{-1}(\boldsymbol{Y} - \boldsymbol{\xi})\} \tag{2.3.3}$$

onde $\boldsymbol{\Omega}_y = \boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}$ é uma matriz de variâncias e covariâncias, e $\boldsymbol{\omega}$ é uma matriz diagonal formada pelos elementos diagonais da matriz $\boldsymbol{\Omega}$ elevados à $\frac{1}{2}$. Para indicar que o vetor aleatório possui a densidade (2.3.3) é usada a notação $\boldsymbol{Y} \sim NA_{d,1}(\boldsymbol{\xi}, \boldsymbol{\Omega}_y, \boldsymbol{\lambda})$. A prova da proposição acima é obtida pela aplicação do método do Jacobiano sobre a densidade do vetor aleatório $\boldsymbol{Z}$.

Na literatura se podem achar outras propostas para expressar a densidade de um vetor aleatório com distribuição NA k-variada considerando parâmetros de localização e escala. Por exemplo, Lachos (2004) propõe uma densidade alternativa àquela da expressão (2.3.1). Na referida definição, um vetor aleatório d-dimensional $\boldsymbol{Y}$ segue uma distribuição normal assimétrica com vetor de

localização $\boldsymbol{\mu}$, matriz de dispersão $\boldsymbol{\Sigma}$ (uma matriz definida positiva de dimensão $d \times d$) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^n$; se sua fdp é dada por

$$f_Y(\boldsymbol{y}) = 2\phi_k(\boldsymbol{Y} - \boldsymbol{\mu}; \Sigma)\Phi(\boldsymbol{\lambda}^t \Sigma^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu})),\ \boldsymbol{Y} \in \mathbb{R}^n. \tag{2.3.4}$$

Para diferenciar um vetor aleatório com função de densidade dada em (2.3.4) com aquela da expressão (2.3.3), se implementará a notação, $\boldsymbol{Y} \sim NA_{d,2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ para indicar que $\boldsymbol{Y}$ possui a densidade (2.3.4). Continuando, Lachos (2004) faz uma observação de como podem ser consideradas outras reparametrizações para representar o parâmetro de assimetria $\boldsymbol{\lambda}$ como, por exemplo

$$\boldsymbol{\lambda} = \frac{\boldsymbol{\Delta}^{-1/2}\boldsymbol{\delta}}{\sqrt{1 - \boldsymbol{\delta}^t \boldsymbol{\Delta}^{-1}\boldsymbol{\delta}}}, \tag{2.3.5}$$

para algum $\boldsymbol{\delta} \in \mathbb{R}^n$ e matriz definida positiva $\boldsymbol{\Delta}$ de dimensão $n \times n$ tal que $\sqrt{\boldsymbol{\delta}^t \boldsymbol{\Delta}^{-1}\boldsymbol{\delta}} < 1$, Lachos (2004) também faz o comentário de que quando $\boldsymbol{\Delta} = \Sigma$, (matriz de correlações) se obtém a parametrização usada em Azzalini e Valle (1996), i.e., aquela da expressão (2.3.5). A diferença entre a distribuição normal em (2.3.4) e aquela introduzida por Azzalini e Valle (1996) e Azzalini e Capitanio (1999), é que estes autores consideram $\boldsymbol{\omega} = \sqrt{\boldsymbol{\Omega}}$ no lugar de $\boldsymbol{\Sigma}$ na parte assimétrica da densidade. A notação $\sqrt{\boldsymbol{\Omega}}$, é usada para representar uma matriz diagonal, cujos elementos são a raiz quadrada dos elementos da diagonal principal de $\boldsymbol{\Omega}$.

Comparando as formas das densidades (2.3.4) e (2.3.3) vemos que no caso de $\boldsymbol{\mu} = \boldsymbol{\xi} = \mathbf{0}$ e $\boldsymbol{\Omega} = \boldsymbol{\Sigma} = \boldsymbol{I}_d$ as duas coincidem. Porem é importante notar que para a definição introduzida por Lachos (2004) se cumpre uma propriedade que não necessariamente é valida para aquela introduzida por Azzalini e Valle (1996), a qual corresponde à versão multivariada da representação estocástica (2.2.2) que se formaliza no seguinte resultado

Proposição 2.3.2. Seja $\boldsymbol{Z} \sim NA_{d,2}(\mathbf{0}, \boldsymbol{I}_d; \boldsymbol{\lambda})$. Então,

$$\boldsymbol{Z} \overset{d}{=} \boldsymbol{\delta}|X_0| + (\boldsymbol{I}_d - \boldsymbol{\delta}\boldsymbol{\delta}^t)^{1/2}\boldsymbol{X}_1, \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^t\boldsymbol{\lambda})^{1/2}}, \tag{2.3.6}$$

$X_0 \sim N(0,1)$, $\boldsymbol{X}_1 \sim N_d(\mathbf{0}, \boldsymbol{I}_d)$ e são independentes.

Algumas das propriedades da distribuição normal assimétrica d-variada ($NA_{k,2}(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$) se resumem no seguinte resultado, que pode ser considerado como o análogo multivariado da proposição (2.2.1). Para as demonstrações correspondentes, sugerimos a leitura de Azzalini e Valle (1996) e Azzalini e Capitanio (1999).

Proposição 2.3.3. Seja $\boldsymbol{Y} \sim NA_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Então,

1. $\boldsymbol{\mu}_y = E(\boldsymbol{Y}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}$, com $\boldsymbol{\delta}$, onde $\boldsymbol{\delta}$ é tal que $\boldsymbol{\lambda} = \frac{\Delta^{-1/2}\boldsymbol{\delta}}{\sqrt{1-\boldsymbol{\delta}^t\Delta\boldsymbol{\delta}}}$,
2. $Cov(\boldsymbol{Y}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^t\boldsymbol{\Sigma}^{1/2}$.

A importância dos resultados (2.3.2) e (2.3.6), radica que em que eles ajudarão no posterior desenvolvimento e implementação dos algoritmos de estimação nos modelos propostos da TRIM.

Outra propriedade de interesse, diz respeito à distribuição de uma transformação linear de um vetor aleatório normal assimétrico, que como Lachos (2004) enfatiza é fechada usando uma matriz de posto completo.

Proposição 2.3.4. Seja $\boldsymbol{Y} \sim NA_{d,2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $\boldsymbol{C}$ uma matriz $d \times q$ de posto completo. Então,

$$\boldsymbol{C}^t\boldsymbol{Y} \sim NA_{q,2}(\boldsymbol{C}^t\boldsymbol{\mu}, \boldsymbol{C}^t\boldsymbol{\Sigma}\boldsymbol{C}, \boldsymbol{\lambda}^*),$$

onde $\boldsymbol{\lambda}^* = \boldsymbol{\delta}_*/(1 - \boldsymbol{\delta}_*{}^t\boldsymbol{\delta}_*)^{1/2}$, com $\boldsymbol{\delta}_* = (\boldsymbol{C}^t\boldsymbol{\Sigma}\boldsymbol{C})^{-1/2}\boldsymbol{C}^t\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. Para a demonstração deste resultado recomenda-se ver Lachos (2004).

Uma característica de interesse da distribuição (2.3.3), está relacionada com suas distribuições marginais, como descrito por Azzalini e Valle (1996) e Azzalini e Capitanio (1999), se $\boldsymbol{Z} \sim NA_{d,1}(\boldsymbol{0}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$, então a distribuição de um subconjunto de componentes de $\boldsymbol{Z}$ ainda é normal assimétrica. A proposição que se encontra continuação pode-se encontrar em Azzalini e Capitanio (1999) proposição (2) junto com sua demonstração.

Proposição 2.3.5. Suponha que $\boldsymbol{Z} \sim NA_{k,1}(\boldsymbol{0}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ e $\boldsymbol{Z}$ é particionado como $\boldsymbol{Z}^t = (\boldsymbol{Z}_1^t, \boldsymbol{Z}_2^t)$ de dimensões h e $k-h$ respectivamente; denotando

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \qquad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix},$$

as correspondentes partições de $\boldsymbol{\Omega}$ e $\boldsymbol{\lambda}$. Então a distribuição marginal de $\boldsymbol{Z}_1$ é $NA_{h,1}(\boldsymbol{0}, \boldsymbol{\Omega}_{11}, \tilde{\lambda}_1)$, onde

$$\tilde{\lambda}_1 = \frac{\lambda_1 + \Omega_{11}^{-1}\Omega_{12}\lambda_2}{(1 + \lambda_2^t\Omega_{22\cdot1}\lambda_2)^{1/2}}, \qquad \Omega_{22\cdot1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}. \tag{2.3.7}$$

2.3.2 Parametrização centrada multivariada de Arellano-Valle e Azzalini (2008)

Devido à crescente utilização das distribuições normais multivariadas com assimetria, e à dificuldade na manipulação das suas densidades, também surgiram novas reparametrizações. Por exemplo, Arellano-Valle e Azzalini (2008) desenvolveram a versão multivariada da parametrização centrada, para definir a formulação segue-se o esquema de Azzalini e Capitanio (1999),

$$\bar{\Omega} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}, \qquad \boldsymbol{\delta} = (1 + \boldsymbol{\lambda}^t\bar{\Omega}\boldsymbol{\lambda})^{-1/2}\bar{\Omega}\boldsymbol{\lambda}$$

e a variável
$\boldsymbol{Z} = \boldsymbol{\omega}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi}) \sim NA_{d,1}(\boldsymbol{0}; \bar{\Omega}, \boldsymbol{\lambda})$, tal que

$$E(\boldsymbol{Z}) = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta} = \boldsymbol{\mu}_z, \qquad Cov(\boldsymbol{Z}) = \bar{\Omega} - \boldsymbol{\mu}_z\boldsymbol{\mu}_z^t = \bar{\Omega} - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^t = \boldsymbol{\Sigma}_z, \tag{2.3.8}$$

com sua versão padronizada $\boldsymbol{Z}_0 = \sigma_z^{-1}(\boldsymbol{Z} - \mu_z)$, onde $\sigma_z = diag(\sigma_{z1}, \ldots, \sigma_{zd})$ é uma matriz diagonal, cujos termos não nulos são os desvios padrão da matriz $\boldsymbol{\Sigma}_z$ de forma tal que o j-ésimo elemento é $\sigma_{zj} = (1 - b^2\delta_{zj}^2)^{1/2}$, em analogia ao caso escalar, esta parametrização é obtida mediante o uso da variável centrada $\boldsymbol{Z}_0$ e dai o seu nome de parametrização centrada, cujos parâmetros são $(\boldsymbol{\mu}_{pc}, \boldsymbol{\Sigma}_{pc}, \boldsymbol{\gamma})$, onde

$$\boldsymbol{\mu}_{pc} = E(\boldsymbol{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\mu}_z, \qquad \boldsymbol{\Sigma}_{pc} = Cov(\boldsymbol{Y}) = \boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\mu}_z\boldsymbol{\mu}_z^t\boldsymbol{\omega}. \tag{2.3.9}$$

Se devem ressaltar dois aspectos importantes da parametrização centrada de Azzalini e Capitanio (1999) ou de Arellano-Valle e Azzalini (2008): (1) Esta é feita sobre a densidade da NA d-variada da expressão (2.3.3), e (2) ela, em algumas situações, incorpora os parâmetros de localização e escala.

2.3.3 Introdução da distribuição NAC_d

Procedemos agora, com a introdução da parametrização da NMA que será utilizada para a modelagem dos traços latentes ($\boldsymbol{\theta}$). Seja $\boldsymbol{Z} \sim NA_{d,2}(\boldsymbol{0}, \boldsymbol{I}_d; \boldsymbol{\lambda})$, da propriedade(2.3.2) sabemos que $Cov(\boldsymbol{Z}) = \boldsymbol{I}_d - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^t$ e $E(\boldsymbol{Z}) = \sqrt{2\pi}\boldsymbol{\delta}$, chamemos estas quantidades de Ψ_z e μ_z respectivamente. Agora consideremos a variável aleatória

$$\boldsymbol{\theta} = (chol(\boldsymbol{\Psi}))^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}_z), \tag{2.3.10}$$

é fácil ver que esta variável terá como parâmetros de localização o vetor $\boldsymbol{0}$, e de escala a matriz identidade de dimensiones apropriadas. Dizemos que o vetor aleatório $\boldsymbol{\theta}$ tem distribuição normal assimétrica d-variada centrada já que esta definida mediante vetores aleatórios centralizados. Utilizar-se-á a notação $\boldsymbol{\theta} \sim NAC_d(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{\lambda})$. Deve-se destacar que a diferença entre a parametrização centrada multivariada aqui apresentada, com aquela de Azzalini e Capitanio (1999), se encontra na densidade considerada para suas construções.

O vector aleatório d-dimensional $\boldsymbol{\theta}$ é obtido mediante uma transformação linear do tipo $\boldsymbol{A}\boldsymbol{Z}+\boldsymbol{B}$ onde $\boldsymbol{A} = \Psi_z^{-1/2}$ e $\boldsymbol{B} = -\sqrt{\frac{2}{\pi}}\boldsymbol{\Psi}_z^{-1/2}\boldsymbol{\delta}$, portanto usando proposição (2.3.1) se obtém que a distribuição de $\boldsymbol{\theta}$ na parametrização introduzida por Lachos (2004) vai ser $NA_{d,2}(-\sqrt{\frac{2}{\pi}}\Psi_z^{-1/2}\boldsymbol{\delta}; \Psi_z^{-1}; \boldsymbol{\lambda})$, este fato é importante já que fornece uma relação entre a parametrização direta e a parametrização centrada aqui usada.

Os primeiros momentos do vetor aleatório d-dimensional $\boldsymbol{\theta}$ são resumidos no seguinte resultado

Proposição 2.3.6. Suponha que $\boldsymbol{\theta} \sim NAC_d(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{\lambda})$. Então

- $E(\boldsymbol{\theta}) = \boldsymbol{0}$
- $Cov(\boldsymbol{\theta}) = \boldsymbol{I}_d$

Demonstração: Como o vetor aleatório $\boldsymbol{\theta}$ pode ser visto como resultado da transformação linear $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{B}$ onde $\boldsymbol{Z} \sim NA_{d,2}(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{\lambda})$, e $\boldsymbol{A} = Cov(\boldsymbol{Z})^{-1/2}$ e $\boldsymbol{B} = -Cov(\boldsymbol{Z})^{-1/2}E(\boldsymbol{Z})$. Então $E(\boldsymbol{Z}_0) = \boldsymbol{A}E(\boldsymbol{Z}) + \boldsymbol{B} = -\boldsymbol{B} + \boldsymbol{B} = \boldsymbol{0}$ e $Cov(\boldsymbol{\theta}) = \boldsymbol{A}Cov(\boldsymbol{Z})\boldsymbol{A}^t = \boldsymbol{I}_d$

□

Outro resultado de interesse obtido com ajuda da representação estocástica vista na expressão (2.3.2) é o seguinte

Proposição 2.3.7. Suponha-se que $\boldsymbol{\theta} \sim NAC_d(\mathbf{0}, \boldsymbol{I}_d, \boldsymbol{\lambda})$, então condicionando o vetor $\boldsymbol{\theta}$ é uma variável aleatória Half-Normal X_0 ($X_0 \sim HN(0,1)$) se vai ter que a distribuição de $\boldsymbol{\theta}|X_0$ vai ser normal simétrica d-variada, em particular

$$\boldsymbol{\theta}|X_0 \sim N_d\left(\boldsymbol{A}\boldsymbol{\delta}X_0, \boldsymbol{A}(\boldsymbol{I}_d - \boldsymbol{\delta}\boldsymbol{\delta}^t)\boldsymbol{A}^t\right), \tag{2.3.11}$$

onde $\boldsymbol{A}$ e $\boldsymbol{B}$ são definidos como antes, isto é, $\boldsymbol{A} = \boldsymbol{\Psi}_z^{-1/2}$ e $\boldsymbol{B} = -\sqrt{\frac{2}{\pi}}\Psi^{-1/2}\boldsymbol{\delta}$.

Um resultado de grande importância, é o (2.3.8), onde se acrescentam parâmetros de localização e escala para a NMAC aqui introduzida. Estes parâmetros são de nosso interesse, porque incorporados em modelos da TRIM, permitem estimas o vetor de médias e a matriz de covariâncias dos traços latentes dos indivíduos.

Proposição 2.3.8. Seja $\boldsymbol{\theta} \sim NAC_d(\mathbf{0}, \boldsymbol{I}_d, \boldsymbol{\lambda})$, e considere-se a transformação

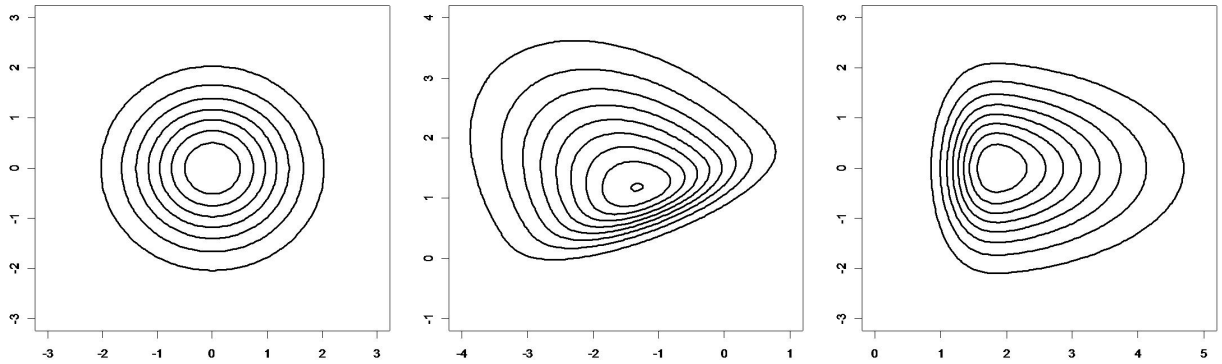
$$\boldsymbol{\theta_1} = \boldsymbol{\mu}_\theta + \boldsymbol{\Psi}_\theta^{\frac{1}{2}}\boldsymbol{\theta}, \tag{2.3.12}$$

então

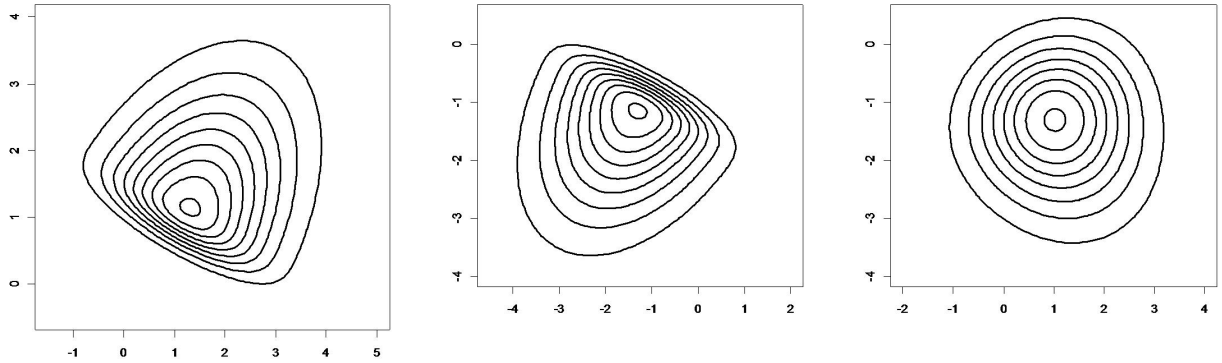
- $E(\boldsymbol{\theta}_1) = \boldsymbol{\mu}_\theta$,
- $Cov(\boldsymbol{\theta}_1) = \boldsymbol{\Psi}_{\boldsymbol{\theta}}$,
- o parâmetro de assimetria de $\boldsymbol{\theta}_1$ será $\boldsymbol{\lambda}_\theta = \boldsymbol{\delta}_{\boldsymbol{\theta}}/(1-\boldsymbol{\delta}_{\boldsymbol{\theta}}{}^t\boldsymbol{\delta}_{\boldsymbol{\theta}})^{1/2}$, com $\boldsymbol{\delta}_{\boldsymbol{\theta}} = (\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{\frac{1}{2}\,t}\boldsymbol{\Psi}_z\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{\frac{1}{2}})^{-1/2}+\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{\frac{1}{2}\,t}\boldsymbol{\Psi}_z^{1/2}\boldsymbol{\delta}$, e $\boldsymbol{\Psi}_z = (\boldsymbol{I}_d - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^t)^{-1}$

O resultado (2.3.7), será importante para o desenvolvimentos dos métodos MCMC para a estimação conjunta dos parâmetros dos items e dos traços latentes $\boldsymbol{\theta}$, no caso em que $\boldsymbol{\theta} \sim NAC_d(\mathbf{0}, \boldsymbol{I}_d; \boldsymbol{\lambda})$. Deve-se notar que para o caso da NMAC aqui introduzida os vetores de localização e escala corresponderão ao vetor de medias e matriz de covariância.

Na Figura 2.3 se apresenta os gráficos de contorno para a distribuição NAC bivariada $NAC_2(\mathbf{0}, \boldsymbol{I}_2, \boldsymbol{\lambda})$ para diferentes valores do parâmetro de assimetria. Pode-se ver como no caso em que $\boldsymbol{\lambda} = (0,0)^t$ o gráfico de contorno corresponde ao de uma distribuição simétrica bivarada.



(a) Contorno NAC$\lambda = (0, 0)$ (b) Contorno NAC$\lambda = (-4, 4)$ (c) Contorno NAC$\lambda = (5, 0)$



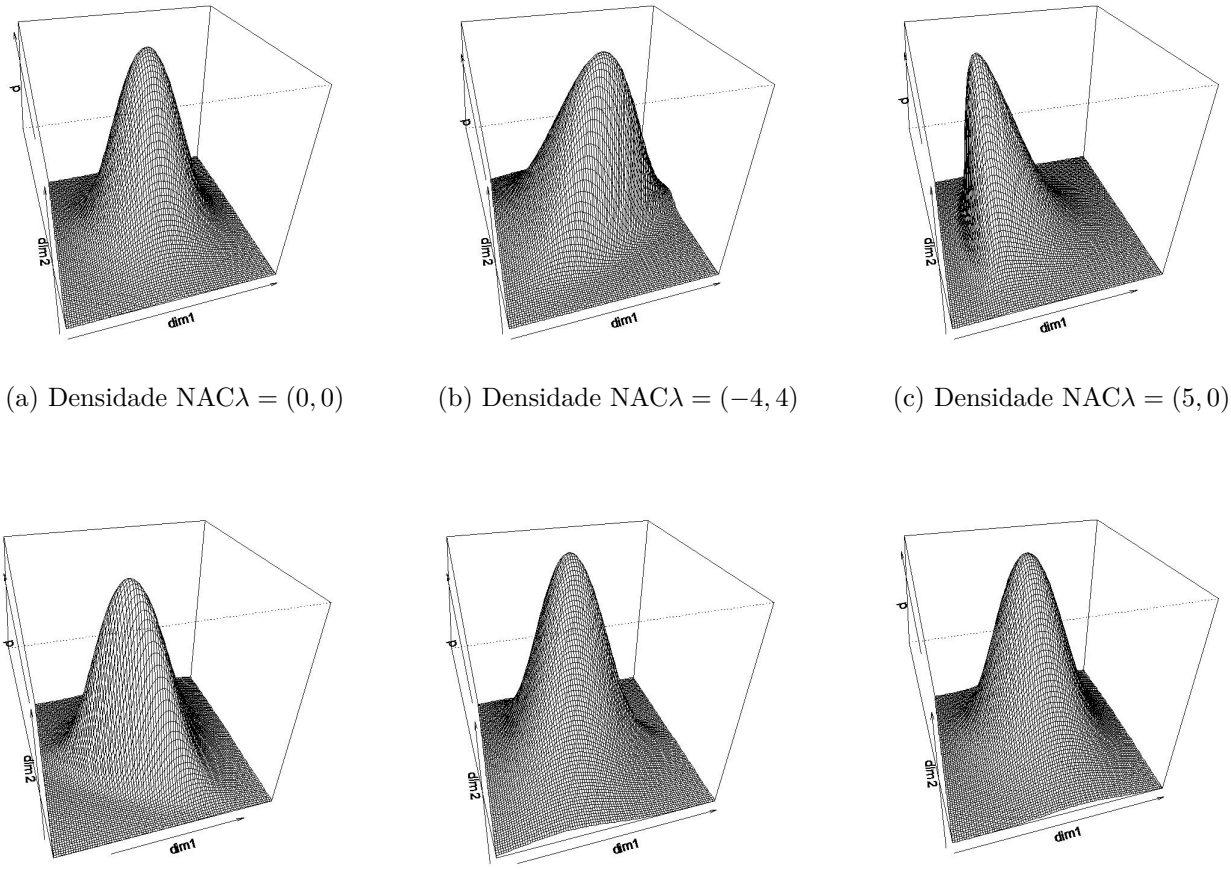
(d) Contorno NAC$\lambda = (5, 5)$ (e) Contorno NAC$\lambda = (-5, -5)$ (f) Contorno NAC$\lambda = (1, -1.3)$

Figura 2.3: Contornos Normal Assimétrica Centrada Multivarada proposta, para diferentes valores de λ

2.4 Métodos de estimação

Nesta seção apresentam-se dois métodos de estimação para a normal multivariada assimétrica, com o intuito de mostrar as vantagens que formas de estimação Bayesiana via MCMC tem em relação a formas de estimação frequentista.

Um dos motivos para a popularidade da distribuição normal assimétrica é o fato dela poder modelar apropriadamente dados reais oriundos de mecanismos especiais de coleta de dados, ver Liseo e Parisi (2013). Porém, uma pesquisa da literatura do tema revela que a maioria dos métodos de inferência para a parametrização usual podem ser problemáticos, ainda para o caso escalar como é explicado em Azzalini (1985), Liseo (1990) e Azzalini e Capitanio (1999). Estes problemas são devidos basicamente às anomalias na verossimilhança: por exemplo, no modelo normal assimétrico univariado padrão, existe uma probabilidade positiva de que o estimador de maxima verossimilhança produzirá valores infinitos; mais especificamente, este fenômeno acontece quando todos os dados compartilham o mesmo sinal. Estas dificuldades tendem a ser ainda maiores no caso



(a) Densidade NAC$\lambda = (0, 0)$ (b) Densidade NAC$\lambda = (-4, 4)$ (c) Densidade NAC$\lambda = (5, 0)$

(d) Densidade NAC$\lambda = (5, 5)$ (e) Densidade NAC$\lambda = (-5, -5)$ (f) Densidade NAC$\lambda = (1, -1.3)$

Figura 2.4: Densidades Normal Assimétrica Centrada diferentes valores de λ

multivariado onde estas situações "problemáticas" não são tao fáceis assim de serem detectadas e muito menos de serem solucionadas.

2.4.1 Estimadores MV

Um resumo sobre metodologias de inferência do ponto de vista frequentista é apresentada aqui, começando inicialmente com o caso univariado da distribuição normal assimétrica. Azzalini e Capitanio (1999) estudaram a estimação clássica via maxima verossimilhança (EMV) nas seções 5 e 6 do seu artigo, e eles notaram que existem vários problemas estatísticos com o procedimento EMV, mesmo para o caso univariado, concluindo que métodos de estimação alternativos deveriam ser explorados.

A parametrização introduzida em Arellano-Valle e Azzalini (2008) evita o problema de singularidade da matriz de informação no ponto $\alpha = 0$, e assim, para os parâmetros centralizados

CP $= (\mu, \sigma, \gamma_1)$ a log-verossimilhança correspondente, no caso univariado é

$$l(\text{CP}) = nlog(\sigma_z/\sigma) - \frac{1}{2}z^t z + \sum \zeta_0(\alpha z_i) \tag{2.4.1}$$

onde

$$z_i = \mu_z + \sigma_z \sigma^{-1}(y_i - x_i^t \beta) = \mu_z + \sigma_z r_i$$

No caso multivariado,

$$l = -\frac{1}{2}nlog|\Omega| - \frac{1}{2}ntr(\Omega^{-1}V) + \sum_i \zeta_0\{\alpha^t \omega^{-1}(y_i - \xi_i)\} \tag{2.4.2}$$

onde

$$V = n^{-1}\sum_i (y_i - \xi_i)(y_i - \xi_i)^t.$$

A manipulação destas verossimilhanças não é trivial, e não existem formas fechadas para os EMV. Além disto os EMV podem ser infinitos em algumas situações. Um exemplo é dado em Liseo e Parisi (2013) onde consideram uma amostra de tamanho 10 de duas dimensões (caso de uma distribuição normal assimétrica bivariada), e todos os parâmetros menos o $\boldsymbol{\lambda}$ conhecidos. A amostra considerada por Liseo e Parisi (2013), é a seguinte

$$\begin{bmatrix} -0.272 & 0.340 & 0.498 & 1.511 & -0.134 & 0.170 & -0.169 & 0.484 & -1.042 & 0.945 \\ 1.421 & 0.668 & 1.610 & -0.610 & 0.577 & -0.168 & 2.222 & -0.606 & 1.789 & 0.361 \end{bmatrix}$$

as estimativas fornecidas pela função msn.mle da livraria sn do **R** são $(\widehat{\lambda}_1, \widehat{\lambda}_2) = (1726754, 1387691)$.

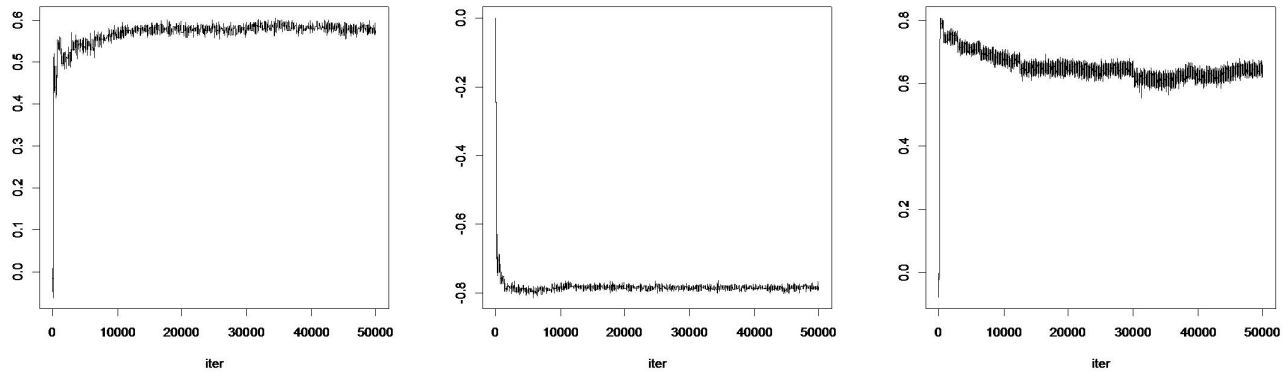
Enquanto no caso escalar o conjunto de amostras que produzem estimativas MV infinitas de $\boldsymbol{\lambda}$ (ou $|\delta| = 1$) podem ser caracterizadas, vide Liseo e Loperfido (2006), no caso multivariado a detecção destes casos não é tao fácil assim. Uma justificativa teórica para o comportamento não satisfatório dos EMV, Liseo e Parisi, 2013, é que a distância de Kullback-Leibler entre duas densidades NA_d com valores próximos de $\boldsymbol{\lambda}$ tende a ser muito pequena. Este fato pode produzir uma verossimilhança perfilada para α que é bastante plana ao longo de um grande porção do espaço paramétrico dele.

2.4.2 Bayesiana

Como é afirmado em Liseo e Parisi, 2013, embora a crescente popularidade de varáveis aleatórias que tomem em conta parâmetros de assimetria, ainda não existe nenhuma solução "satisfatória" para problemas de estimação e de julgamento de hipóteses. A abordagem bayesina destes problemas foi estudada por Liseo e Loperfido (2006) para o caso escalar. No caso multivariado, importantes avanços em formas de estimação bayesiana, podem ser encontrados em Fung e Seneta (2010).

Nesta sub-seção um mecanismo de estimado baseado no algoritmo Metropolis-Hastings é apresentado para estimar todos os parâmetros para um vetor aleatório com distribuição $NAC_d(\mathbf{0}, \boldsymbol{\Psi}_\theta, \boldsymbol{\lambda}_\theta)$ com o intuito de mostrar a eficiência dos métodos MCMC na estimação dos parâmetros, já que com distribuições deste estilo nos capítulos seguintes, se modelará os traços latentes dos itens.

A continuação exibimos um exemplo de dados simulados considerando a matriz $\boldsymbol{\Psi}_\theta$ como uma matriz de correlações 2×2 com elemento fora da diagonal igual a 0,6 e vetor de parâmetros de assimetria igual a $(-0,7; 0,7)^t$. As a prioris consideradas são, para a matriz $\boldsymbol{\Psi}_\theta$ uma a priori Wishart inversa não informativa, e para o vetor $\boldsymbol{\delta}$ foram usadas distribuições uniformes no intervalo $(-0.99527; 0.99527)$. Nas Figuras (2.5) se apresentam os traceplot resultantes da implementação do algoritmo para estimar os parâmetros de uma simulação de 1000 dados. Nas Figuras (2.5) podemos conferir não só a convergência das cadeias para todos os parâmetros, mas também a acurácia das estimativas.



(a) Traceplot para λ_1 quando o valor verdadeiro é 0.7

(b) Traceplot para λ_2 quando o valor verdadeiro é -0.7

(c) Traceplot para $\Psi_{1,2}$ quando o valor verdadeiro é 0.6

Figura 2.5: Traceplot dos parametros da NAC_2

Capítulo 3

Modelos da TRI Multidimensionais Assimétricos para vários grupos e o Problema da Identificabilidade

3.1 Introdução

Neste capitulo, o principal objetivo é apresentar uma nova classe de modelos multidimensionais da TRI para vários grupos com distribuições normais multivariadas centradas assimétricas para modelar os traços latentes. Uma base teórica para o desenvolvimento destes modelos é apresentada. Com este intuito se apresentam as expressões, suposições e demais aspectos necessários para uma análise de dados por meio da Teoria de Resposta ao Item Multidimensional. Se faz uma discussão do problema de identificabilidade dos modelos trabalhados nesta dissertação, apresentando um resultado o qual garante a identificabilidade deles sob certas restrições. Primeiro apresentar-se-ão os resultados no caso de um único grupo para posteriormente apresentar os resultados de grupos múltiplos.

3.2 Introdução à Teoria de Resposta ao Item Multidimensional TRIM

A Teoria de Resposta ao Item Multidimensional (TRIM) consiste em um conjunto de modelos estatísticos que representam a probabilidade de indivíduos obterem certo escore em cada um dos itens que compõem o instrumento de medida (prova, questionário, etc), aos quais os indivíduos são submetidos considerando-se mais de um traço latente, além, os parâmetros dos itens inerentes à estrutura multidimensional, ver M. Reckase (2009). Nos modelos da TRIM para cada indivíduo em particular considera-se valor para cada uma das variáveis latentes de interesse, e o conjunto destas variáveis latente forma o espaço latente, veja Jiang (2005). No caso do modelos da TRI unidimendional, o espaço latente completo consta de uma única variável. Para os modelos da TRIM, é feita a suposição de que dois ou mais traços latentes são necessários para caracterizar o

desempenho de cada indivíduo no instrumento de medida.

Na literatura existem vários exemplos onde a suposição de unidimensionalidade não se verifica e que sua utilização simplifica, por demais, a natureza dos instrumentos de medida, ver por exemplo: Ackerman (1994), Reckase, Ackerman e Carlson (1988) e mais recentemente M. Reckase (2009) e Montenegro(2010). Isto ocorre pelo fato, da propria estrutura dos itens, a qual permite medir (ou exige) mais de um traço latente. Por exemplo, em testes de Matemática, em geral, os itens exigem uma certa capacidade de compreensão e interpretação de textos (e possivelmente outras competências), além do próprio conhecimento em Matemática. Tais aspectos motivam o estudos dos modelos da TRIM.

3.3 Modelos da TRIM para itens dicotômicos

Para descrever os modelos da TRIM, é necessário antes a introdução do conceito de *espaço latente completo*. Em Lord, Novick e Birnbaum (1968) ele é definido como a coleção de todas as variáveis latentes θ_d com $d = 1, 2, \ldots, D$ que discriminam os grupos de examinados, onde D será a quantidade de dimensões de proficiência. Denotaremos assim, com o vetor $\boldsymbol{\theta}$ o espaço latente completo, onde

$$\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_D)^t. \tag{3.3.1}$$

Para os modelos da TRIM é feita a suposição que, mediante funções continuas de probabilidade, se podem relacionar os traços latentes de cada individuo $\boldsymbol{\theta}$ com sua probabilidade de obter um certo escore em um item. O modelo dicotômico da TRIM, seguindo a notação de M. Reckase (2009), pode ser representado na seguinte forma

$$p(Y_{ij} = 1|\boldsymbol{\theta}) = F(\boldsymbol{\theta}, \boldsymbol{\zeta}_j), \tag{3.3.2}$$

para o individuo i no item j e para F uma função de ligação, probito ou logito por exemplo.

Na expressão (3.3.2) nota-se que esta probabilidade depende tanto do vetor de parâmetros incidentais $\boldsymbol{\theta}$ quanto dos parâmetros estruturais do item j, contidos em $\boldsymbol{\zeta}_j$. Estes parâmetros estruturais dependerão dos diferentes modelos considerados.

A grande maioria dos modelos dicotômicos da TRIM fazem o pressuposto de que a probabilidade de se obter uma resposta correta a um item aumenta conforme algum elemento do vetor de habilidades $\boldsymbol{\theta}$ também aumente. Este pressuposto é usualmente conhecido como monotonicidade. Além, se faz a suposição de que, condicionando aos vetores $\boldsymbol{\theta}$ e $\boldsymbol{\zeta}$, o grupo de respondentes vão responder a cada item de forma independente, i.e., a resposta de qualquer individuo a qualquer item não tem nenhum impacto na resposta de outro individuo a um item. Também, e novamente condicionando a $\boldsymbol{\theta}$ e $\boldsymbol{\zeta}$, a resposta de um indivíduo a um item não altera as tendências de resposta desse mesmo indivíduo, em nenhuma forma particular, a outros itens. A resposta de todo indivíduo a qualquer item do teste vai depender só do vetor de habilidades associado a este individuo e ao vetor de parâmetros próprios de item. Estes pressupostos fazem sentido, já que implicam que o processo de construção de testes precisa ser controlado de forma que os indivíduos, para os quais estão dirigidos, não compartilhem informação durante o teste, e que os testes devem ser construídos

de forma tal que a informação em um item não incremente nem diminua as probabilidades de responder corretamente a outro item. Em conjunto, a suposição de respostas independentes a todos os ítens de um teste por todos os indivíduos que o responderam é conhecido como o pressuposto de independência local ou independência condicional, ver M. Reckase (2009).

Suponhamos que temos N indivíduos e que para cada um deles são apresentados I ítens (perguntas) com o objetivo de avaliar o traço latente destes indivíduos em D áreas. A resposta y_{ij} do i-ésimo estudante para o j-ésimo item é medida como 1 para uma resposta correta e 0 para uma resposta incorreta. Os dados a serem tratados estarão representados por uma matriz $\boldsymbol{Y}$ de zeros e uns com entradas y_{ij}.

A independência local ou condicional, embora fora citada como suposição, em outros trabalhos como F.M. Lord (1980), Nojosa (2001), se considera que ela é mais que uma suposição e pode ser vista como uma conseqüência da correta determinação da dimensionalidade do instrumento de medida. Segundo M. Reckase (2009), a dimensionalidade é uma propriedade dos traços latentes dos indivíduos que compõem a amostra e que se submeteram ao teste, e não uma propriedade do teste em si mesmo. Uma definição do conceito de dimensionalidade, na psicometría, pode ser a encontrada em Montenegro (2010), "(dimensionalidade) é a dimensão mínima no espaço de habilidades requerida para se obter independência condicional". A correta especificação da dimensão do teste implica o fato de se poder tratar de forma condicionalmente independente as respostas aos diferentes item do teste, de um mesmo individuo, i.e., a correta especificação da dimensão do teste, implica a independência local.

Na TRIM, podem ser identificados dois topos de modelos, vide M. Reckase (2009). Estes tipos são definidos pela forma em que a informação proveniente do vetor de traços latentes é combinado com as características de cada item. Um dos tipos de modelos é baseado na combinação linear dos traços latentes, esta combinação linear é usada com uma função de ligação, usualmente probito ou logito, para especificar a probabilidade de resposta. Outros exemplos de funções de ligação podem ser encontrados em Bazan, Branco e Bolfarine (2006). Para diferentes valores do vetor de traços latentes, a combinação linear pode ser a mesma. Assim, se algum dos traços latentes for baixo, a soma vai ser a mesma se outra das coordenadas for o suficientemente alta. Esta característica nos modelos tem sido rotulada de compensação, e os modelos com esta propriedade são comumente chamados de modelos compensatórios.

O segundo tipo de modelos separa as partes cognitivas de um teste em partes e usa um modelo unidimensional para cada um deles, tais modelos são conhecidos como não-compensatórios. Os modelos considerados nesta dissertação, pertencem à classe dos compensatórios.

Antigamente, os modelos dicotômicos mais usuais na TRI unidimensional, propostos em Rasch (1960), eram tais que a probabilidade de resposta correta ao item j do individuo i (p_{ij}), é modelada como

$$p_{ij} = F(\theta_i - \beta_j), \quad j = 1, \ldots, J, \quad i = 1, \ldots, N, \tag{3.3.3}$$

onde F é alguma função de ligação, usualmente logito ou probito. Os parâmetros θ_i representam o traço latente do i-ésimo estudante e o parâmetro β_i representa o nível de dificuldade do i-ésimo item.

Os modelos de dois parâmetros na TRI se obtém ao se introduzir um parâmetro de inclinação

$\alpha_j (> 0)$ para cada item. Ou seja,

$$\begin{aligned} p_{ij} &= F(\theta_i \alpha_j - \beta_j), \quad \alpha_j > 0, \quad j = 1, \ldots, J, \quad i = 1, \ldots, N, \\ p_{ij} &= F(\eta_{ij}), \end{aligned} \tag{3.3.4}$$

a restrição $\alpha_j > 0$, $j = 1, \ldots, k$ assegura que o estudante com maior traço latente θ_i obtenha uma maior probabilidade de responder corretamente o item j.

Os modelos de três parâmetros consideram uma assíntota c_j para p_{ij}, neste caso

$$\begin{aligned} p_{ij} &= c_j + (1 - c_j) F(\theta_i \alpha_j - \beta_j), \quad 0 \leq c_j < 1, \quad \alpha_j > 0, \quad j = 1, \ldots, J, \quad i = 1, \ldots, N, \\ p_{ij} &= c_j + (1 - c_j) F(\eta_{ij}), \end{aligned} \tag{3.3.5}$$

Estes são os três modelos fundamentais na TRI unidimensional para ítens dicotômicos e as diferenças ente eles são claras. A probabilidade de uma resposta correta sob o modelo de três parâmetros é explicada por um fator adicional, a probabilidade aproximada de indivíduos com baixo nível de traço latente responderem corretamente ao item.

A extensão multidimensional destes modelos se obtém quando o preditor linear η_{ij} nas expressões (3.3.4)-(3.3.5) é definido como

$$\begin{aligned} \eta_{ij} &= \boldsymbol{\alpha}_j \boldsymbol{\theta}_i^t - \beta_j, \\ &= \sum_{d=1}^{D} \theta_{id} \alpha_{jd} - \beta_j \end{aligned} \tag{3.3.6}$$

em que

- $\boldsymbol{\theta}_j^t = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{Dj})$ é o vetor de habilidades do i-ésimo individuo, $i = 1, 2, \ldots, N$, e onde D é a dimensão do item;
- $\boldsymbol{\zeta}_{j\cdot} = (b_j, \boldsymbol{\alpha}_j^t)^t$ é o vetor de parâmetros do item j;
- $\boldsymbol{\alpha}_j^t = (\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jD})$ é o vetor relacionado à discriminação do item j;
- β_j é o parâmetro relacionado à dificuldade do item j;
- $P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \boldsymbol{\zeta}_j) = p_{ij}$ é a probabilidade do individuo i com vetor proficiência $\boldsymbol{\theta}_i$ responder corretamente o item j, com vetor de parâmetros $\boldsymbol{\zeta}_j$.

Em (3.3.5), se $c_j = 0$ temos a versão multidimensional do modelo de dois parâmetros. Camilli (1994) resumiu os trabalhos de comparação das funções logito e probito. Ele incluiu a prova matemática de que a função probito e a função logito diferem em menos de 0,01 na probabilidade quando a constate 1,702 é incluída no exponente da função logito.

Os modelos de ogiva normal são amplamente utilizados e encontram-se implementados em diversos programas existentes, como o NOHARM (Fraser e McDonald (1988)), que estima os parâmetros $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ por uma adaptação do método dos mínimos quadrados e o TESTFACT (Wilson et al. (1991)) que estima os parâmetros dos items segundo o método da máxima verossimilhança marginal e os parâmetros dos indivíduos por métodos bayesianos de esperança ou moda a posteriori a serem expostos no Capítulo 3.

3.3.1 Parâmetros usuais nos modelos usuais da TRIM

Um interpretação intuitiva dos parâmetros dos modelos pode ser determinada a partir de uma inspeção cuidadosa da Figura 3.2. Os parâmetros $\boldsymbol{\alpha}$ indicam a direção dos contornos equiprováveis. A seguir apresentamos as definições de parâmetros unidimensionais que resumem as informações sobre discriminação e dificuldade. Para mais discussões recomenda-se a leitura do trabalho de Nojosa (2001).

Parâmetro de dificuldade

A definição para a medida de dificuldade de ítens multidimensionais (DIFICM) é baseada no intuito de determinar o ponto no espaço latente em que a superfície de resposta ao item apresente inclinação maxima, ressaltando que a inclinação em um ponto qualquer da superfície é diferente dependendo da direção na qual a inclinação é obtida. O procedimento usado para encontrar o ponto de maior inclinação a partir da origem envolve dois passos. Primeiro, o ponto de máxima inclinação em uma particular direção é determinado. Segundo, as inclinações em cada direção são analisadas para determinar a direção que fornece o máximo absoluto da superfície de resposta.

Em Nojosa (2001) é provado, usando por conveniência o modelo logito multidimensional de 3 parâmetros, que a distância ao ponto de maior inclinação na superfície de resposta ao item é

$$DIFICM_j = \frac{\beta_j}{\sqrt{\sum_{d=1}^{D} \alpha_{jd}^2}}. \tag{3.3.7}$$

Este parâmetro pode ser interpretado da mesma forma que o parâmetro de dificuldade nos modelos unidimensionais.

Parâmetro de discriminação

O vetor de parâmetros $\boldsymbol{\alpha}_j$ indicam a direção dos contornos equiprováveis, ver Figuras 3.2, e a taxa com a que a probabilidade de resposta correta muda de um ponto a outro no espaço latente. Isto pode ser visto tomando a primeira derivada parcial na expressão seguinte

$$p_{ij} = c_j + (1 - c_j)F(\boldsymbol{\alpha}_j \boldsymbol{\theta}_i^t - \beta_j), \quad 0 \leq c_j < 1, \quad \alpha_j > 0, \quad j = 1, \ldots, J, \quad i = 1, \ldots, N, \tag{3.3.8}$$

com respeito àlguma dimensão especifica d. Já que os parâmetros $\boldsymbol{\alpha}$ estão relacionados com a inclinação da superfície e à taxa de mudança da probabilidade com respeito aos eixos, o parâmetro $\boldsymbol{\alpha}$ é usualmente chamado parâmetro de discriminação, vide Reckase e McKinley (1983). Para uma dedução formal do parâmetro de discriminação multidimensional se recomenda ao leitor ver Nojosa (2001). O poder de discriminação, de modo geral, de um item indica quão rápido é a mudança da probabilidade de resposta correta a um item. Itens com valores das discriminações muito grandes vão dividir de forma clara a região espacial em duas partes no caso bidimensional, ver Nojosa (2001).

No caso multidimensional, o parâmetro de discriminação de um item (DISCM), fornece o mesmo tipo de informação que é fornecida pelo parâmetro de discriminação. No caso da TRI

unidimensional, o parâmetro de discriminação está relacionado à inclinação da curva característica do item no ponto onde a inclinação é maxima, ou seja o ponto de inflexão.

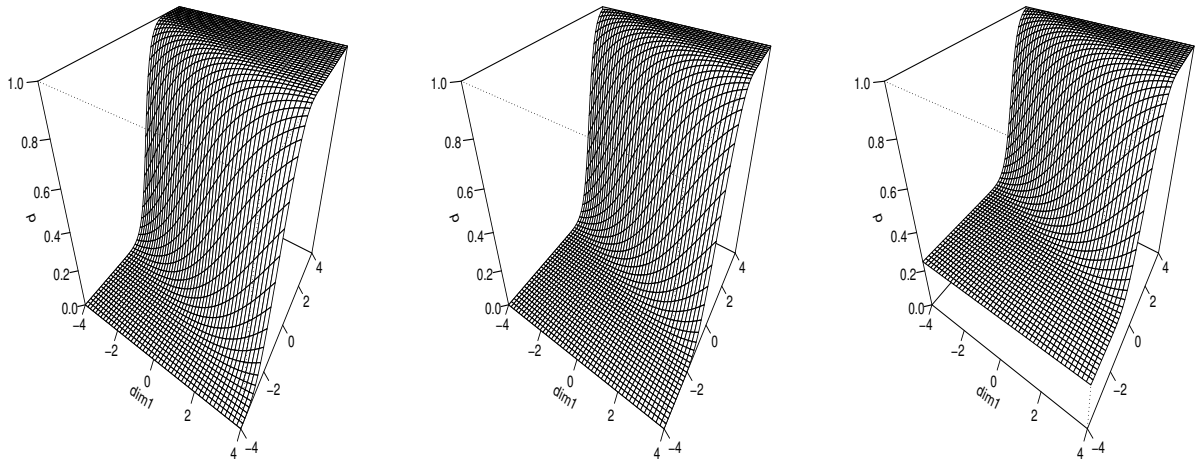
Em Nojosa (2001) é provado que o parâmetro de discriminação para ítens multidimensioais é

$$DISCM_j = \sqrt{\sum_{d=1}^{D} \alpha_{jd}^2}. \tag{3.3.9}$$

Probabilidade de resposta correta de indivíduos com baixo nível do traço latente (Acerto casual)

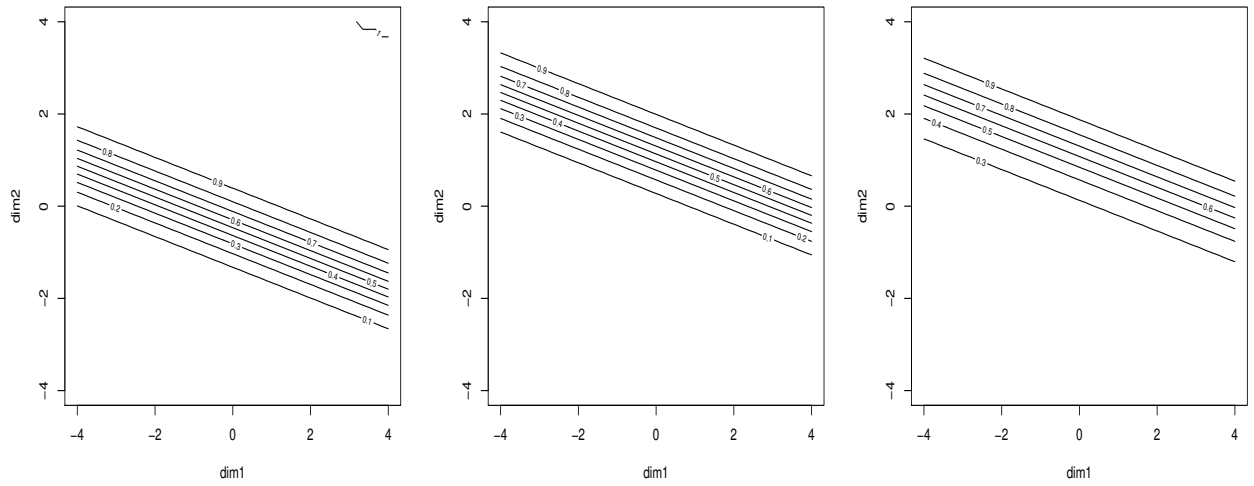
Este parâmetro é usualmente denotado pela letra c e fornece uma (possível) assíntota inferior não nula para a curva característica do item. Ele representa a probabilidade dos indivíduos com baixo traço latente em todas as dimensões de responderem ao item de forma correta. Este parâmetro é incluído dentro do modelo para tomar em conta o desempenho de indivíduos com traços latentes pertencentes às regiões inferiores das escalas, escolher ao acaso a resposta correta ou eliminar distratores "improváveis" são fatores que devem ser tomado em conta.

Nas Figuras 3.2 e 3.4 apresentam-se exemplos de superfícies e contornos de resposta ao item para vários itens com diferentes parâmetros. As duas representações mostram que a probabilidade de resposta correta aumenta monotonicamente com um aumento do valor de e um ou mais traços latentes. Além disto, é claro que os contornos de probabilidades iguais formam linhas retas. Estas linhas mostram a natureza compensatória do modelo. Vemos também o efeito que tem a inclusão do parâmetro c_j, como uma assíntota da superfície.



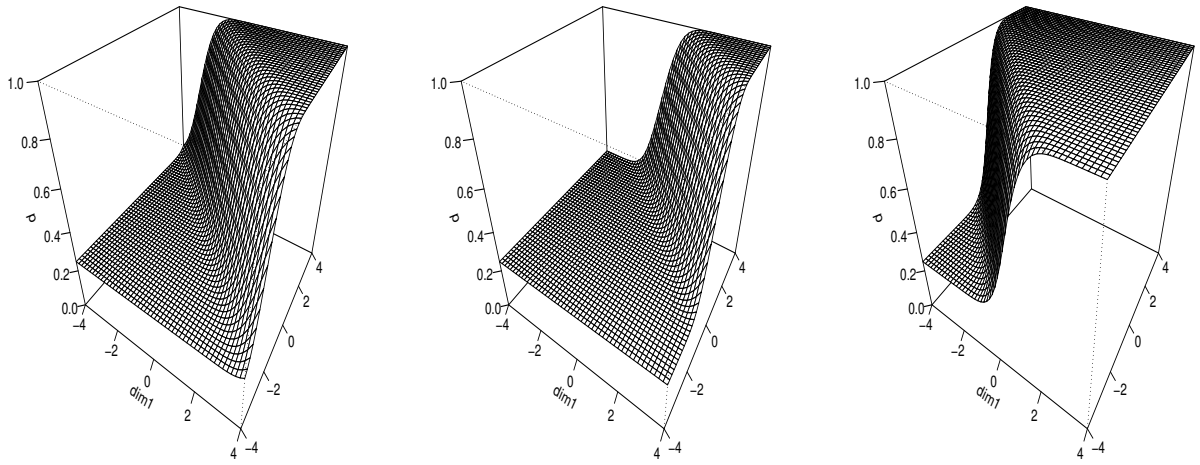
(a) $a_1 = 0.5; a_2 = 1.5; b = -0.7; c = 0$ (b) $a_1 = 0.5; a_2 = 1.5; b = 0.7; c = 0$ (c) $a_1 = 0.5; a_2 = 1.5; b = 0.7; c = 0.25$

Figura 3.1: Superfícies de resposta ao item para itens com diferentes parâmetros



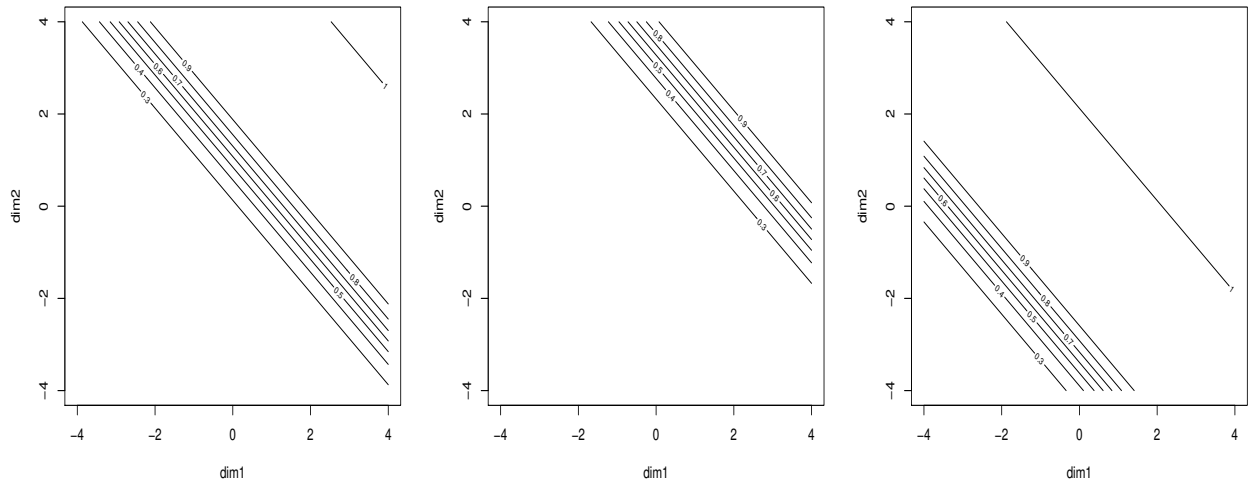
(a) $a_1 = 0.5; a_2 = 1.5; b = -0.7; c = 0$ (b) $a_1 = 0.5; a_2 = 1.5; b = 0.7; c = 0$ (c) $a_1 = 0.5; a_2 = 1.5; b = 0.7; c = 0.25$

Figura 3.2: Contornos para probabilidades de respostas corretas para itens com com diferentes parâmetros



(a) $a_1 = 1.5; a_2 = 1.5; b = 0.7; c =$(b) $a_1 = 1.5; a_2 = 1.5; b = 5; c = 0.25$(c) $a_1 = 1.5; a_2 = 1.5; b = -5; c = 0.25$

Figura 3.3: Superfícies de resposta ao item para itens com diferentes parâmetros



(a) $a_1 = 1.5; a_2 = 1.5; b = 0.7; c =$(b) $a_1 = 1.5; a_2 = 1.5; b = 5; c = 0.25$(c) $a_1 = 1.5; a_2 = 1.5; b = -5; c = 0.25$

Figura 3.4: Contornos para probabilidades de respostas corretas para itens com com diferentes parâmetros (continuação)

3.4 Identificabilidade dos Modelos

No decorrer deste capitulo se apresentou de forma bastante geral alguns dos modelos multidimensionais da TRI. Agora discutiremos alguns problemas referentes à identificabilidade destes modelos. O problema da identificabilidade dos modelos da TRI é inerente à estimação conjunta dos parâmetros dos ítens, e dos traços latentes. No caso em que nem os parâmetros incidentais nem estruturais são conhecidos, se faz necessário o estabelecimento de uma métrica (escala) para que todos os parâmetros possam ser estimados. No caso multidimensional, isto deve ser feito para cada dimensão. O problema conhecido por **falta de identificabilidade do modelo**, vide Matos (2008), induz à necessidade de estabelecimento desta métrica para os parâmetros, com exceção do parâmetro c, a qual inicialmente era arbitrária. A falta de identificabilidade ocorre porque mais de um conjunto de parâmetros pode produzir o mesmo valor da verossimilhança. À seguir, apresentamos uma série de definições que poderão ser de ajuda na compreensão do texto e que podem ser vistas, por exemplo, em, Rivers (2003) e Matos (2008).

Definição 3.4.1. (**Parâmetros Observacionalmente Equivalentes**.) Dois pontos do espaço paramétrico $\boldsymbol{\Upsilon}_1$ e $\boldsymbol{\Upsilon}_2 \in \boldsymbol{\Theta}$ são ditos observacionalmente equivalentes se $p_{\boldsymbol{\Upsilon}_1} = p_{\boldsymbol{\Upsilon}_2}$.

A igualdade $p_{\boldsymbol{\Upsilon}_1} = p_{\boldsymbol{\Upsilon}_2}$ equivale a dizer que dois modelos probabilísticos são ponto a ponto iguais, ou seja $p_{\boldsymbol{\Upsilon}_1}(\boldsymbol{y}) = p_{\boldsymbol{\Upsilon}_2}(\boldsymbol{y})$ para todo valor amostral $\boldsymbol{y}$.

Os modelos probabilísticos referidos na Definição 3.4.1, podem ser, por exemplo, a Função de Resposta ao Item (FRI) multidimensional

Definição 3.4.2. (**Parâmetro e Modelo Globalmente Identificáveis**.) Um ponto do espaço paramétrico $\boldsymbol{\Upsilon}_0 \in \boldsymbol{\Theta}$ é dito identificável ou globalmente identificável se não existe nenhum outro ponto do espaço paramétrico observacionalmente equivalente à $\boldsymbol{\Upsilon}_{0\Theta}$. Neste caso, dizemos que o modelo probabilístico $p_{\boldsymbol{\Upsilon}}$ é globalmente identificável.

Definição 3.4.3. (**Parâmetro e Modelo Localmente Identificável**.) Um ponto do espaço paramétrico $\boldsymbol{\Upsilon}_0 \in \boldsymbol{\Theta}$ é dito ser localmente identificável se existe uma vizinhança $\boldsymbol{V}$ de $\boldsymbol{\Upsilon}_0$ tal que não exista nenhum outro ponto do espaço paramétrico $\boldsymbol{\Upsilon}_0 \in \boldsymbol{\Theta} \cap \boldsymbol{V}$ observacionalmente equivalente à $\boldsymbol{\Upsilon}_0$. Neste caso, dizemos que o modelos probabilístico $p_{\boldsymbol{\Upsilon}}$ é localmente identificável na vizinhança de $\boldsymbol{\Upsilon}_0$.

Continuando com o problema da não-identificabilidade, e seguindo o desenvolvimento de Andrade, Tavares e Valle (2000) vamos ver o caso dos modelos probito unidimensionais, tomando a e b constantes, se $\theta_i^* = a\theta_i + b$, $\beta_j^* = a\beta_j + b$, $\alpha_j^* = \alpha_j/a$ e $c_j^* = c_j$ então

$$\begin{aligned} p(Y_{ij} = 1; \theta_i^*, \alpha_j^*, \beta_j^*, c_j^*) &= c_j^* + (1 - c_j^*)F(\eta_{ij}^*) \\ &= c_j^* + (1 - c_j^*)F(\alpha_j^*\theta_i^* - \beta_j^*) \\ &= c_j + (1 - c_j)F(\frac{\alpha_j}{a}(a\theta_i + b) - \beta_j^* + \frac{\alpha_j b}{a}) \\ &= c_j + (1 - c_j)F(\alpha_j\theta_i - \beta_j) \\ &= c_j + (1 - c_j)F(\eta_{ij}) \\ &= p(Y_{ij} = 1; \theta_i, \alpha_j, \beta_j, c_j), \end{aligned} \tag{3.4.1}$$

com F alguma função de ligação, logito ou probito, por exemplo.

Para diferentes valores dos parâmetros do traço latente, discriminação, dificuldade e de acerto casual, pode-se gerar a mesma probabilidade de acerto ao item j. No caso multidimensional, podem-se distinguir pelo menos dois casos onde diferentes valores dos parâmetros produzem a mesma probabilidade de acerto. Estes dois tipos de não-identificabilidade são devidos à invariância a transformações lineares e rotações ortogonais. Um exemplo do primeiro tipo para o caso de somente uma dimensão é o discutido acima. No caso em que mais dimensões são consideradas, basta considerar uma matriz não-singular $\boldsymbol{A}$ e um vetor de reais $\mathbf{b}$, tais que $\boldsymbol{\theta}_i^* = \boldsymbol{A}\boldsymbol{\theta}_i + \boldsymbol{b}$, e então seguindo o esquema de Matos (2008), ou seja:

$$\begin{aligned}\Phi(\boldsymbol{\alpha}_j^t\boldsymbol{\theta}_i - \beta_j) &= \Phi(\boldsymbol{\alpha}_j^t\boldsymbol{A}^{-1}(\boldsymbol{\theta}_i^* - \boldsymbol{b}) - \beta_j) \\ &= \Phi(\boldsymbol{\alpha}_j^t\boldsymbol{A}^{-1}\boldsymbol{\theta}_i^* - (\boldsymbol{\alpha}_j^t\boldsymbol{A}^{-1}\boldsymbol{b} + \beta_j)) \\ &= \Phi((\boldsymbol{\alpha}_j^*)^t\boldsymbol{\theta}_i^* - \beta_j^*),\end{aligned} \tag{3.4.2}$$

em que $\beta_j^* = \beta_j + \boldsymbol{\alpha}_j^t\boldsymbol{A}^{-1}\boldsymbol{b}$ e $\boldsymbol{\alpha}_j^* = (\boldsymbol{A}^{-1})^t\boldsymbol{\alpha}_j$, e isto é verdade para quaisquer valor de i e j, fato que implica $P(Y_{ij} = 1; \boldsymbol{\theta_i}^*, \boldsymbol{\alpha}_j^*, \beta_j^*, c_i) = P(Y_{ij} = 1; \boldsymbol{\theta_i}, \boldsymbol{\alpha}_j, \beta_j, c_i)$ o que por sua vez implicará que $\boldsymbol{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{y}) = \boldsymbol{L}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*|\boldsymbol{y})$.

Este problema de invariância à transformações lineares é mais conhecido como indeterminação da escala de medida, ver Andrade, Tavares e Valle (2000), Matos (2008), Nojosa (2001), já que é um problema de indeterminação da métrica dos traços latentes dos indivíduos.

A segunda causa de não-identificabilidade dos modelos, por invariância rotacional deve-se à possibilidade de rotação dos traços latentes em torno da origem. Formalmente, se $\boldsymbol{M}$ é uma matriz ortogonal qualquer, tal que $\boldsymbol{M}^t\boldsymbol{M} = \boldsymbol{M}\boldsymbol{M}^t = \boldsymbol{I}_K$ e $\boldsymbol{\theta}_i^* = \boldsymbol{M}\boldsymbol{\theta}_i$,
então

$$\Phi(\boldsymbol{\alpha}_j^t\boldsymbol{\theta}_i - \beta_j) = \Phi(\boldsymbol{\alpha}_j^t\boldsymbol{M}^t\boldsymbol{\theta}_i^* - \beta_j) = \Phi((\boldsymbol{\alpha}_j^*)\boldsymbol{\theta}_i^* - \beta_j) \tag{3.4.3}$$

com $\boldsymbol{\beta}_j^* = \boldsymbol{B}\boldsymbol{M}_j^*$.

Na tentativa de garantir a identificabilidade dos modelos da TRIM, na literatura podem-se encontrar várias propostas. Uma delas é a de Fraser e McDonald (1988) que emprega $\boldsymbol{\mu}_\theta = \mathbf{0}$, $\boldsymbol{\Sigma}_\theta = \boldsymbol{I}$ e $a_{jq} = 0$, para $j = 1, \ldots, k-1$ e $q = j+1, \ldots, k$. Holman et al. (2004) utilizam $\boldsymbol{\mu}_\theta = \mathbf{0}$, $a_{jq} = 0$, para $j \neq q$ e $a_{jq} = 1$, para $j = q$, $j = 1, \ldots, k$ e $q = 1, \ldots, k$, $a_{jq} = 0$, para $j \neq q$ e $a_{jq} = 1$, para $j = q, j = 1, \ldots, k$ e $q = 1, \ldots, k$. Podem-se encontrar outros estudos sobre a falta de identificabilidade nos modelos da TRIM, como por exemplo Rivers (2003) o qual estabelece condições necessárias e suficientes para se obter identificabilidade local, vide Definição 3.4.3, de um modelo D-dimensional de 2 parâmetros. Mais recentemente, em Matos (2008) aborda-se o problema sob pontos de vista da inferência frequentista e bayesiana.

Uma suposição bastante questionada ma TRIM, é o fato das matrizes de covariância dos traços latentes serem fixadas como a matriz identidade. Como já foi comentado, a definição de uma métrica esta intrinsecamente relacionada com a identificabilidade dos modelos. No caso em que se deseje estimar as "relações" entre os traços latentes seria necessário, por exemplo, a estimação da matriz Σ_θ de covariâncias dos traços latentes.

Apresentamos agora um resultado que ajuda a garantir a identificar os modelos da TRIM, perante transformes lineares e rotacionais.

Proposição 3.4.4. Suponha uma matriz ortogonal $\boldsymbol{R}$, diferente da matriz identidade, ou uma matriz permutações, e seja $\boldsymbol{\Gamma}$ uma matriz de correlações. Então o produto $\boldsymbol{\Gamma}^* = \boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^t$ é tal que todos os elementos da diagonal principal não são todos iguais a 1. Isto é, a matriz $\boldsymbol{\Gamma}^*$, não vai ter a estrutura de uma matriz de correlações.

Demonstração. Queremos provar, que para toda matriz ortogonal $\boldsymbol{R}$, diferente da identidade e a uma matriz de permutações, e toda matriz de correlações, $\boldsymbol{\Gamma}$, a matriz $\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^t$ não será uma matriz de correlações. Apresentamos a prova para 2 e 3 dimensões. Para provar que $\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^t$ não preserva as propriedades de uma matriz de correlações, basta provar com que seus elementos da diagonal principal não são todos iguais a um.

- **Caso de matrizes 2×2**:

$$\begin{aligned}
\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^t &= \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} 1 & g \\ g & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix} \right) \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} r_{12}g & r_{11}g \\ r_{22}g & r_{21}g \end{pmatrix} \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} r_{12}g & r_{11}g \\ r_{22}g & r_{21}g \end{pmatrix} \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2r_{11}r_{12}g & r_{12}r_{21}g + r_{11}r_{22}g \\ r_{12}r_{21}g + r_{11}r_{22}g & 2r_{11}r_{12}g \end{pmatrix}.
\end{aligned}$$

Assim, queremos provar que não existem elementos, reais, r_{11}, r_{12} e g para os quais seja verdade que

$$\begin{aligned}
2gr_{12}r_{11} &= 0; \\
2gr_{21}r_{22} &= 0,
\end{aligned} \tag{3.4.4}$$

sob as restrições de:

$$\begin{aligned}
r_{11}^2 r_{12}^2 &= 1; \\
r_{21}^2 r_{22}^2 &= 1; \\
r_{11}r_{21} + r_{12}r_{22} &= 0,
\end{aligned}$$

as anteriores restrições sobre os elementos da matriz $\boldsymbol{R}$ são verdade, pelo fato da matriz $\boldsymbol{R}$ ser ortogonal. Fazendo a prova por absurdo, vamos supor que existem valores r_{11}, r_{12} e g para os quais seja verdade (3.4.4). Mas isto só é possível se $b = 0$ em cujo caso a matriz $\boldsymbol{\Gamma}$ seria a matriz identidade, fato que violaria uma das hipóteses do teorema, ou se algum par de elementos da matriz $\boldsymbol{R}$ são nulos, em cujo caso teríamos que a matriz $\boldsymbol{R}$ é ou a matriz identidade ou uma matriz de permutações, fato que viola outra das hipóteses do teorema. Por redução ao absurdo, temos, que não existem valores de r_{11}, r_{12} e g para os quais seja verdade (3.4.4), o qual implica que a matriz $\boldsymbol{R\Gamma R}^t$ não é uma matriz de correlações.

- **Caso de matrizes 3×3**:

  Para as matrizes

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 1 & g_1 & g_3 \\ g_2 & 1 & g_3 \\ g_3 & g_2 & 1 \end{pmatrix},$$

  os elementos da diagonal principal da matriz, $\boldsymbol{R\Gamma R}^t$ vão ser:

$$\begin{aligned} &1 + r_{12}r_{11}g_1 + r_{11}r_{13}g_2 + r_{12}r_{13}g_3; \\ &1 + r_{21}r_{22}g_1 + r_{21}r_{23}g_2 + r_{23}r_{22}g_3; \\ &1 + r_{31}r_{32}g_1 + r_{31}r_{33}g_2 + r_{32}r_{33}g_3, \end{aligned} \tag{3.4.5}$$

  queremos provar que sob as restrições necessárias para que a matriz $\boldsymbol{R}$ seja uma matriz ortogonal, e tais que $g_i \neq 0$, $\forall i = 1, 2, 3$, não existem valores de r_{ij}, $\forall i, j = 1, 2, 3$ nem g_i, $\forall i = 1, 2, 3$, para os quais seja verdade que

$$\begin{aligned} &r_{12}r_{11}g_1 + r_{11}r_{13}g_2 + r_{12}r_{13}g_3 = 0; \\ &r_{21}r_{22}g_1 + r_{21}r_{23}g_2 + r_{23}r_{22}g_3 = 0; \\ &r_{31}r_{32}g_1 + r_{31}r_{33}g_2 + r_{32}r_{33}g_3 = 0. \end{aligned} \tag{3.4.6}$$

  Mas por um raciocínio análogo ao caso de matrizes 2×2, ter-se-ia que as únicas matrizes ortogonais $\boldsymbol{R}$ que satisfazem (3.4.6), são matrizes de permutações ou a matriz identidade, o qual implicaria uma contradição. Assim provamos este resultado.

□

Proposição 3.4.5. Seja $\boldsymbol{\theta}_j$ o vetor $(D \times 1)$ de traços latentes associado ao individuo j, tal que $E(\boldsymbol{\theta}_j) = \mathbf{0}$ e $Cov(\boldsymbol{\theta}_j) = \boldsymbol{\Sigma}_\theta$, $\forall j, j = 1, \ldots, N$. Suponha, que a matriz $\boldsymbol{\Sigma}_\theta$ tem a estrutura de uma matriz de correlações, com nenhum dos seus elementos, fora da diagonal principal igual a 0. Então o modelo (3.3.8) vai ser identificável.

Demonstração: Para que o modelo (3.3.8) seja identificável, deve-se garantir a invariância ante transformações lineares e rotacionais:

- **Invariância ante transformações lineares:**

  Para este tipo de invariância, basta provar que para uma matriz não-singular $\boldsymbol{A}$ e um vetor de reais $\mathbf{b}$, não é possível fazer a transformação $\boldsymbol{\theta}_i^* = \boldsymbol{A}\boldsymbol{\theta}_i + \boldsymbol{b}$. Isto é verdade, porque para esta transformação ser possível, deve-se ter que $Cov(\boldsymbol{A}\boldsymbol{\Gamma})$ seja uma matriz de correlações, e isto só vai ser verdade no caso em que a matriz $\boldsymbol{A}$ seja ortogonal (ver caso a seguir de invariância rotacional) ou se a matriz $\boldsymbol{A}$ não é de posto completo, mas isto não é possível porque $\boldsymbol{A}$ deve ser não singular, isto é, deve ser de posto completo.

- **Invariância rotacional:**

  A prova para este caso, se obtém de forma direta após a aplicação da Proposição (3.4.4), já que ela diz, que não existe matriz ortogonal $\boldsymbol{R}$, diferente da matriz identidade ou uma matriz de permutações, nem uma matriz de correlações $\boldsymbol{\Gamma}$ com nenhum elemento igual a 0, para a qual a matriz $\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^t$ seja uma matriz de correlações, isto é, a covariância da matriz $\boldsymbol{R}\boldsymbol{\Gamma}$ deve não vai ter a estrutura de uma matriz de correlações.

□

A importância da proposição (3.4.5), se evidencia no fato em que impõe restrições aos traços latentes, definindo uma métrica para eles do tipo $(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$, com $\boldsymbol{\Sigma}_\theta$ uma matriz de correlações, garantindo a identificabilidade dos modelos.

Um fato que merece ser comentado à respeito da matriz de covariâncias dos traços latentes, $\boldsymbol{\Sigma}_\theta$, é que ela não pode ter nenhum elemento igual a 0. Se de fato acontecer que o elemento da linha h e coluna l da matriz $\boldsymbol{\Sigma}_\theta$ é igual a 0, então o modelo não vai ser mais identificável, já que não será invariante ante rotações. Um exemplo é o seguinte, suponha que

$$\boldsymbol{\Sigma}_\theta = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \text{ e } \boldsymbol{R} = \begin{pmatrix} 1 & 0,5 & 0 \\ 0,5 & 1 & 0,3 \\ 0 & 0,3 & 1 \end{pmatrix},$$

agora, se $\boldsymbol{\theta}_i^* = \boldsymbol{R}\boldsymbol{\theta}_i$, então

$$\Phi(\boldsymbol{\alpha}_j^t\boldsymbol{\theta}_i - \beta_j) = \Phi(\boldsymbol{\alpha}_j^t\boldsymbol{R}^t\boldsymbol{\theta}_i^* - \beta_j) = \Phi((\boldsymbol{\alpha}_j^*)\boldsymbol{\theta}_i^* - \beta_j) \tag{3.4.7}$$

a covariância de $\boldsymbol{\theta}_i^*$ vai ser

$$\begin{pmatrix} 1 & 0,14 & 0,56 \\ 0,14 & 1 & 0 \\ 0,56 & 0 & 1 \end{pmatrix},$$

a qual de fato tem uma estrutura de matriz de correlação. Portanto para esta matriz $\boldsymbol{\Sigma}_\theta$ o modelo probito multidimensional não seria identificável.

A restrição imposta pela proposição (3.4.5), de que as correlações entre os traços latentes sejam não nulas, pode parecer bastante forte no sentido de impor a condição dos traços latentes sempre tiverem algum tipo de "dependência". Porem, se os elementos nulos da matriz de covariâncias do exemplo anterior forem substituídos, por exemplo pelo valor $0,01$, a diagonal da matriz de covariâncias de $\boldsymbol{\theta}_i^*$ vai ser $(1, 0, 91, 01)$, não preservando, então, a estrutura de uma matriz de correlações. As correlações embora não possam ser exatamente iguais a zero, podem ser próximas ao zero, e assim, a restrição da proposição (3.4.5) é mais razoável, o que nos leva à conseguir identificar correlações nulas, através das próprias estimativas e dos respectivos intervalos de credibilidade.

3.5 Modelos da TRIM com distribuições NMAC nos traços latentes

Na pressente seção desenvolvemos o modelo probito de 3 parâmetros, com distribuição dos traços latentes normal multivariada centrada apresentada no capítulo 2. Inicialmente para um único grupo e posteriormente para grupos múltiplos. Fornecemos uma base teórica para o desenvolvimento dos algoritmos MCMC para estimação dos parâmetros, que será feito no Capitulo 4. Esta seção introduz um novo modelo assimétrico para a TRIM, denominado modelo probito normal multivariado assimétrico centrado. Este modelo estende ao caso multidimensional os modelo probito normal assimétrico introduzido por Bazan (2005) Capítulo 3, o modelo assimétrico introduzido por Matos (2008), e os modelos considerando parametrizações centradas da normal assimétrica univariada propostos em Azevedo, Bolfarine e Andrade (2011), Azevedo, Bolfarine e Andrade (2012), Santos (2012) e Santos, Azevedo e Bolfarine (2013).

Considerando, inicialmente, o caso de um único grupo, vamos definir o o modelo probito D-dimensional de 3 parâmetros assimétrico (D-MP3AC) da seguinte forma

Modelo de resposta ao item (parte funcional)

$$\begin{aligned} Y_{ij}|\boldsymbol{\theta}_{i\cdot}, \boldsymbol{\zeta}_j, c_j &\sim \text{Bernoulli}(P_{ij}), \\ P_{ij} = P(Y_{ij}|\boldsymbol{\theta}_{i\cdot}, \boldsymbol{\zeta}_j) &= c_j + (1 - c_j)\Phi(\boldsymbol{\eta}_{ij}), \end{aligned} \tag{3.5.1}$$

Distribuição Latente (parte estrutural)

$$\boldsymbol{\theta}_{j\cdot}|\boldsymbol{\Theta} \sim NAC_D(\mathbf{0}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\delta}_\theta), \tag{3.5.2}$$

em que

$$\boldsymbol{\zeta}_j = (\boldsymbol{\alpha}_j, \beta_j)^t, \quad \boldsymbol{\Theta} = (\mathbf{0}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\lambda}_{\theta_k}) \text{ e } \eta_{ij} = \sum_{d=1}^{D} \alpha_{jd}\theta_{di} - \beta_j,$$

ou seja, seguindo a notação do capítulo 2, o vetor $\boldsymbol{\theta}_{i\cdot}$ segue uma distribuição normal multivariada assimétrica centrada, na parametrização ali introduzida. Sob esta parametrização, teremos que

$E(\boldsymbol{\theta}_{i\cdot}|\boldsymbol{\Theta}_i) = \mathbf{0}$ e $Cov(\boldsymbol{\theta}_{i\cdot}|\boldsymbol{\Theta}_i) = \boldsymbol{\Sigma}_\theta$, com a estrutura de $\boldsymbol{\Sigma}_\theta$ a de uma de matriz de correlações. Isto, e com ajuda da proposição (3.4.5), garante que o modelo (3.5.2) esta identificado.

É importante destacar, que o parâmetro de assimetria de nosso interesse, não é $\boldsymbol{\lambda}$, mas sim é $\boldsymbol{\delta}$ definido como no Capítulo 2, isto é, $\frac{\boldsymbol{\lambda}}{(1+\boldsymbol{\lambda}^t\boldsymbol{\lambda})^{1/2}}$. A escolha deste parâmetro, é feita por suas componentes ter uma interpretação semelhante ao coeficiente de assimetria de Pearson,

3.5.1 Caso de grupos múltiplos

Para a definição do modelo no caso de grupos múltiplos, deve-se considerar a situação em que se tem K grupos compostos de n_k indivíduos cada, para um total de $N = \sum_{k=1}^{K} n_k$. Estes são submetidos a diferentes testes com J_k ítens cada. Os testes devem possuir certa quantidade de ítens em comum e são escolhidos de um total de J ítens, com $\sum_{k=1}^{K} J_k \leq J$. Estando no caso multidimensional, é necessário impor outra condição aos ítens, esta é que para o conjunto de J ítens, eles devem medir o mesmo vetor, de dimensão $D \times 1$, de proficiências para cada indivíduo, isto é, se assume que os testes aos quais os diferentes grupos são submetidos medem as mesmas D dimensões. Esta restrição é feita, para garantir a determinação da dimensão de cada teste e subseqüentemente garantir a independência local, vide Seção 3.3. Define-se, agora, a seguinte notação: θ_{dik} é o d-ésimo, $(d = 1, \ldots, D)$ traço latente do individuo $i (i = 1, \ldots, n_k)$ pertencente ao grupo $k, (k = 1, \ldots, K)$. $\boldsymbol{\theta}_{\cdot jk} = (\theta_{1jk}, \ldots \theta_{Dik})^t$ é o vetor de traços latentes do individuo i do grupo k. $\boldsymbol{\theta}_{\cdot\cdot k} = (\boldsymbol{\theta}_{\cdot 1k}, \ldots \boldsymbol{\theta}_{\cdot n_k k})^t$ é a matriz de traços latentes dos indivíduos do grupo k. $\boldsymbol{\theta}^t_{\cdots} = (\boldsymbol{\theta}_{\cdot\cdot 1}, \ldots, \boldsymbol{\theta}_{\cdot\cdot K})^t$ é a matriz de todos os traços latentes: Y_{ijk} é a resposta do individuo i do grupo k no item $j (j = 1, \ldots, J)$. $\boldsymbol{Y}_{i\cdot k} = (Y_{i1k}, \ldots, Y_{iJk})^t$ será o vetor de respostas em todos os itens do individuo j do grupo k. $\boldsymbol{\zeta}_j$ é o vetor de parâmetros do item j. $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^t, \ldots, \boldsymbol{\zeta}_J^t)^t$ é o vetor com todos os parâmetros de todos os itens. $\boldsymbol{\Theta}_k$ é o vetor de parâmetros populacionais do grupo k e $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1^t, \ldots, \boldsymbol{\Theta}_k^t)^t$ é o vetor com os parâmetros populacionais de todos os grupos.

Desta forma baseados na notação de Santos (2012), definimos o modelo probito D-dimensional de 3 parâmetros assimétrico para grupos múltiplos (D-MP3ACGM) da seguinte forma

Modelo de resposta ao item (parte funcional)

$$\begin{aligned} Y_{ijk}|\boldsymbol{\theta}_{\cdot jk}, \boldsymbol{\zeta}_j, c_j &\sim \text{Bernoulli}(P_{ijk}), \\ p_{ijk} = P(Y_{ijk}|\boldsymbol{\theta}_{i\cdot k}, \boldsymbol{\zeta}_j) &= c_j + (1 - c_j)\Phi(\boldsymbol{\eta}_{ijk}), \end{aligned} \tag{3.5.3}$$

Distribuição Latente (parte estrutural)

$$\boldsymbol{\theta}_{j\cdot k}|\boldsymbol{\Theta}_k \sim NAC_D(\boldsymbol{\mu}_{\theta_k}, \boldsymbol{\Sigma}_{\theta_k}, \boldsymbol{\delta}_{\theta_k}), \tag{3.5.4}$$

em que $\boldsymbol{\zeta}_j = (\boldsymbol{\alpha}_j, \beta_j)^t$, $\boldsymbol{\Theta}_k = (\boldsymbol{\mu}_{\theta_k}, \boldsymbol{\Sigma}_{\theta_k}, \boldsymbol{\delta}_{\theta_k})$ e $\boldsymbol{\eta}_{ijk} = \sum_{k=1}^{K} \alpha_{iq}\theta_{djk} - \beta_j$.

Deve-se notar que para os traços latentes esta-se considerando distribuições NMAC com vetor de médias diferente ao vetor nulo, contrario a como era feito no caso de um único grupo. Este fato se deve à identificabilidade do modelo no caso de grupos múltiplos. Como já foi visto na Seção 3.4 do

presente capítulo, a identificabilidade dos modelos no âmbito da TRI está fortemente relacionada com o estabelecimento da métrica. No caso de grupos múltiplos, é de real interesse poder comparar as estimativas dos diferentes grupos e isto só será possível, se estabelecermos uma métrica de modo que as estimativas, estejam na mesma métrica, para cada dimensão. Como é explicado em Santos (2012), no modelo de grupos múltiplos se procede da seguinte forma: primeiro se faz com que os testes aplicados aos diferentes grupos possuam ítens em comum, sendo estes ítens são necessários para criar uma ligação entre os grupos. Segundo, é escolhido um grupo de referência, para o qual o vetor de médias e a matriz de covariâncias dos traços latentes, sejam fixados, com isto fixamos também a escala dos outros grupos, veja Bock e Zimowski (1987).

No caso de modelos unidimensionais, Santos (2012) apresenta uma discussão à respeito à identificabilidade dos modelos de K grupos. Consideremos, sem perda de generalidade, que o grupo de referência seja o primeiro. Assim deve-se considerar as seguintes restrições:

$$\begin{aligned} &\boldsymbol{\theta}_{\cdot j1} \sim NAC_d(\mathbf{0}, \boldsymbol{\Sigma}_{\theta_1}, \boldsymbol{\lambda}_{\theta_1}), \\ &\boldsymbol{\theta}_{\cdot jk} \sim NAC_d(\boldsymbol{\mu}_{\theta_k}, \boldsymbol{\Sigma}_{\theta_k}, \boldsymbol{\lambda}_{\theta_k}), \quad k = 2, \ldots, K. \end{aligned} \tag{3.5.5}$$

Diferente ao caso unidimensional, para a situação onde se tem D dimensões, dois cenários serão considerados:

- Se fixarmos a matriz de covariância do grupo de referencia como a matriz identidade. Neste caso, a matriz de covariância dos outros grupos devem ser iguais à matrizes diagonais,
- Se fixarmos a matriz de covariância do grupo de referencia como uma matriz de correlações, como foi feito no caso de um único grupo. Neste caso, as matrizes dos outros grupos devem ser matrizes de covariância.

Analogamente ao modelo de um grupo, note que não é possível considerar transformações do tipo (3.4.3) nos parâmetros do grupo de referência e, conseqüentemente, também não será possível fazê-lo nos parâmetros dos demais grupos, uma vez que, a escala desses estará ligada através dos ítens comuns. Isso garante que o modelo (3.5.3) está identificado.

Neste capítulo foi feito um breve resumo de alguns dos modelos clássicos da TRIM, junto com as definições dos seus parâmetros. Também foi feita uma discussão sobre a identificabilidade deles , além de permitir a modelagem da matriz de covariâncias dos traços latentes. O modelo probito D-dimensional de 3 parâmetros assimétrico para grupos múltiplos (D-MP3ACGM) foi definido, e apresentou-se as condições para garantir a sua identificabilidade.

Capítulo 4

Estimação via MCMC para Modelos da TRIM Assimétricos de Grupos Múltiplos

4.1 Introdução

Nas últimas décadas observou-se um crescimento no uso de métodos bayesianos baseados em Monte Carlo via Cadeias de Markov (MCMC) na área da TRI. No entanto os modelos da TRI também foram crescendo em complexidade, o que fez com que muitos dos mecanismos de estimação, usualmente empregados, se tornassem proibitivos. Assim os métodos MCMC constituem uma alternativa bastante útil para estes casos. Albert, no seu artigo pioneiro de 1992, marcou o inicio da implementação dos métodos MCMC na TRI. Albert (1992) empregou o esquema de dados aumentados para a etapa de estimação do vetor de parâmetros do modelo, em conjunto com o amostrador de Gibbs , ver Geman e Geman (1984); Gelfand e Smith (1990). Na literatura, podem-se identificar vários tipos de dados aumentados (Beguin e Glas (2001) e Sahu (2002)). Assim, uma motivação deste capitulo é comparar estes esquemas de dados aumentados no caso de modelos da TRIM. Também queremos avaliar se nossos algoritmos estão convergindo e obter os parâmetros MCMC: burn-in, espaçamento e tamanho total das cadeias, para se ter amostras validas de tamanho desejado. É considerada uma aplicação da proposta de Gonzalez (2004) para melhorar a convergência das cadeias, o que representa mais uma contribuição desta dissertação.

Para uma simulação de um teste multidimensional, os algoritmos apresentados são comparados. Aquele que segundo alguns critérios de comparação, forneça uma melhor alternativa, será escolhido para implementar os modelos mais complexos introduzidos no capítulo 3.

4.2 Dados aumentados

O esquema de dados aumentados (Tanner e Wong (1987)) foi uma importante contribuição para a utilização dos algoritmos Monte Carlo via Cadeias de Markov MCMC. Quando combinada com as ferramentas introduzidas nos trabalhos de Metropolis, Rosenbluth e Rosenbluth (1972) e Hastings (1970) torna possível a analise bayesiana de modelos mais complexos. O esquema de dados aumentados consiste em tomar dados latentes ou faltantes como parâmetros a serem estimados.

Embora isto introduza mais parâmetros para se estimar, as densidades condicionais ficam, em geral, mais simples de serem simuladas. ë importante destacar que a extensão dos esquemas para o caso de grupos múltiplos é imediata.

Na TRI, para os modelos dicotômicos de 3 parâmetros, se destacam dois esquemas de dados aumentados, o introduzido em Beguin e Glas (2001) e o introduzido em Sahu (2002). Primeiro se apresentara o algoritmo desenvolvido segundo o esquema de Beguin e Glas (2001).

4.2.1 Beguin e Glass

O procedimento introduzido em Beguin e Glas (2001), é um procedimento bayesiano para estimar o modelo multidimensional probito de três parâmetros. O procedimento é uma generalização daquele proposto por Albert(1992). Como foi visto no capitulo 3, no modelo multidimensional de 3 parâmetros con enlace probito a probabilidade de uma resposta correta é dada por

$$p(Y_{ij} = 1|\boldsymbol{\theta}_i, \boldsymbol{\zeta}_j, c_j) = c_j + (1 - c_j)\boldsymbol{\Phi}(\eta_{ij}), \tag{4.2.1}$$

A expressão (4.2.1) pode ser interpretada como uma combinação linear entre a probabilidade $\boldsymbol{\Phi}(\eta_{ij})$ de que o indivíduo de fato sabe a resposta correta e responder corretamente com probabilidade um, e a probabilidade $(1 - \boldsymbol{\Phi}(\eta_{ij}))$ de que o individuo não saiba a resposta correta mas responda corretamente ao acaso com probabilidade c_j. Assim a probabilidade de uma resposta correta é a soma de um termo $\boldsymbol{\Phi}(\eta_{ij})$, com um termo $c_j(1-\boldsymbol{\Phi}(\eta_{ij}))$. De acordo com esta interpretação, Beguin e Glas (2001) introduzem o vetor de variáveis binárias W_{ij} tal que

$$W_{ij} = \begin{cases} 1, \text{ se o individuo } i \text{ conhece a resposta correta ao item } j \\ 0, \text{ se o individuo } i \text{ não conhece a resposta correta ao item } j. \end{cases} \tag{4.2.2}$$

De tal forma que se $W_{ij} = 0$, o individuo i vai "chutar" a resposta ao item j; e se $W_{ij} = 1$, o individuo i sabe a resposta correta e portanto responderá o item corretamente. Conseqüentemente, a probabilidade condicional de $W_{ij} = w_{ij}$ dado Y_{ij} é dada por

$$\begin{aligned} &p(W_{ij} = 1|Y_{ij} = 1, \eta_{ij}, c_j) \propto \boldsymbol{\Phi}(\eta)_{i,j} \\ &p(W_{ij} = 0|Y_{ij} = 1, \eta_{ij}, c_j) \propto c_j(1 - \boldsymbol{\Phi}(\eta)_{i,j})) \\ &p(W_{ij} = 1|Y_{ij} = 0, \eta_{ij}, c_j) = 0 \\ &p(W_{ij} = 0|Y_{ij} = 0, \eta_{ij}, c_j) = 1. \end{aligned} \tag{4.2.3}$$

Para a implementação do amostrador de Gibbs, aos dados $\boldsymbol{y}$ utiliza-se também os dados latentes $\boldsymbol{w} \equiv (w_{11}, \ldots, w_{nk})^t$. Os dados também são aumentados com outro conjunto de dados latentes $\boldsymbol{z} \equiv (z_{11}, \ldots, z_{nk})^t$, onde as variáveis Z_{ij} são independentes e normalmente distribuídas com média η_{ij} e desvio padrão 1. Os conjuntos de variáveis aumentados se relacionam no sentido em que $z_{ij} > 0$ se $w_{ij} = 1$ e $z_{ij} \leq 0$ e $w_{ij} = 0$. Isto é

$$p(z_{ij}|w_{ij}, \eta_{ij}) \propto \phi(z_{ij}; \eta_{ij}, 1)[\boldsymbol{I}(z_{ij} > 0)\boldsymbol{I}(w_{ij} = 1) + \boldsymbol{I}(z_{ij} \leq 0)\boldsymbol{I}(w_{ij} = 0)], \tag{4.2.4}$$

onde $\boldsymbol{I}(\cdot)$ é a função indicadora usual que assume o valor um se o seu argumento for verdadeiro e o valor zero caso contrario.

Para os α_{jk} e os β_j se consideram distribuições a priori normais multivariadas simétricas . Para o parâmetro de acerto casual se considera a priori conjugada $Beta(\kappa_{1c}, \kappa_{2c})$. A distribuição a posteriori conjunta de $\boldsymbol{\zeta}$, $\boldsymbol{c}$, $\boldsymbol{\theta}$, $\boldsymbol{z}$ e $\boldsymbol{w}$ é dada por

$$\begin{aligned} p(\boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w}|\boldsymbol{y}) &= p(\boldsymbol{z},\boldsymbol{w}|\boldsymbol{y}; \boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) p(\boldsymbol{\zeta}) p(\boldsymbol{c}) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{k} p(z_{ij}|w_{ij}; \eta_{ij}) p(w_{ij}|y_{ij}; \eta_{ij}, c_j) p(\boldsymbol{\theta}_i) p(\boldsymbol{\zeta}) p(\boldsymbol{c}) \end{aligned} \tag{4.2.5}$$

onde se define a $p(z_{ij}|w_{ij}; \eta_{ij})$ como em (4.2.4) e $p(w_{ij}|y_{ij}; \eta_{ij}, c_j)$ como em (4.2.5).

Denotando por $(\cdot)$ o conjunto de todos os outros parâmetros, temos que o algoritmo proposto em Beguin e Glas (2001) para $t = 1, 2, \ldots, B, \ldots, M$, onde B é o burn in e M é o tamanho de amostra gerada: simula iterativamente todas as quantidades desconhecidas na seguinte ordem:

- Inicializar o algoritmo escolhendo valores iniciais convenientes;
- Simule as variáveis aumentadas W_{ij} e Z_{ij} de $W_{ij}|(\cdot)$ e $Z_{ij}|(\cdot)$ respectivamente. $i = 1, \ldots, N$ e $j = 1, \ldots, J$;
- Simule os traços latentes $\boldsymbol{\theta}_i$ de $\boldsymbol{\theta}_i|(\cdot)$. $i = 1, \ldots, N$;
- Simule $\boldsymbol{\zeta}_j$ de $\boldsymbol{\zeta}_j|(\cdot)$. $j = 1, \ldots, J$;
- Simule c_j de $c_j|(\cdot)$. $j = 1, \ldots, J$,

para maior profundidade no algoritmo se recomenda ao leitor o trabalho Beguin e Glas (2001).

4.2.2 Sahu

Outro esquema de dados aumentados bastante referenciado na literatura para ajustar o modelo de ogiva normal é aquele introduzido em Sahu (2002), porem o algoritmo referenciado foi implementado para o caso unidimensional. Uma extensão para o caso multidimensional com ligação logito pode ser encontrada em exemplo Fu, Tao e Shi (2009). Neste trabalho é feita uma extensão multidimensional que difere daquela apresentada em Fu, Tao e Shi (2009). O algoritmo proposto por Sahu (2002) introduz dois tipos de variáveis aleatórias correspondentes a cada observação y_{ij}. A primeira é uma variável aleatória (aumentada) Bernoulli, que é denotada por U_{ij} com probabilidade de sucesso c_j. A segunda variável aumentada é Z_{ij} como definida em Beguin e Glas (2001), isto é, uma variável aleatória com distribuição normal com média η_{ij} e variância 1. A relação destas variáveis aumentadas com os dados $\boldsymbol{Y}$ é

$$y_{ij} = u_{ij} + (1 - u_{ij}) \times \boldsymbol{I}(z_{ij} > 0). \tag{4.2.6}$$

Novamente, para os α_{jd} e os $\beta_j$se consideram distribuições a priori normais multivariadas simétricas, com matriz de covariancias uma matriz diagonal. Para o parâmetro e acerto casual se

considera uma a priori $Beta(a, b)$. A distribuição a posteriori conjunta de $\boldsymbol{\zeta}$, $\boldsymbol{c}$, $\boldsymbol{\theta}$, $\boldsymbol{z}$ e $\boldsymbol{u}$ é dada por

$$p(\boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{u}|\boldsymbol{y}) = p(\boldsymbol{z},\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta})p(\boldsymbol{\theta})p(\boldsymbol{\zeta})p(\boldsymbol{c}|\boldsymbol{u}) \tag{4.2.7}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{k} p(z_{ij}|u_{ij}; \eta_{ij})p(u_{ij}|y_{ij}; \eta_{ij}, c_j)p(\boldsymbol{\theta}_i)p(\boldsymbol{\zeta})p(\boldsymbol{c}) \tag{4.2.8}$$

O procedimento consiste na simulação passo a passo das condicionais a posteriori das componentes $\boldsymbol{\zeta}$, $\boldsymbol{c}$, $\boldsymbol{\theta}$, $\boldsymbol{z}$ e $\boldsymbol{w}$.

O algoritmo MCMC usando o esquema de dados aumentados de Sahu (2002), é bastante similar ao de Beguin e Glas (2001). Novamente denotando por $(\cdot)$ o conjunto de todos os outros parâmetros, temos que o algoritmo proposto em Sahu para $t = 1, 2, \ldots, B, \ldots, M$, onde B é o burn in e M é o tamanho de amostra gerado: simula iterativamente todas as quantidades desconhecidas na seguinte ordem:

- Inicializar o algoritmo escolhendo valores iniciais convenientes;
- Simule as variáveis aumentadas U_{ij} e Z_{ij} de $U_{ij}|(\cdot)$ e $Z_{ij}|(\cdot)$ respectivamente. $i = 1, \ldots, N$ e $j = 1, \ldots, J$;
- Simule os traços latentes $\boldsymbol{\theta}_i$ de $\boldsymbol{\theta}_i|(\cdot)$. $i = 1, \ldots, N$;
- Simule $\boldsymbol{\zeta}_j$ de $\boldsymbol{\zeta}_j|(\cdot)$. $j = 1, \ldots, J$;
- Simule c_j de $c_j|(\cdot)$. $j = 1, \ldots, J$,

para maior profundidade no algoritmo se recomenda ao leitor a referencia Sahu(2001).

4.3 Metropolis Hasting MCMC

Em Patz e Junker (1999b) é descrito um método MCMC baseado no algoritmo de Metropolis-Hasting, para alguns modelos da TRI. Em Patz e Junker (1999a) os mecanismos de estimação foram estendidos a casos de não resposta e modelos politómicos, entre outros.

Sob a suposição de independência local, a verossimilhança total das respostas para os N indivíduos e os J itens pode ser escrita como

$$\begin{aligned} L(\boldsymbol{y}|\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c}) &= \prod_{i=1}^{N} L(y_{i\cdot}|\boldsymbol{\theta}_{i\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c}) \\ &= \prod_{j=1}^{J} L(y.j|\boldsymbol{\theta}_{\cdot j}, \boldsymbol{\alpha}_{j\cdot}, \beta_j, c_j) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{n} p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}, \end{aligned} \tag{4.3.1}$$

estas equações nos dizem respeito a verossimilhança da matriz de resposta de dimensão $N \times J$, e ela pode ser expressa como o produto das verossimilhanças entre os respondentes ou como o produto das verossimilhanças entre os items.

Denotamos a distribuição a priori conjunta de todos os parâmetros e traços latentes por $\pi(\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c})$. Então a distribuição a priori pode ser escrita como

$$\begin{aligned} \pi(\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c}) &= \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{\cdot\cdot})\pi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{\cdot\cdot})\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\pi_{\boldsymbol{c}}(\boldsymbol{c}) \\ &= \Big(\prod_{i=1}^{N} \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{i\cdot})\Big) \prod_{j=1}^{J} \pi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{j\cdot})\pi_{\beta}(\beta_j)\pi_{c}(c_j), \end{aligned} \tag{4.3.2}$$

assim a distribuição a posteriori conjunta para o modelo pode-se expressar como

$$p(\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c})\pi(\boldsymbol{\theta}_{\cdot\cdot}, \boldsymbol{\alpha}_{\cdot\cdot}, \boldsymbol{\beta}, \boldsymbol{c}). \tag{4.3.3}$$

Não é possível obter amostras simuladas diretamente desta distribuição a posteriori conjunta, ver Jiang (2005). Assim, com o objetivo de obter valores amostrais para os parâmetros do modelo a partir da distribuição conjunta, deve-se implentar o algoritmo de Metropolis Hastings dentro do amostrador de Gibbs (MHGS) [1]

No algoritmo de Metropolis-Hastings, diferentemente do amostrador de Gibbs, escolhe-se um estado via um esquema de passeio aleatório e este estado candidato nem sempre é aceito, diferentemente do amostrados de Gibbs. Mais precisamente, seja r uma densidade, tal que a função de transição seja definida como $R(y, B) = \int_B r(z-y)dz$. Então, define-se a probabilidade de aceitação α como

$$\alpha(y, z) = min\left\{\frac{g(z)r(y-z)}{g(y)r(z-y)}, 1\right\}, \tag{4.3.4}$$

onde $g(\cdot)$ é a distribuição objetivo. Por exemplo, as distribuições a posteriori para cada traço latente dos indivíduos, ou a distribuição condicional completa para cada parâmetro do item. Se o denominador em (4.3.4) for igual a 0, simplesmente se fixa a probabilidade de aceitação $\alpha = 1$.

A seguir apresentam-se as densidades propostas correspondentes aos traços latentes e aos parâmetros dos ítens. Elas sao escolhidas tanto por conveniência como por eficiência.

- Densidade proposta para $\boldsymbol{\theta}^{t+1}$ é $N_D(\boldsymbol{\theta}^t, \boldsymbol{I}_D)$.

- Densidade proposta para β^{t+1} é $N(\beta^t, \sigma^2_\beta)$.

- Densidade proposta para c^{t+1} é $U(c^t - \delta_c, c^t + \delta_c)$;

[1] Sigla do inglês: Metropolis-Hastings within Gibbs sampling.

onde σ_β^2, δ_c são constantes. Aqui iguais a 1 e $\frac{1}{4}$.

O algoritmo então é:

- Simule $\boldsymbol{\theta}_{i\cdot}^*$ de uma $N_D(\boldsymbol{\theta}_{i\cdot}^t, \boldsymbol{I})$, $\forall i = 1, 2, \ldots, N$. Aceite $\boldsymbol{\theta}_{i\cdot}^{t+1} = \boldsymbol{\theta}_{i\cdot}^*$ com probabilidade $\alpha(\boldsymbol{\theta}_{i\cdot}^t, \boldsymbol{\theta}_{i\cdot}^*)$,
- Simule $\boldsymbol{\theta}_{i\cdot}^*$ de uma $N_D(\boldsymbol{\theta}_{i\cdot}^t, \boldsymbol{I})$, $\forall i = 1, 2, \ldots, N$. Aceite $\boldsymbol{\theta}_{i\cdot}^{t+1} = \boldsymbol{\theta}_{i\cdot}^*$ com probabilidade $\alpha(\boldsymbol{\theta}_{i\cdot}^t, \boldsymbol{\theta}_{i\cdot}^*)$,
- Simule β_j^* de uma $N(\beta_j^t, \sigma_\beta^2)$, $\forall j = 1, 2, \ldots, J$. Aceite $\beta_j^{t+1} = \beta_j^*$ com probabilidade $\alpha(\beta_j^t, \beta_j^*)$,
- Simule c_j^* de uma $U(c_j^t - \delta_c, c_j^t + \delta_c)$, $\forall j = 1, 2, \ldots, J$. Aceite $c_j^{t+1} = c_j^*$ com probabilidade $\alpha(c_j^t, c_j^*)$.

Maiores detalhes podem ser encontrados no Apêndice.

4.4 Proposta de aceleração de convergência

Como vimos nas seções anteriores, os dados aumentados podem facilitar a forma dos algoritmos MCMC porém, a convergência destes algoritmos pode ser lenta devido às altas correlações entre os parâmetros do modelo e os dados latentes. Correlações muito grandes implicam que as cadeias se movem lentamente ao longo do espaço paramétrico fazendo necessário uma grande quantidade de iterações para se obter uma amostra válida, no sentido dos algoritmos MCMC, da densidade a posteriori. Além disso, um número grande de iterações será necessário para determinar a convergência da cadeia. Isto motiva o desenvolvimento de ferramentas para reduzir as autocorrelações da cadeia e, conseqüentemente, melhorar a convergência do algoritmo MCMC.

Uma proposta de redução das autocorrelações das cadeias é dada em Gonzalez(2004). Tal artigo é orientado a redução das autocorrelações para modelos probito e probito multivariado. O algoritmo que foi proposto é um resultado de acrescentar um passo ao algoritmo de amostrador de Gibbs convencional. A continuação se apresentara o algoritmo para o caso de um modelo probito, assim como introduzido em Gonzalez (2004).

Primeiro se supõe a variável y_i que pode ser zeros ou um e que a sua vez é determinada por uma variável latente continua y_i^* tal que

$$y_i^* = X_i a + e_i, \quad i = 1, \ldots, N$$

a variável e_i se distribui segundo uma lei normal com média zero e variância igual a 1. A variável binária y_i é igual a um se e só se $y_i^* \geq 0$, e será igual a zero no caso contrario. X_i é um vetor $1 \times l$ de regressoras e a é um vetor de parâmetros, isto é, se supõe que se tem l variáveis exploratórias na equação, com parâmetros $a = (a_1, \ldots, a_l)^t$. Conisedere-se a seguinte reparametrização do modelo,

$$\overline{\pi} = \left(a_1, \frac{a_2}{a_1}, \frac{a_3}{a_1}, \ldots, \frac{a_l}{a_1}\right) = (a_1, \overline{a}_2, \ldots, \overline{a}_l) \tag{4.4.1}$$

$$y_i^* = X_{i1}\pi_1 + X_{i2}\overline{a_2}a_1 + X_{i3}\overline{a_3}a_1 + \cdots + X_{il}\overline{a_l}a_1 \tag{4.4.2}$$

A distribuição a posteriori de $\overline{\pi}$, que será denotada $\overline{a}_M(\cdot)$ é derivada da densidade a posteriori de a (a_M) desta forma:

$$\overline{a}_M(a_1, \overline{a}_2, \ldots, \overline{a}_l) = a_M(a_1, \overline{a}_2 a_1, \ldots, \overline{a}_l)|a_1|^{l-1},$$

onde $|a_1|^{l-1}$ é o Jacobiano da transformação.

Aplicando esta proposta aos modelos da TRIM, especialmente ao esquema de dados aumentados de Beguin e Glass (2001) para o modelo probito de 3 parâmetros multidimensional vamos ter que o vetor de parâmetros a_j será agora $(\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jD}, \beta)^t$. Assim o algoritmo de Beguin e Glass incluindo este passo de aceleração de convergência será:

para $t = 1, 2, \ldots, B, \ldots, M$, onde B é o burn in e M é o tamanho de amostra gerado: simula iterativamente todas as quantidades desconhecidas na seguinte ordem:

- Inicializar o algoritmo escolhendo valores iniciais convenientes;
- Simule as variáveis aumentadas W_{ij} e Z_{ij} de $W_{ij}|(\cdot)$ e $Z_{ij}|(\cdot)$ respectivamente. $i = 1, \ldots, N$ e $j = 1, \ldots, J$;
- Simule os traços latentes $\boldsymbol{\theta}_i$ de $\boldsymbol{\theta}_i|(\cdot)$. $i = 1, \ldots, N$;
- Simule $\boldsymbol{\zeta}_j$ de $\boldsymbol{\zeta}_j|(\cdot)$. $j = 1, \ldots, J$;
- Simule α_{j1} de $\alpha_{j1}|(\overline{\alpha}_{j2}, \ldots, \overline{\alpha}_{jD}, \overline{\beta})$. $j = 1, \ldots, J$;
- Simule c_j de $c_j|(\cdot)$. $j = 1, \ldots, J$,

Note que este algoritmo é o mesmo que aquele apresentado por Beguin e Glass (2001), comum passo extra. Portanto, a diferença consiste em que todos os parâmetros $(\alpha_{jd}, \ldots, \alpha_{jD}, \beta)^t$ são multiplicados por um fator aleatório. Este passo extra acelera o algoritmo porque ele propõe uma mudança nos parâmetros a qual não depende dos dados latentes. A aplicação desta proposta para o esquema de dados aumentados de Sahu(2001) se construíram de forma análoga. Uma descrição mais detalhada deste algoritmo pode ser vista no apêndice.

4.5 Estudo de simulação

Nesta seção avaliaremos os cinco algoritmos MCMC discutidos, em termos da qualidade de convergência, avaliando a magnitude das autocorrelações. As análises serão feitas com base em um único conjunto de respostas que foram simuladas considerando-se a seguinte situação: 2000 indivíduos, 30 itens, 2 dimensões, e tais que $\boldsymbol{\theta}_{id} \sim N_2(\mathbf{0}; \boldsymbol{I}_2)$, $\forall i$, $\forall d$. Assim poderemos comparar os algoritmos sob uma circunstância bastante simples no sentido de que não assumimos assimetria para os traços latentes, não temos dados faltantes, e fixamos a métrica dos traços latentes como $(\mathbf{0}, \boldsymbol{I}_2)$. Os parâmetros dos itens foram fixados nos seguintes intervalos: $\alpha_{jd} \in [0; 1, 4], \forall j = 1, 2, \ldots 30$, $\forall d = 1, 2$. $\beta_j \in [-3; 3]$ e $c_j \in [0, 20; 0, 25]$.

4.5.1 Comparação dos algoritmos de estimação

Foram gerados 50000 valores utilizando os algoritmos de Beguin e Glas (2001), Sahu (2002), Metropolis Hastings, Beguin e Glass com o passo de aceleração de convergência de Gonzalez (2004) e Sahu com o passo de aceleração de convergência de Gonzalez (2004). Consideramos um período de aquecimento (*burn in*) de 10020 iterações, já que a convergência de todas as cadeias parecia ocorrer após este numero. Para todas as cadeias pegamos valores de 40 em 40, para assim reduzir a autocorrelação e em todos os casos obter um tamanho da amostra de 1000. Simulamos 3 cadeias com diferentes valores iniciais, para cada algoritmo, com o intuito de verificar a mistura entre elas. No apêndice apresentamos as simulações para as três cadeias, o gráfico do critério para monitorar a convergência de simulações iterativas proposto por Brooks e Gelman (1998) e os correlogramas sem espaçamento para os parâmetros dos itens 10, 20 e 30.

Nas Figuras 4.1 apresentamos os gráficos de valores simulados dos parâmetros do item 16 que foi escolhido aleatóriamente. Os gráficos mostram os valores após o período de aquecimento considerando os três conjuntos de valores iniciais com o espaçamento, isto para facilitar a visualização. Podemos ver que as cadeias se misturam bem, indicando que os diferentes valores iniciais não alteram, de forma significativa, os resultados. Nas figuras 4.2 a 4.5 mostram os correlogramas, com e sem espaçamento, das amostras geradas para os parâmetros do ítem 16, para um dos conjuntos de valores iniciais.

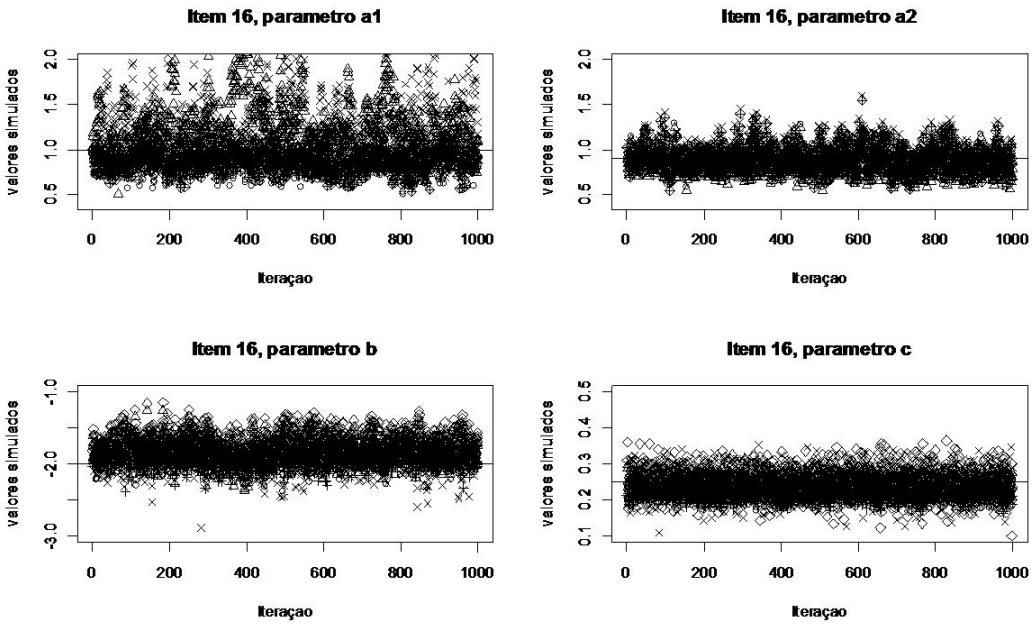


Figura 4.1: Valores simulados para o parâmetros do ítem 16. Legenda: + Sahu-Gonzalez, × Beguin e Glass-Gonzalez, ∘ Sahu, △ Beguin e Glas, ⊕ Metropolis Hasting

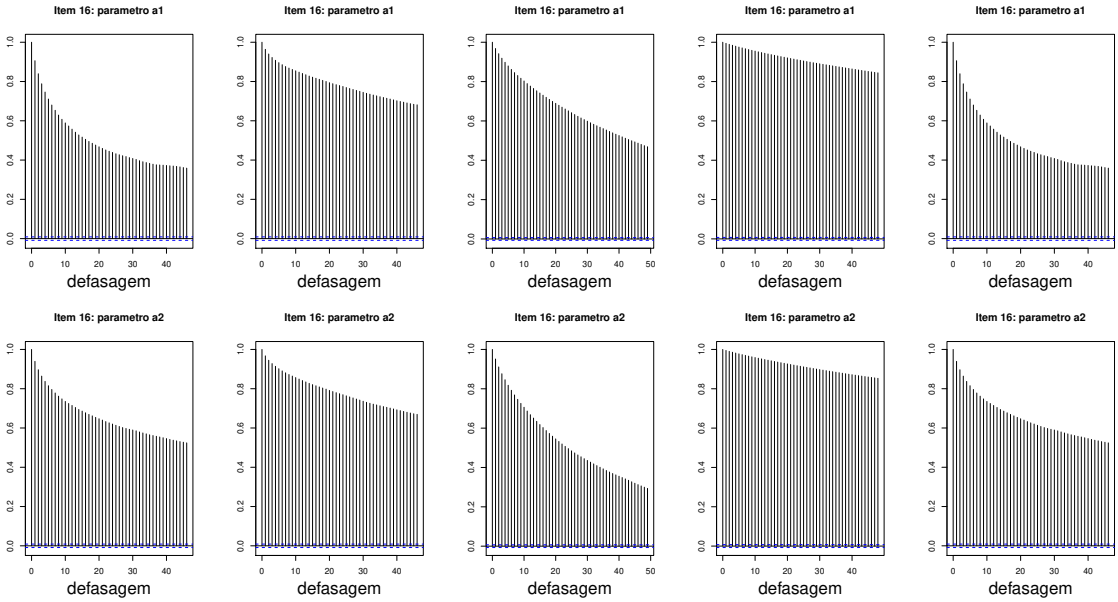


Figura 4.2: Correlogramas sem espaçamento das amostras geradas para os parâmetros do item 16. Ordem de esquerda a direita: Sahu-Gonzalez, Beguin Glass-Gonzalez, Sahu, Beguin Glass, Metropolis Hasting

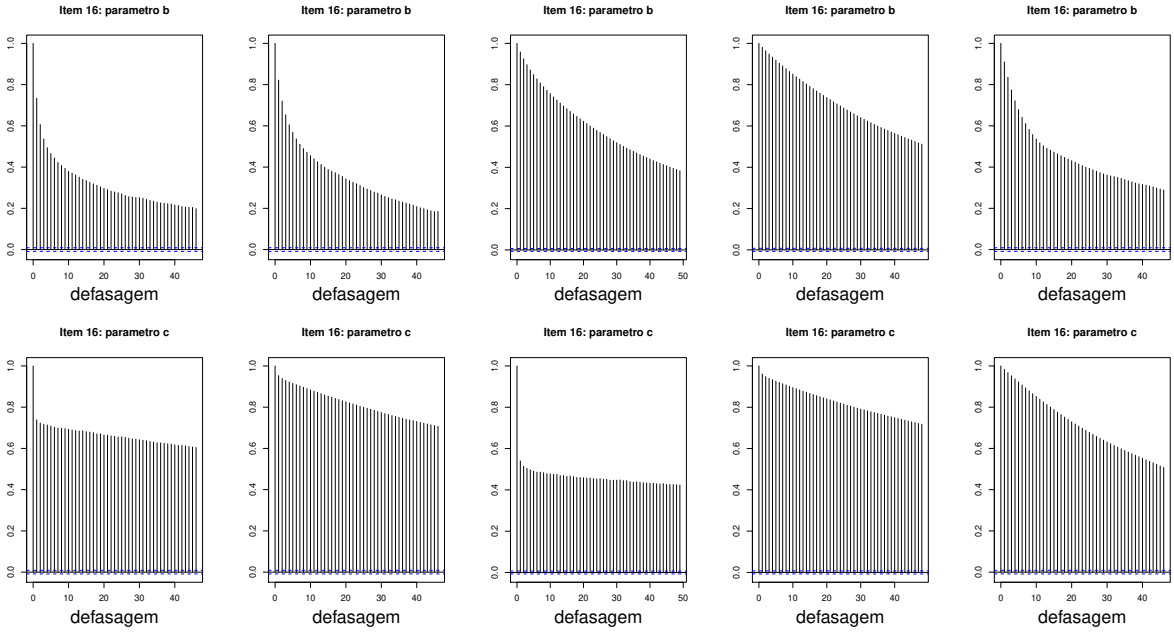


Figura 4.3: Correlogramas sem espaçamento das amostras geradas para os parâmetros do item 16 (continuação). Ordem de esquerda a direita: Sahu-Gonzalez, Beguin Glass-Gonzalez, Sahu, Beguin Glass, Metropolis Hasting

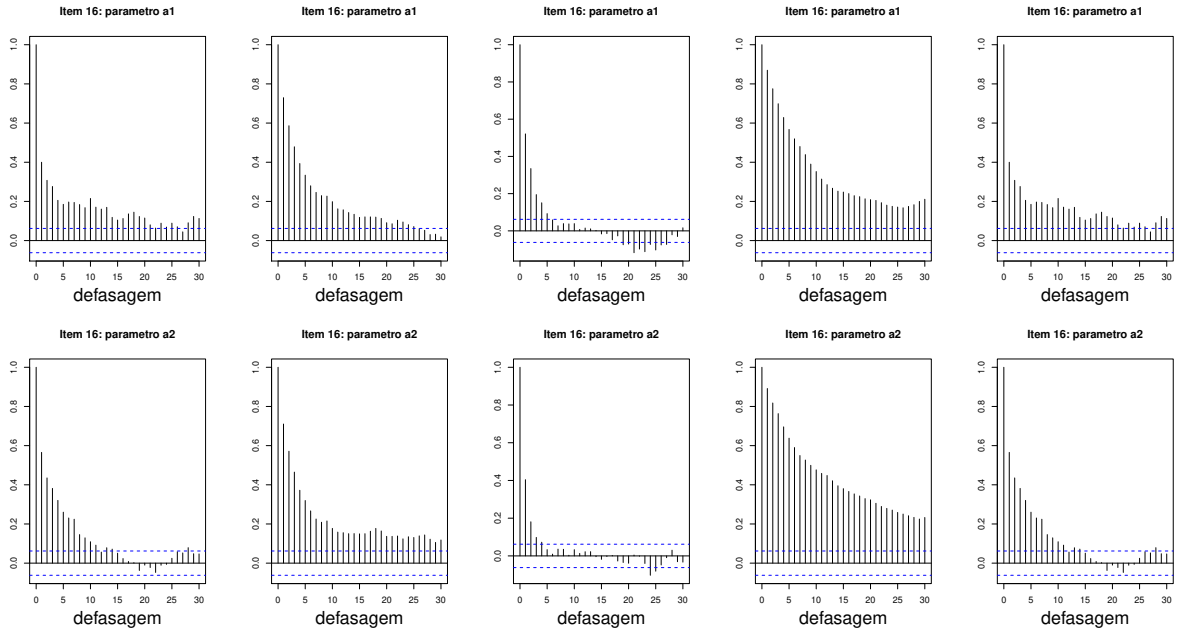


Figura 4.4: Correlogramas com espaçamento das amostras geradas para os parâmetros do item 16. Ordem de esquerda a direita: Sahu-Gonzalez, Beguin Glass-Gonzalez, Sahu, Beguin Glass, Metropolis Hasting

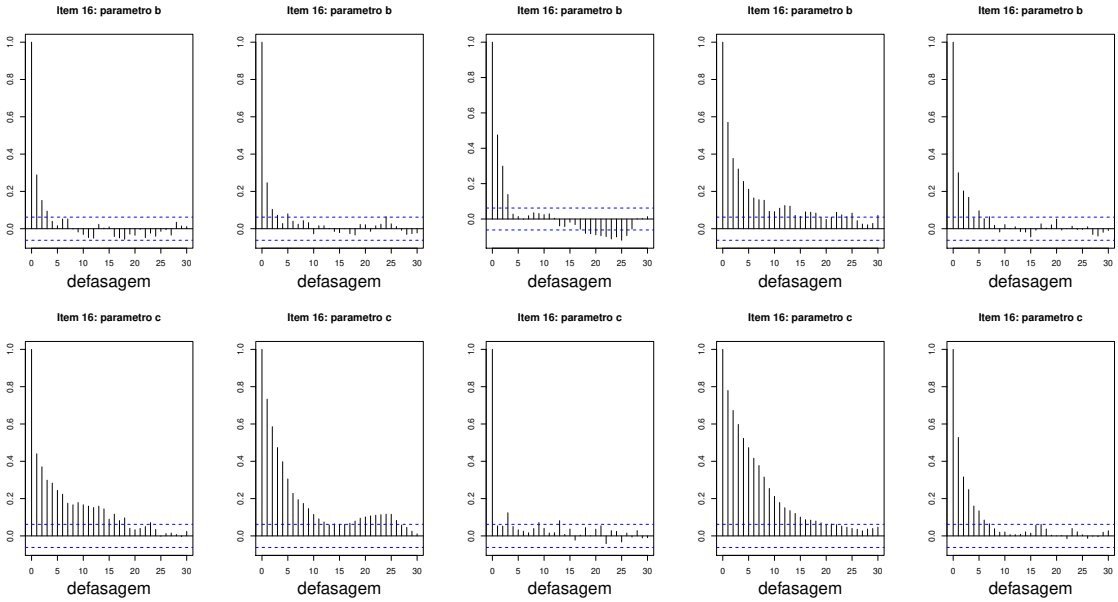


Figura 4.5: Correlogramas com espaçamento das amostras geradas para os parâmetros do item 16 (continuação). Ordem de esquerda a direita: Sahu-Gonzalez, Beguin Glass-Gonzalez, Sahu, Beguin Glass, Metropolis Hasting

Os resultados anteriores confirmam a convergência de todos os cinco algoritmos após um certo número de iterações. Porém, como discutido em Santos, Azevedo e Bolfarine (2013), os longos períodos de computação que são demandados pelos métodos MCMC, fazem necessaria realizar a comparação dos algoritmos de forma que o tempo para atingir a convergência seja tomado em conta. O tamanho efetivo de amostra (ESS) [2] representa um critério de comparação bastante conveniente para este objetivo. Particularmente neste trabalho utilizamos o ESS como feito em Sahu (2002). O ESS se define para cada parâmetro como o tamanho das amostras MCMC dividido pela autocorrelação, $\gamma = 1 + 2\sum_{k=1}^{\infty} \rho_k$, onde ρ_k é a autocorrelação com defasagem k. Portanto, quanto maior for o valor do critério ESS, mais proxima de uma amostra aleatória da distribuição a posteriori conjunta esta a amostra gerada e mais representativa esta sera, ver Gammerman e Lopes (2006). Em Sahu (2002), foi proposto como estimativa da autocorrelação o valor $(1+\rho^*)/(1-\rho^*)$, onde $\rho^* = |\rho_1|$.

Com as amostras obtidas para os cinco algoritmos MCMC calculamos os ESS. A Tabela 4.1 mostra as médias do ESS para cada um dos algoritmos. Em termos do ESS, o esquema de dados aumentados de Sahu (2002), com a proposta de aceleração de convergência de Gonzalez (2004) teve a melhor performance. No decorrer do pressente trabalho, para implementação dos algoritmos MCMC trabalho se usará o esquema de dados aumentados de Sahu (2002) com a proposta de aceleração de convergência de Gonzalez (2004).

Tabela 4.1: Comparação dos diferentes algoritmos MCMC para os dados simulados, segundo o critério do ESS

Algoritmo de estimação	Sahu-Gonzalez	BG-Gonzalez	Sahu	BG	MH
ESS	3265.666	845.6012	3210.733	949.8026	1058.033

4.6 Algoritmo MCMC para o modelo D-MP3NAC

Nesta seção, apresentamos uma *versão básica* do algoritmo ADMHGSSR [3], ver Azevedo, Bolfarine e Andrade (2012) eSantos (2012), o qual permite obter amostras marginais dos parâmetros do modelo D-MP3AC introduzido no Capítulo 3. Diz-se versão básica pelo fato do algoritmo não considerar a introdução de qualquer restrição sob os parâmetros do item, necessária, por exemplo, para contornar possíveis problemas de não identificabilidade.

Estamos assumindo como distribuição a priori para os traços latentes uma $NAC_D(\mathbf{0}, \Sigma_\theta, \boldsymbol{\lambda}_\theta)$ como vista no Capítulo 2. Podemos descrever a distribuição dos traços latentes de forma hierárquica da seguinte forma.

$$\boldsymbol{\theta}_{i\cdot}|(H_i, \Omega_i) \sim N_D\Big(\boldsymbol{A\delta}H_i + \boldsymbol{B}, \boldsymbol{A}(\boldsymbol{I}_D - \boldsymbol{\delta\delta}^t)\boldsymbol{A}^t\Big), \tag{4.6.1}$$

$$H_{jk} \sim HN(0,1), \tag{4.6.2}$$

[2]Sigla do inglês: effective sample size

[3]Sigla do inglês: Augmented data Metropolis-Hastings within Gibbs sampling stochastic representation

em que,

$$\begin{aligned}\boldsymbol{A} &= \Psi_z^{-1/2},\\ \boldsymbol{B} &= -\sqrt{\frac{2}{\pi}}\boldsymbol{\Psi}_z^{-1/2}\boldsymbol{\delta},\\ \boldsymbol{\delta} &= \frac{\boldsymbol{\lambda}}{(1+\boldsymbol{\lambda}^t\boldsymbol{\lambda})^{1/2}}.\end{aligned} \tag{4.6.3}$$

Deve-se notar que estamos modelando a matriz de covariâncias para os traços latentes e o vetor de parâmetros de assimetria, para eles serão preciso passos Metropolis. Consideramos agora as densidades propostas para os parâmetros populacionais: para a matriz de covariâncias se considera uma Wishart inversa como sugerido por Gelman et al. (2013). Para cada componente do parâmetro de assimetria $\boldsymbol{\delta}$ consideram-se a prioris uniformes. Assim,

$$q(\delta_d^{(t-1)}, \delta_d) = U(\nu_1((\delta_d^{(t-1)})), \nu_2(\delta_d^{(t-1)})), \tag{4.6.4}$$

em que

$$\begin{aligned}\nu_1(\delta_d^{(t-1)}) &= max\{-0,99527; \delta_d^{(t-1)}) - \Delta_\delta\}, \quad \Delta_\delta > 0,\\ \nu_2(\delta_d^{(t-1)}) &= min\{-0,99527; \delta_d^{(t-1)}) + \Delta_\delta\},\end{aligned}$$

para a matriz de covariância,

$$q(\Sigma_d^{(t-1)}, \Sigma_d) = IW(m+D, (D-1)\Sigma_d^{-1(t)}), \tag{4.6.5}$$

com m e Δ_δ constantes previamente definidas. Assim, estamos assumindo a seguinte forma geral para a priori conjunta dos parâmetros.

$$\begin{aligned}p(\boldsymbol{\theta}_{..}, \boldsymbol{h}_{..}\boldsymbol{\zeta}_{.}, \boldsymbol{c}, \boldsymbol{\Theta}|\boldsymbol{\Xi}_{\boldsymbol{\theta}}, \boldsymbol{\Xi}_{\boldsymbol{\zeta}}, \boldsymbol{\Xi}_c) = &\left\{\prod_{i=1}^{N} p(\boldsymbol{\theta}_i|\boldsymbol{\Theta})\right\}\left\{\prod_{i=1}^{N} p(h_i)\right\}\left\{\prod_{j=1}^{J} p(\boldsymbol{\zeta}_j|\boldsymbol{\Xi}_\zeta)\right\}\\ &\times\left\{\prod_{j=1}^{J} p(\boldsymbol{c}_j|\boldsymbol{\Xi}_c)\right\}\{p(\boldsymbol{\Theta}|\boldsymbol{\Xi}_\theta)\}\end{aligned} \tag{4.6.6}$$

O algoritmo ADMHGSRS para $t = 1, 2, \ldots, B, \ldots, M$, onde B é o burn in e M é o tamanho de amostra gerado: simula iterativamente todas as quantidades desconhecidas na seguinte ordem:

- Inicializar o algoritmo escolhendo valores iniciais convenientes;
- Simule as variáveis aumentadas U_{ij} e Z_{ij} de $U_{ij}|(\cdot)$ e $Z_{ij}|(\cdot)$ respectivamente. $i = 1, \ldots, N$ e $j = 1, \ldots, J$;
- Simule h_i de $H_i|(\cdot).i = 1, \ldots, N$;

- Simule os traços latentes $\boldsymbol{\theta}_i$ de $\boldsymbol{\theta}_i|(\cdot)$. $i = 1, \ldots, N$;
- Simule δ_j de $\delta_j|(\cdot).j = 1, \ldots, J$;
- Simule $\boldsymbol{\Sigma}_{\theta_j}$ de $\boldsymbol{\Sigma}_{\theta_j}|(\cdot).i = 1, \ldots, N$;
- Simule $\boldsymbol{\zeta}_j$ de $\boldsymbol{\zeta}_j|(\cdot)$. $j = 1, \ldots, J$;
- Simule α_{j1} de $\alpha_{j1}|(\overline{\alpha}_{j2}, \ldots, \overline{\alpha}_{jD}, \overline{\beta})$. $j = 1, \ldots, J$;
- Simule c_j de $c_j|(\cdot)$. $j = 1, \ldots, J$,

4.7 Comentários finais do capítulo

Neste capitulo descrevemos vários algoritmo MCMC para estimação de parâmetros de item e traços latentes, concluindo que o esquema de dados aumentados de Sahu (2002) com a proposta de aceleração de convergência de Gonzalez (2004) era o melhor em termos do ESS. Neste capitulo também apresentamos o algoritmo MCMC para o modelo D-MP3NAC.

Capítulo 5

Estudo de Simulação

5.1 Introdução

Neste capítulo apresentamos os resultados de estudos de simulação realizados com intuito de avaliar a acurácia das estimativas e, também, o impacto de importantes fatores na recuperação dos verdadeiros valores dos parâmetros. Realizamos este estudo utilizando o esquema de dados aumentados de Sahu (2002) com o passo de acelerado de convergência de Gonzalez (2004), algoritmo ADGS. Esta escolha foi baseada nos resultados obtidos na no Capítulo anterior, que apontaram este algoritmo como sendo o de melhor performance, em termos de convergência.

A estimação bayesiana utilizando métodos de Monte Carlo baseados na simulação de cadeias de Markov (MCMC; Gelfand e Smith (1990), Geman e Geman (1984)), representam uma excelente alternativa aos métodos clássicos via Maxima Verossimilhança Marginal (MVM; Darrell Bock e Lieberman (1970), Bock e Aitkin (1981)) no contexto da TRI, particularmente em modelos complexos. Nesta seção consideramos diversas situações de interesse prático, formadas pela escolha das distribuições dos traços latentes, numero de itens (tamanho do teste), dimensionalidade do teste, números de indivíduos em cada grupo, os quais compartilham diferentes percentuais de itens comuns.

De modo a comparar as diferentes situações, devem-se usar estatísticas convenientes. Seja $\boldsymbol{\vartheta}_l \in (\theta_{dj}, \alpha_{di}, b_i)$, onde l é um índice conveniente (i, j ou k) e $\widehat{\vartheta}_{lr}$ sua respectiva estimativa obtida na réplica r, $r = 1, \ldots, R$. Defina também $\widehat{\overline{\vartheta}}_l = \frac{1}{R}\sum_{r=1}^{R} \widehat{\vartheta}_{lr}$. A raiz quadrada do erro quadrático médio (REQM) e o viés relativo absoluto (VRA) são definidos da seguinte forma:

$$\text{REQM} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\widehat{\vartheta}_{lr} - \vartheta_l)^2} \text{ e}$$

$$\text{VRA} = \frac{|\widehat{\overline{\vartheta}}_l - \vartheta_l|}{|\vartheta|}.$$

As respostas aos itens geradas sob as diferentes condições serão analisadas sob o modelo simétrico e o modelo assimétrico proposto no capítulo 4. Em vista disso, espera-se detectar as

vantagens e/ou desvantagens ao se adotar tais modelos no processo de estimação, quando a verdadeira distribuição dos traços latentes possui diferentes comportamentos, além de medir a influência dos fatores na acurácia associada as estimativas. Conduzir-á o estudo para um conjunto de $R = 10$ replicas, considerando os seguintes fatores (e seus níveis):

- **Único grupo:**
  - Número de respondentes (NR): 500 e 2000,
  - Número de itens (NI): 20 e 40;
  - Número de dimensões (ND): 2 e 3;
  - Distribuição verdadeira dos traços latentes: $N_D(\mathbf{0}, \boldsymbol{I})$, $N_D(\mathbf{0}, \boldsymbol{R})$, $NAC_D(\mathbf{0}, \boldsymbol{I}, \boldsymbol{\lambda})$ e $NAC_D(\mathbf{0}, \boldsymbol{R}, \boldsymbol{\lambda})$;
  - Modelos: simétrico e assimétrico.

- **Grupos múltiplos (três grupos):**
  - Número de respondentes por grupo (NR): 500 e 1000,
  - Número de itens por grupo (NI): 20 e 40;
  - Número de dimensões (ND): 2 e 3;
  - Percentual de itens comuns entre os testes (NIC): 25% e 50%;
  - Distribuição verdadeira dos traços latentes: $N_D(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $N_D(\mathbf{0}, \boldsymbol{R})$, $NAC_D(\mathbf{0}, \boldsymbol{I}, \boldsymbol{\lambda}_k)$ e $NAC_D(\mathbf{0}, \boldsymbol{R}_k, \boldsymbol{\lambda}_k)$;
  - Modelos: simétrico e assimétrico.

De forma, totaliza-se 32 situações simuladas para o caso de único grupo e 64 para o caso de grupos múltiplos. Os parâmetros populacionais considerados, para o caso de um único grupo, foram os seguintes:

$$\boldsymbol{\delta} = \begin{pmatrix} -0.7 \\ 0.7 \end{pmatrix} \text{ e } \boldsymbol{\delta} = \begin{pmatrix} -0.7 \\ 0 \\ 0.7 \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0,9 \\ 0,9 & 1 \end{pmatrix} \text{ e } \boldsymbol{R} = \begin{pmatrix} 1 & 0,9 & 0,9 \\ 0,9 & 1 & 0,7 \\ 0,9 & 0.7 & 1 \end{pmatrix}$$

lembrando que $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^t\boldsymbol{\lambda}}}$, é o nosso parâmetro de assimetria de interes. Os parâmetros populacionais considerados, para o caso de grupos múltiplos, foram os seguintes:

$$\boldsymbol{\delta}_1 = \begin{pmatrix} -0,5 \\ 0,7 \end{pmatrix},\ \boldsymbol{R}_1 = \begin{pmatrix} 1,2 & 0,8 \\ 0,8 & 1,2 \end{pmatrix}$$
$$\boldsymbol{\delta}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \boldsymbol{R}_2 = \begin{pmatrix} 1,0 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\boldsymbol{\delta}_3 = \begin{pmatrix} 0,7 \\ -0,5 \end{pmatrix},\ \boldsymbol{R}_3 = \begin{pmatrix} 1,8 & 0,95 \\ 0,95 & 1,8 \end{pmatrix}$$

o grupo de referência considerado foi o segundo.

5.2 Análises das estimativas de único grupo

Os resultados desse estudo foram resumidos através dos gráficos das estatísticas REQM (Figuras 5.1 a 5.9), calculados para cada conjunto de parâmetros. Pelos resultados, de uma forma geral, nota-se que o algoritmo de DAMHGS gerou estimativas próximas dos valores originais de cada parâmetro.

O modelo assimétrico obteve vantagem sobre o simétrico na maioria das situações, especialmente, nas quais a distribuição dos traços latentes apresentava certa assimetria. Além disso, verifica-se que com o aumento do número dos indivíduos, a imprecisão associada as estimativas dos parâmetros dos itens diminui. O contrário ocorre quando se aumenta o numero de itens ou categorias, neste caso, para um mesmo conjunto de traços latentes, suas estimativas tornam-se menos acuradas. Já com relação aos traços latentes, o aumento nos itens contribui positivamente na sua precisão

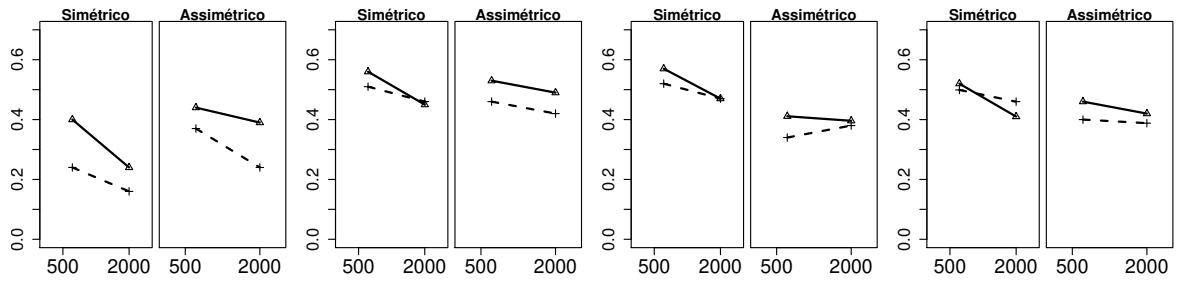


Figura 5.1: REQM das estimativas dos parâmetros α_1. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

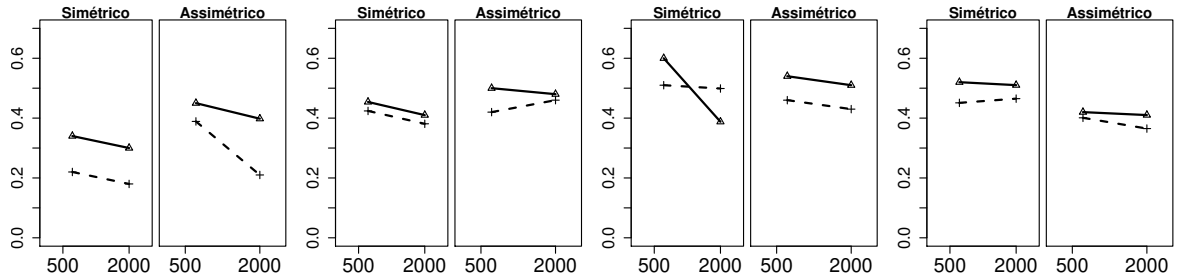


Figura 5.2: REQM das estimativas dos parâmetros α_2. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

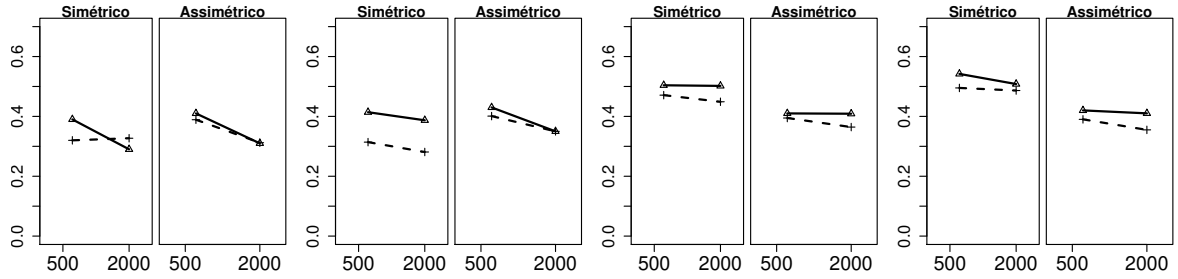


Figura 5.3: REQM das estimativas dos parâmetros β. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

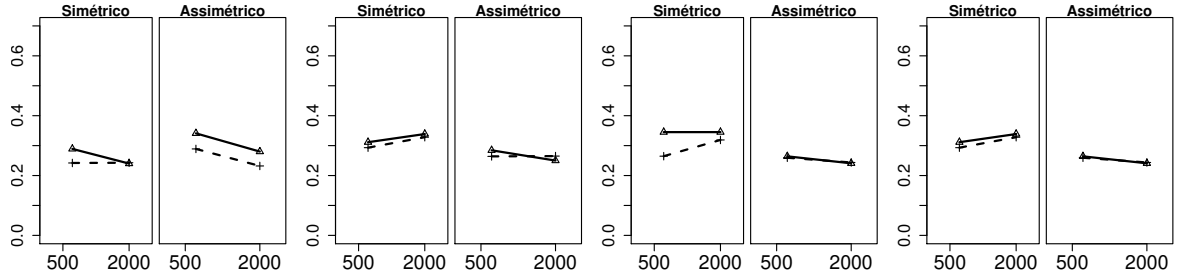


Figura 5.4: REQM das estimativas dos parâmetros c. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

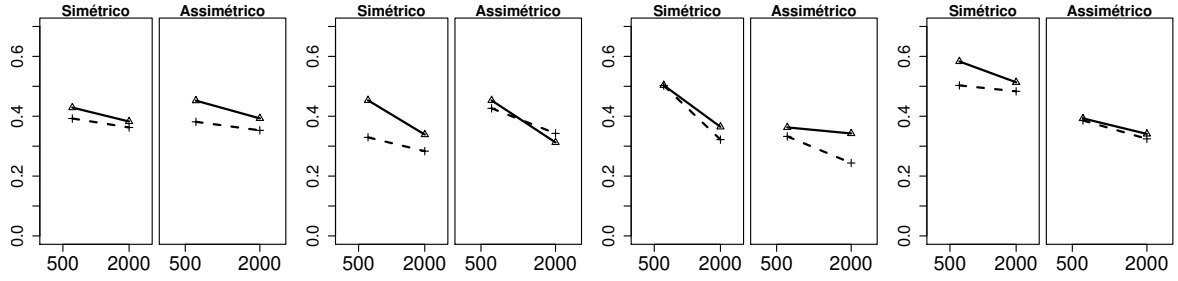


Figura 5.5: REQM das estimativas dos parâmetros α_1. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

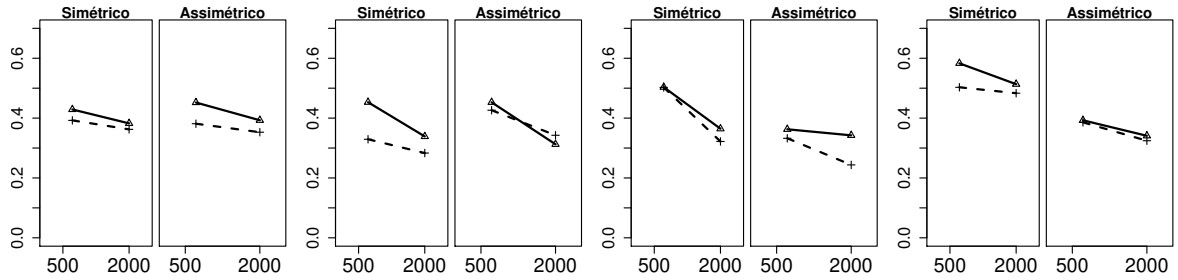


Figura 5.6: REQM das estimativas dos parâmetros α_2. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

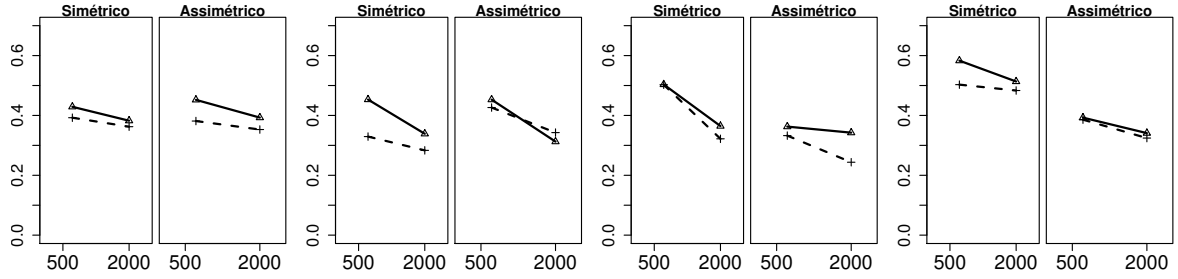


Figura 5.7: REQM das estimativas dos parâmetros α_3. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

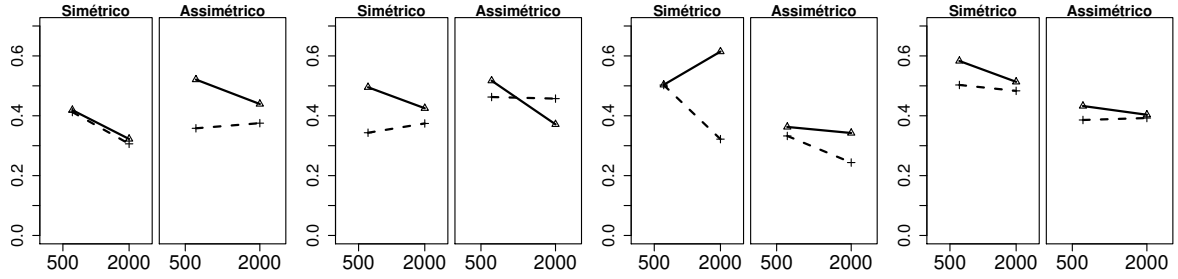


Figura 5.8: REQM das estimativas dos parâmetros β. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

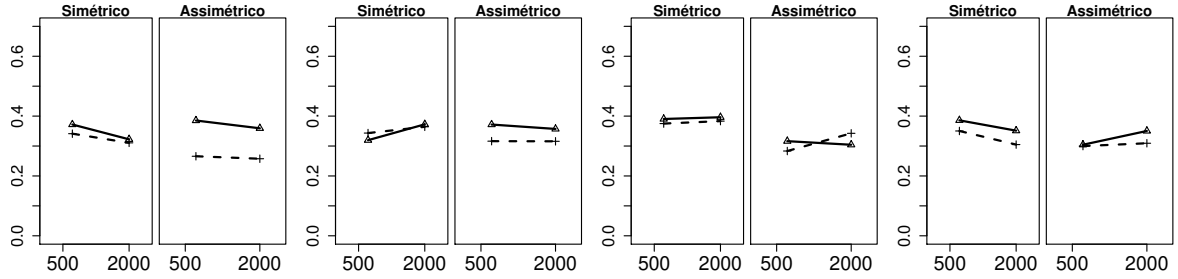


Figura 5.9: REQM das estimativas dos parâmetros c. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

Nas Tabelas 5.1 a 5.5 são apresentados as medias do parâmetro de assimetria, em geral, nota-se uma razoável recuperação dos valores originais.

Tabela 5.1: Média das estimativas do parâmetro de assimetria δ_1. Modelos de 2 dimensões (o valor verdadeiro será 0 para as distribuições simétricas e para as distribuições assimétricas, o verdadeiro valor é $-0,7$)

	N(0,I)	N(0,I)	N(0,R)	N(0,R)	NAC(0,I)	NAC(0,I)	NAC(0,R)	NAC(0,R)
	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens
500	-0,043	-0,065	0,003	0,01	-0,65	-0,73	-0,76	-0,63
2000	-0,08	0,051	0,018	0,008	-0,61	-0,76	-0,73	-0,78

Tabela 5.2: Média das estimativas do parâmetro de assimetria δ_2. Modelos de 2 dimensões (o valor verdadeiro será 0 para as distribuições simétricas e para as distribuições assimétricas, o verdadeiro valor é 0,7)

	N(0,I)	N(0,I)	N(0,R)	N(0,R)	NAC(0,I)	NAC(0,I)	NAC(0,R)	NAC(0,R)
	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens
500	0,062	0,03	-0,021	0,07	0,63	0,65	0,62	0,66
2000	0,01	-0,041	-0,003	0,058	0,67	0,69	0,73	0,63

Tabela 5.3: Média das estimativas do parâmetro de assimetria δ_1. Modelos de 3 dimensões (o valor verdadeiro será 0 para as distribuições simétricas e para as distribuições assimétricas, o verdadeiro valor é $-0,7$)

	N(0,I)	N(0,I)	N(0,R)	N(0,R)	NAC(0,I)	NAC(0,I)	NAC(0,R)	NAC(0,R)
	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens
500	0,03	-0,092	0,019	-0,03	-0,79	-0,77	-0,74	-0,75
2000	0,012	-0,101	-0,025	0,002	-0,81	-0,79	-0,77	-0,76

Tabela 5.4: Média das estimativas do parâmetro de assimetria δ_2. Modelos de 3 dimensões (o valor verdadeiro será 0 para as distribuições simétricas e para as distribuições assimétricas, o verdadeiro valor é 0)

	N(0,I)	N(0,I)	N(0,R)	N(0,R)	NAC(0,I)	NAC(0,I)	NAC(0,R)	NAC(0,R)
	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens
500	0,096	0,073	0,054	0,013	-0,109	-0,12	-0,127	-0,11
2000	0,047	0,046	0,015	0,023	-0,12	-0,102	-0,125	-0,10

Tabela 5.5: Média das estimativas do parâmetro de assimetria δ_3. Modelos de 3 dimensões (o valor verdadeiro será 0 para as distribuições simétricas e para as distribuições assimétricas, o verdadeiro valor é 0,7)

	N(0,I)	N(0,I)	N(0,R)	N(0,R)	NAC(0,I)	NAC(0,I)	NAC(0,R)	NAC(0,R)
	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens	20 Itens	40 Itens
500	0,093	0,042	0,050	0,042	0,688	0,742	0,726	0,683
2000	0,021	0,005	0,089	0,033	0,672	0,626	0,663	0,766

5.3 Grupos múltiplos

Novamente, através da análise das Figuras 5.10 a 5.13 nota-se que a raiz do erro quadrático médio das estimativas dos parâmetros dos itens diminui com o aumento do teste, ao contrario das estimativas dos traços latentes, que apresentam um ligeiro aumento. Nas Figuras 5.14 a 5.18 apresenta-se a situação onde se tem 3 dimensões e uma estrutura do 25% de itens em comum. Os resultados para o 50% de itens em comum mantiveram o mesmo comportamento descrito, apresentando uma redução geral do erro quadrático médio.

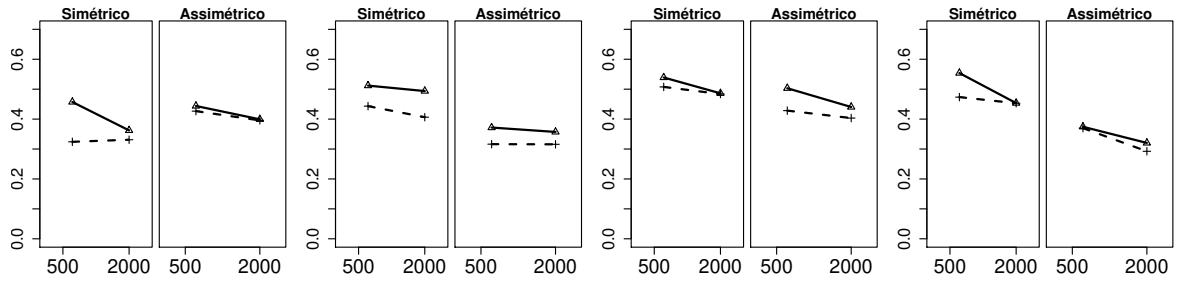


Figura 5.10: REQM das estimativas dos parâmetros α_1, com 25% de itens comuns. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

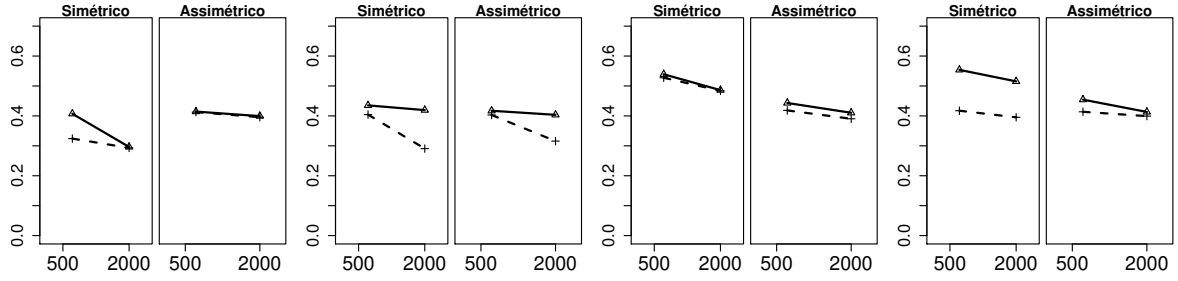


Figura 5.11: REQM das estimativas dos parâmetros α_2, com 25% de itens comuns. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

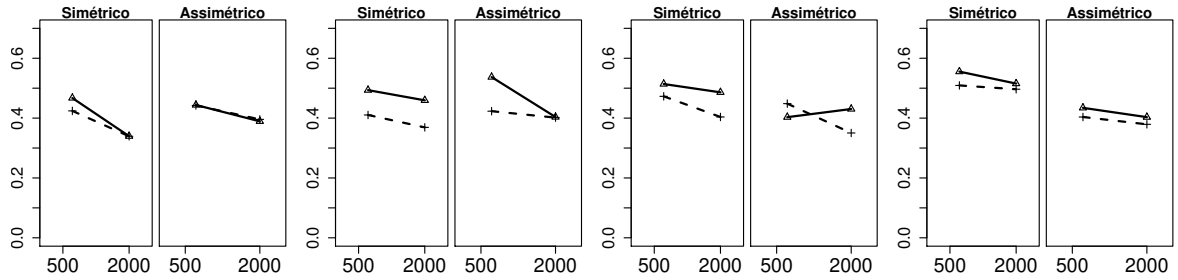


Figura 5.12: REQM das estimativas dos parâmetros β, com 25% de itens comuns. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

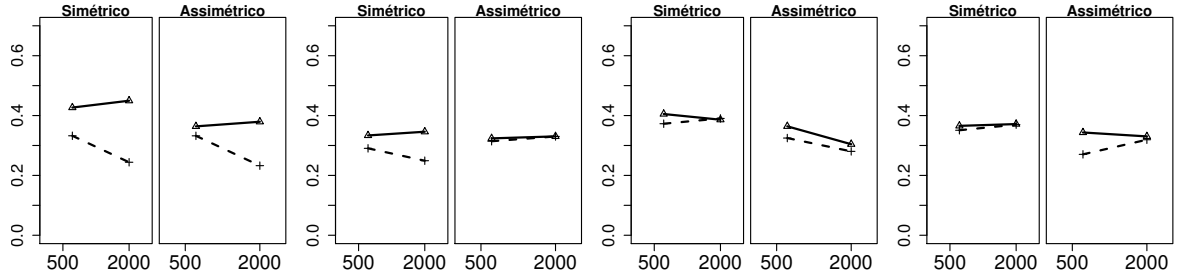


Figura 5.13: REQM das estimativas dos parâmetros c, com 25% de itens comuns. Ordem de esquerda a direita: $N_2(0, \boldsymbol{I})$, $N_2(0, \boldsymbol{R})$, $NAC_2(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_2(0, \boldsymbol{R}, \boldsymbol{\lambda})$

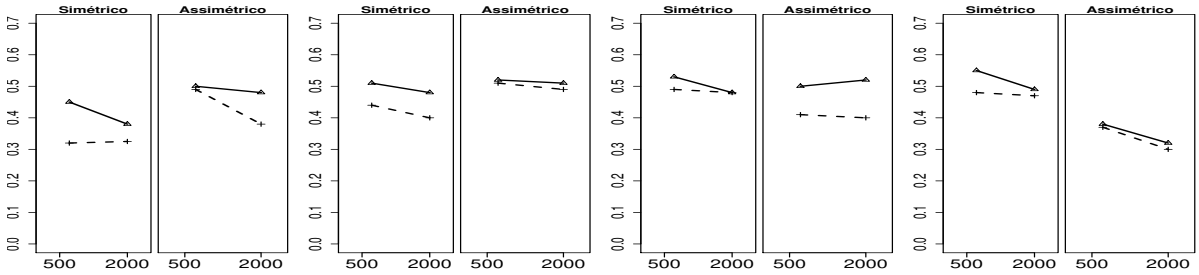


Figura 5.14: REQM das estimativas dos parâmetros α_1, com 25% de itens comuns. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

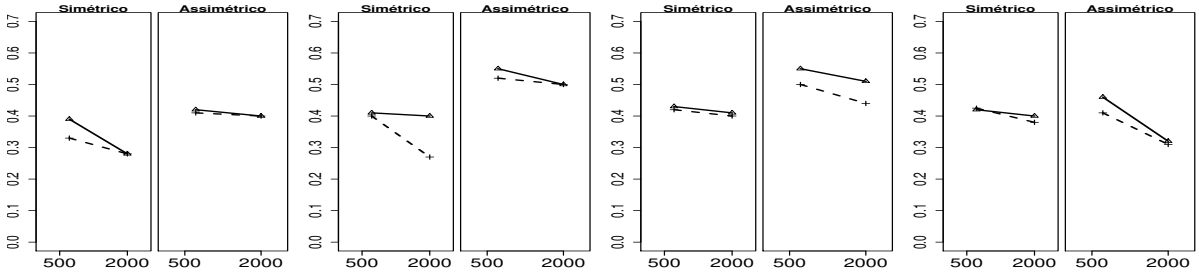


Figura 5.15: REQM das estimativas dos parâmetros α_2, com 25% de itens comuns. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

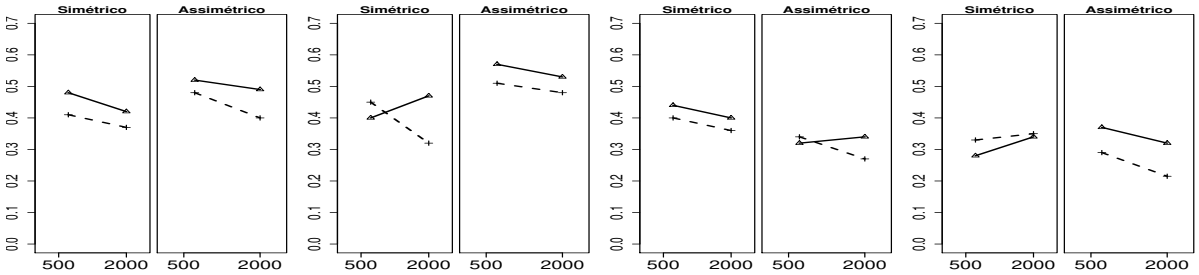


Figura 5.16: REQM das estimativas dos parâmetros α_3, com 25% de itens comuns. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

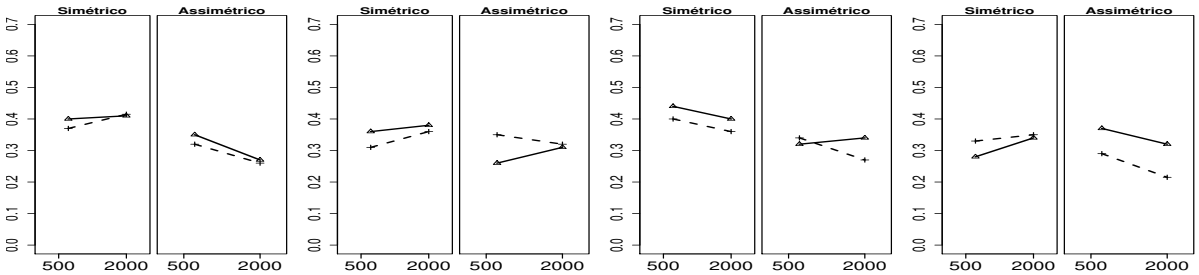


Figura 5.17: REQM das estimativas dos parâmetros β, com 25% de itens comuns. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

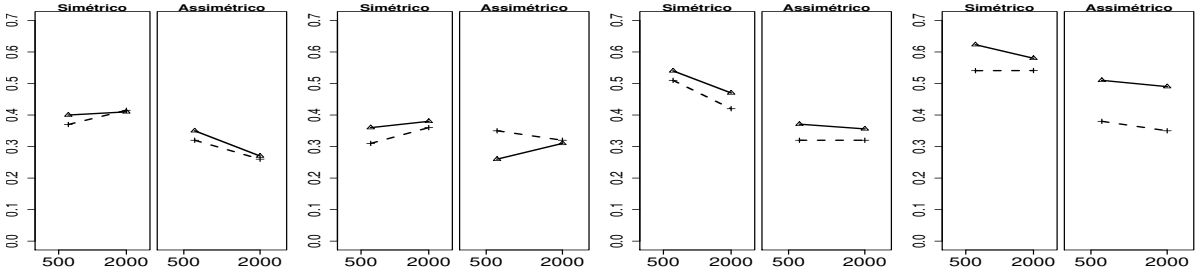


Figura 5.18: REQM das estimativas dos parâmetros c, com 25% de itens comuns. Ordem de esquerda a direita: $N_3(0, \boldsymbol{I})$, $N_3(0, \boldsymbol{R})$, $NAC_3(0, \boldsymbol{I}, \boldsymbol{\lambda})$,$NAC_3(0, \boldsymbol{R}, \boldsymbol{\lambda})$

5.4 Comentários e conclusões

Após ter concluído a etapa de simulação podemos concluir que o modelo assimétrico se ajusta razoavelmente bem a todas as situações consideradas mostrando a versatilidade do modelo. Para o coeficiente de assimetria notamos que foi bem recuperado independentemente de se for nulo ou não. A grande dificuldade apresentada no processo de simulação, foi os grandes tempos de computação, sendo necessárias varias semanas para a conclusão dos algoritmos. Porem este fato é uma das grandes dificuldades dos algoritmos MCMC em geral, abrindo a possibilidade a considerar outras técnicas de estimação com custos computacionais menores.

Capítulo 6

Análise dos dados do vestibular da Unicamp

6.1 Introdução

Neste Capítulo utilizaremos os modelos de grupos múltiplos assimétrico e simétrico considerando: uma, duas e três dimensões, tendo ao todo 6 modelos, em uma análise de dados reais. A implementação é feita considerando uma amostra referente à primeira fase do vestibular da Universidade estadual de Campinas (UNICAMP) do ano 2013. O presente capítulo está dividido da seguinte forma: começamos fazendo uma breve descrição dos dados reais, assim como das características do teste. Na seção 6.2 analisamos os dados, comparando os resultados obtidos após àlicação dos seis modelos considerados e dedicamos uma subseção para apresentar algumas medidas de diagnostico. Avaliamos a qualidade de ajuste do modelo escolhido, utilizando mecanismos de diagnóstico baseados na distribuição preditiva de medidas de discrepância adequadas. Por último, na seção 6.5 apresentamos conclusões e comentários.

6.2 Vestibular UNICAMP 2013

O Vestibular nacional da UNICAMP, aplicado pela Comissão Permanente para os Vestibulares (COMVEST), classifica e seleciona candidatos para a matrícula inicial na UNICAMP e nos cursos de Medicina e Enfermagem da Faculdade de Medicina de São José do Rio Preto (Famerp), toda a informação aqui apresentada é extraída do site da COMVEST (*www.comvest.unicamp.br*), se pode entrar mais detalhes.

O vestibular da UNICAMP está constituído por duas fases. A primeira fase, a qual é obrigatória para todos os candidatos, é constituída de uma única prova composta de duas partes:

- Redação,
- Conhecimentos gerais.

Na parte de conhecimentos gerais, são apresentadas 48 questões de múltipla escolha relacionadas à diferentes áreas do conhecimento desenvolvidas no ensino médio. Cada questão da parte de conhecimentos gerais vale um ponto e os resultados dessa prova serão analisados.

No ano 2013, os candidatos podiam escolher entre 20 locais para se submeterem à prova da primeira fase. Foram inscritos 67.408 pessoas, que concorreram a 3.320 vagas na UNCIAMP, divididas em 68 opções de curso, e 144 vagas na FAMERP, divididas em duas opções de curso. Dos 67.408 inscritos só 60.806 se apresentaram ao local da prova e responderam todas as 48 perguntas. Para realizar uma analise descritiva dos dados e posteriormente ajustar os modelos desenvolvidos no Capítulo 4, agrupamos os indivíduos inscritos segundo o código da diretoria acadêmica da UNICAMP (DAC) do curso da primeira opção. . Tal agrupamento foi realizado pois espera-se que os indivíduos apresentem uma maior variabilidade entre esses grupos do que internamente. Os grupos são, a saber: artes, ciências biológicas e da saúde, ciências exatas e da terra, ciências humanas, medicina e tecnológicas. Espera-se, por exemplo, espera-se que a média das habilidades dos inscritos no curso de medicina seja maior que para os inscritos e, outros cursos, e também se espera que para os inscritos a medicina, as habilidades apresentem algum tipo de assimetria, já que pode acontecer o fato que para este curso se inscrevam indivíduos com níveis de traços latentes altos em geral.

Na Tabela 6.1 apresentamos algumas medidas de resumo, relativas aos escores observados, para os seis grupos. Podemos ver como o escore dos inscritos aos cursos relacionados com a area de medicina são maiores com respeito aos outros grupos. No entanto, as médias dos escores dos grupos estão muito próximas.

Tabela 6.1: Estatísticas descritivas escores observados dados COMVEST 2013

Área DAC	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	Indivíduos
Artes	6	20	25	24,85	29	44	2555
Biológicas e saúde	4	20	24	24,26	28	46	6013
Exatas e da terra	5	23	28	28,5	34	48	31648
Humanas	5	21	25	25,79	30	46	6838
Medicina	7	25	32	30,92	38	48	12779
Tecnológicas	9	19	23	23,38	27	44	973

Complementamos a informação da Tabela 6.1, com as distribuições dos escores para cada grupo, apresentadas na Figura 6.1. Nesta figura, é possível ver como, para alguns grupos, o pressuposto de assimetria parece não ser valido, em particular para os grupos referentes a "Medicina" e "Humanas", fato que motiva a implementação de modelos que tomem em conta a assimetria nos traços latentes.

Para a implementação dos modelos, tomamos uma amostra em cada grupo, a fim de obter um tamanho total de tamanho 2000 indivíduos. A amostragem, aleatória simples sem reposição, foi feita de tal forma que a proporção dos inscritos em cada grupo com relação ao total de inscritos fosse conservada. Na Figura 6.7 apresentam-se as distribuições de escores. Para cada grupo da amostra, podemos ver como o comportamento assimétrico dos indivíduos presentes na população é mantido na amostra.

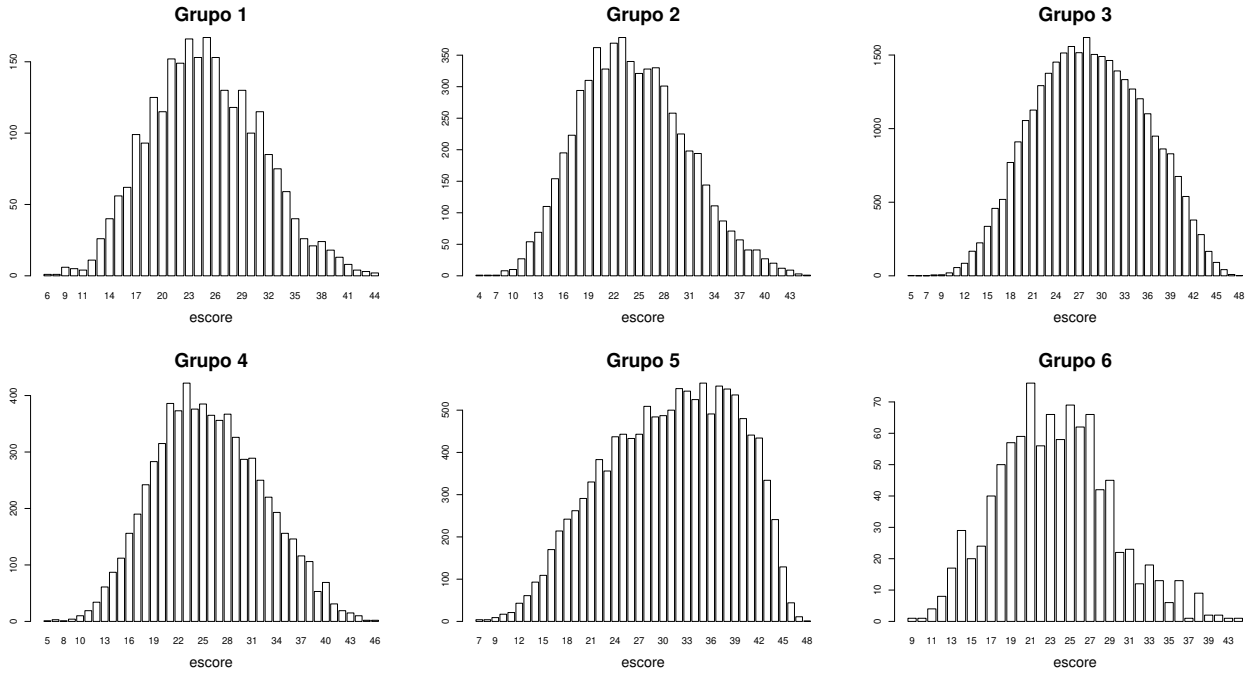


Figura 6.1: Escores observados por grupo

Na Tabela 6.2 apresentamos os tamanhos da amostra em relação aos grupos considerados.

Tabela 6.2: Tamanho de cada grupo na amostra de 2000 indivíduos

Grupo	Porcentagem da amostra	Tamanho do grupo
Grupo 1 (Artes)	4,20%	84
Grupo 2 (Biológicas e saúde)	9,89%	198
Grupo 3 (Exatas e da terra)	52,04%	1041
Grupo 4 (Humanas)	11,24%	225
Grupo 5 (Medicina)	21,01%	420
Grupo 6 (Tecnológicas)	1,6%	32

Nas Figuras 6.3 a 6.6 encontram-se os gráficos de colunas relativos aos percentuais de alunos que escolheram cada categoria, por item, no qual o gabarito é a coluna em destaque. Procedendo-se uma análise previa, detectou-se que no item 43 a resposta correta teve menor proporção de respostas que qualquer outra das opções. Este fato pode significar várias coisas, por exemplo, que o item não é adequado, ou que ele foi respondido de forma displicente pelos candidatos. Por este motivo o item 43 foi desconsiderado nos ajustes dos modelos.

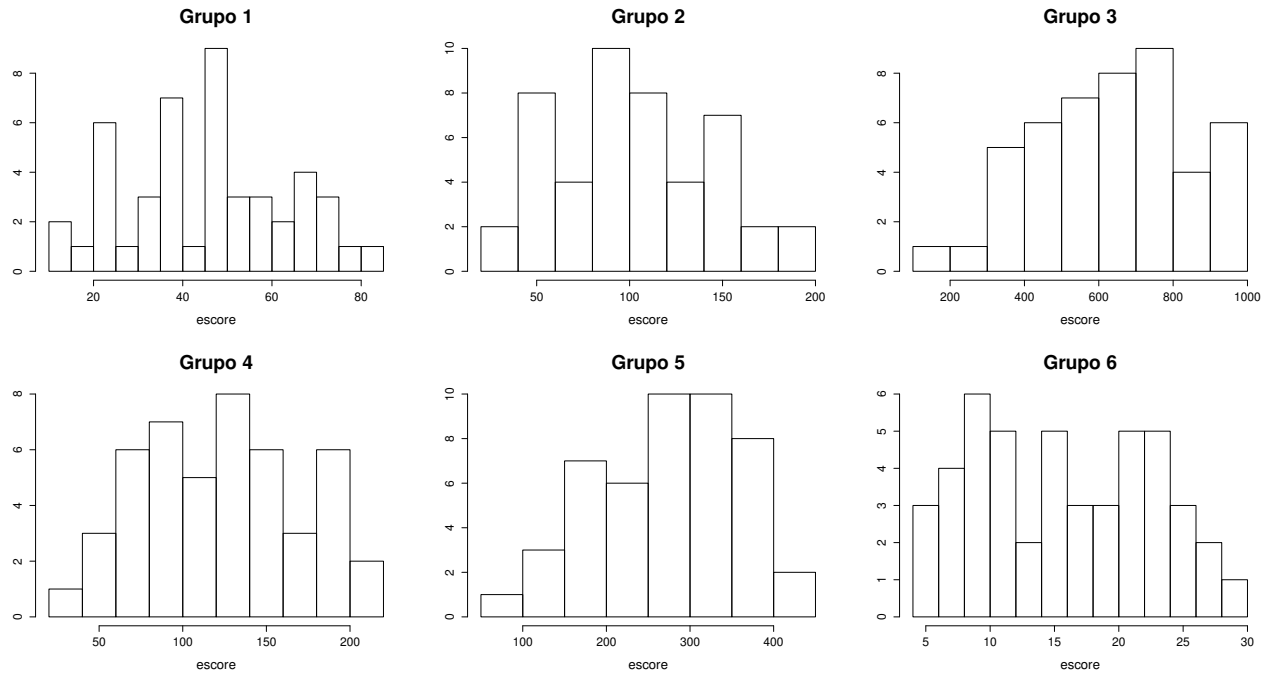


Figura 6.2: Escores observados por grupo, dados da amostra

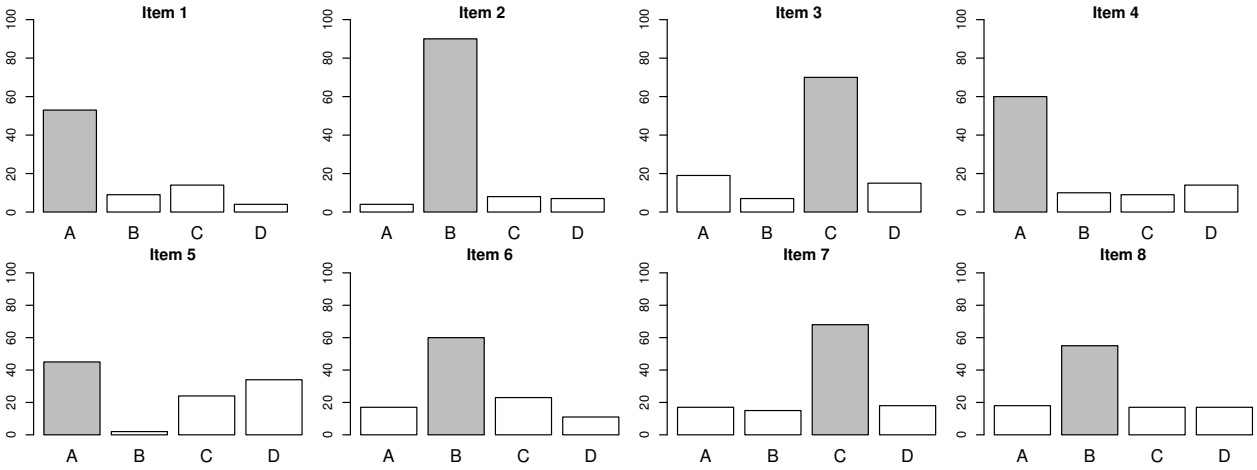
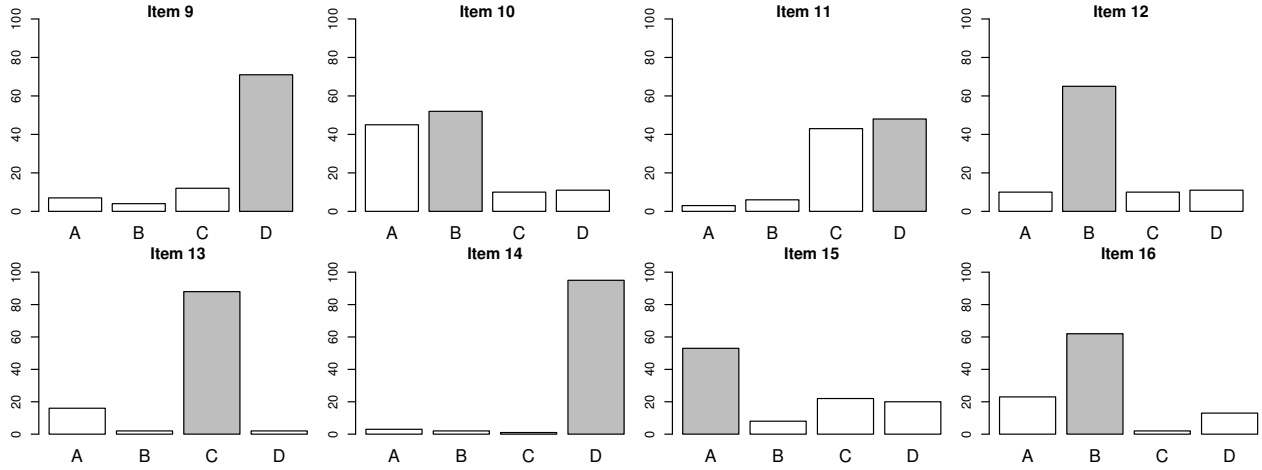
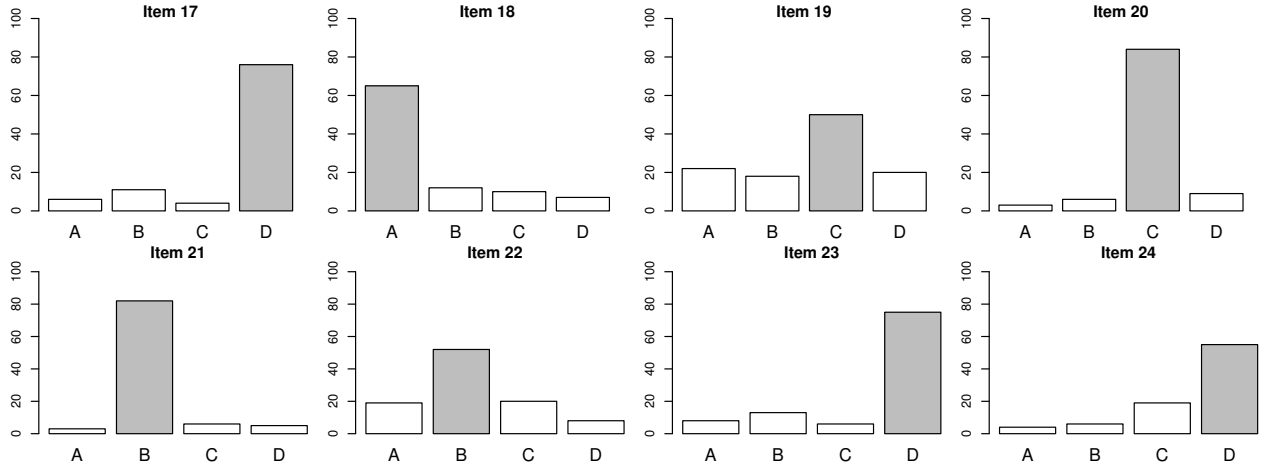


Figura 6.3: Histograma dos itens.

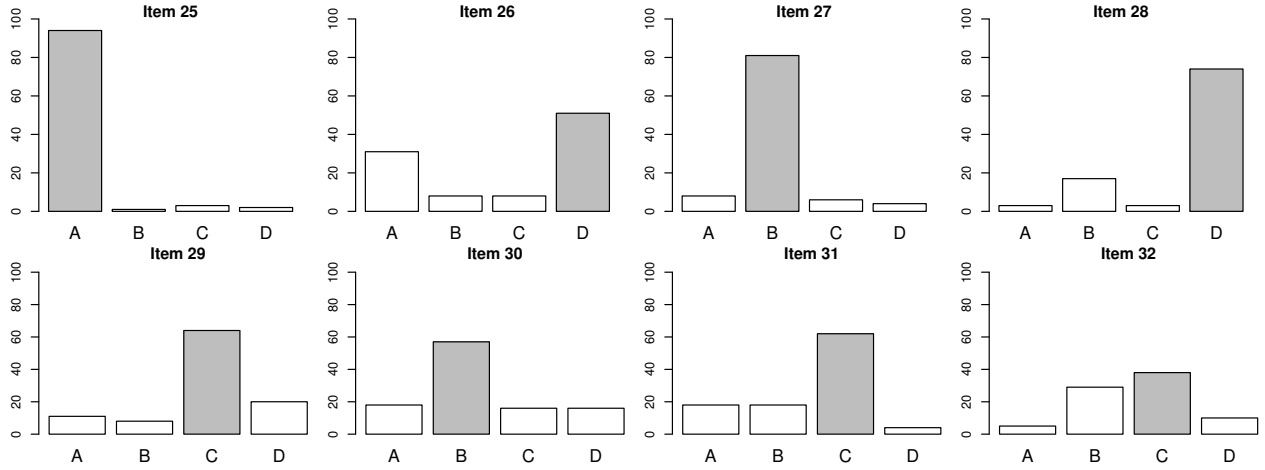


(a) Itens 9 a 16

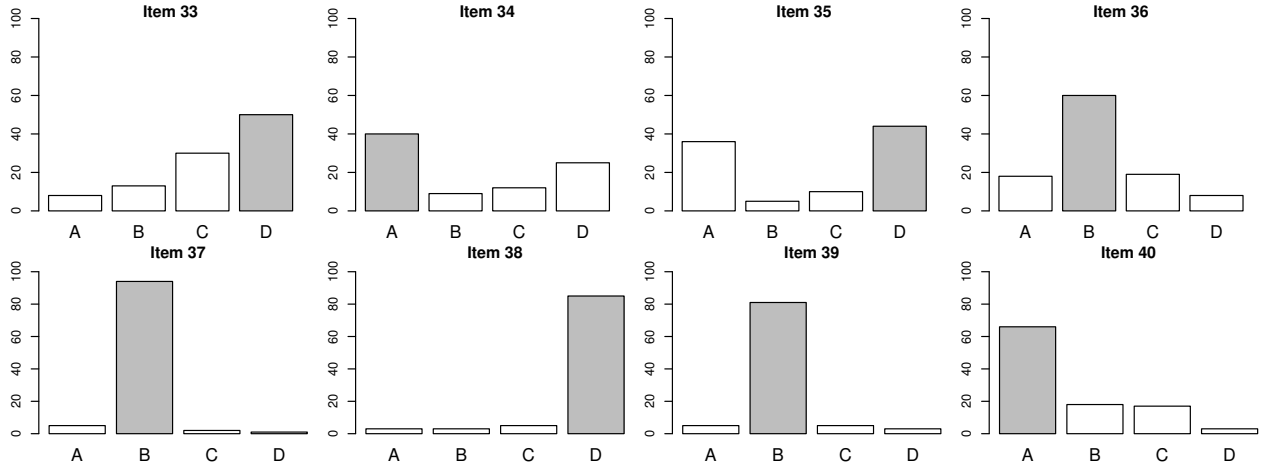


(b) Itens 17 a 24

Figura 6.4: Histograma dos itens.



(a) Itens 25 a 32



(b) Itens 33 a 40

Figura 6.5: Histograma dos itens.

6.3 Comparação dos modelos

Implementamos seis modelos, considerando diferentes dimensões, bem como diferentes distribuições, ambas relativas aos traços latentes. Os modelos considerados se resumem na Tabela 6.3

Deve-se notar que estamos considerando um caso particular, em termos de grupos múltiplos onde temos 6 grupos não balanceados, no qual as provas respondidas pelos diferentes grupos são compostas pelos mesmos 47 itens. É importante lembrar, que o parâmetro de assimetria que estamos modelando, na notação do capitulo 2, é o parâmetro $\boldsymbol{\delta}$, isto é, o parâmetro cujas componentes oscilam entre -1 e 1, e que no caso univariado está intrinsecamente relacionado ao coeficiente de assimetria de Pearson.

De um ponto de vista estatístico, para selecionar o modelo mais adequado, utilizamos estatísticas de ajuste consideradas por Bazán et al. (2006) e Azevedo et al. (2011). As quais foram

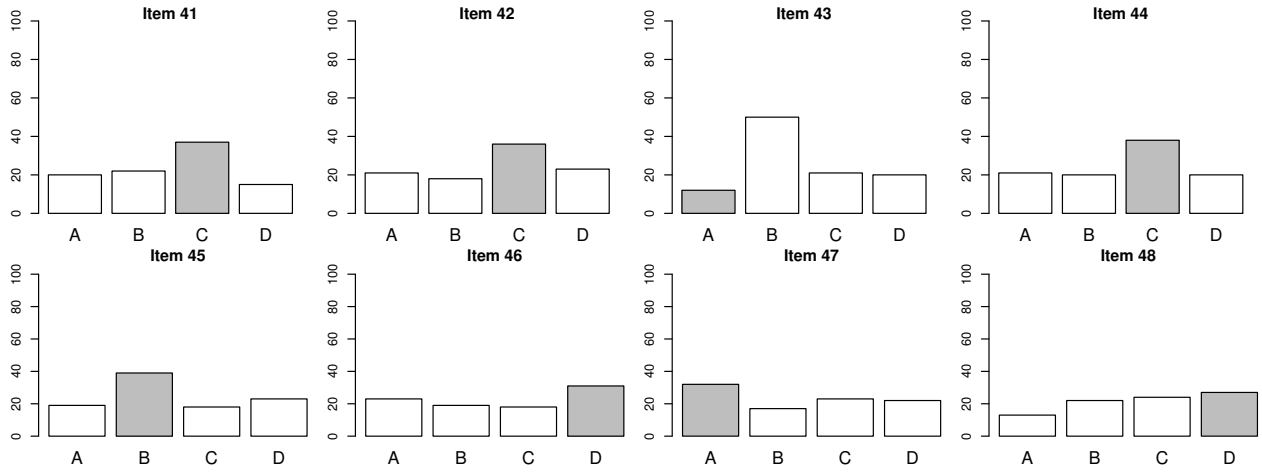


Figura 6.6: Histograma dos itens.

Tabela 6.3: Modelos considerados dados COMVEST 2013

Nome do modelo	Dimensão	Distribuição traços latentes
1-MP3PNS	1	Normal simétrica
1-MP3PNAC	1	Normal assimétrica centrada
2-MP3PNS	2	Normal simétrica
2-MP3PNAC	2	Normal assimétrica centrada
3-MP3PNS	3	Normal simétrica
3-MP3PNAC	3	Normal assimétrica centrada

estudadas em detalhe por Spiegelhalter et al. (2002) e estão baseadas na noção do "Desvio Bayesiano", veja Dempster (1997).

O desvio bayesinao $D(\boldsymbol{\vartheta})$ é definido como -2log-verossimilhança. No nosso caso temos que

$$D(\boldsymbol{\vartheta}) = -2ln(L(\boldsymbol{\theta}_{\cdots}, \boldsymbol{\zeta}_{\cdot}, \boldsymbol{c}, \boldsymbol{\Omega})) = -2ln(L(\boldsymbol{\theta}_{\cdots}, \boldsymbol{\zeta}_{\cdot}, \boldsymbol{c})p(\boldsymbol{\theta}_{\cdots}|\boldsymbol{\Omega})). \tag{6.3.1}$$

É importante a introdução da estatística ρ_D, a qual é definida como:

$$\rho_D = \overline{D(\vartheta)} - D(\overline{\vartheta}), \tag{6.3.2}$$

na prática, podemos utilizar amostras MCMC para estimar ρ_D. Por exemplo

$$\widehat{\overline{D(\boldsymbol{\vartheta})}} = \frac{1}{T}\sum_{t=1}^{T} D(\boldsymbol{\vartheta}^{(t)}) \tag{6.3.3}$$

em que o índice t representa a t-ésima realização de um total de T realizações simuladas, em termos da amostra válida MCMC. Para estimar $D(\overline{\vartheta})$ utilizamos as médias a posteriori dos parâmetros. Isto leva ao critério DIC, o qual penaliza o esperado do desvio bayesiano pela complexidade do

modelo, ou número de parâmetros efetivos (ρ_D),

$$\widehat{DIC} = D(\overline{\vartheta}) + 2\widehat{\rho}_D. \tag{6.3.4}$$

Outro critério de informação considerado o Esperado do Critério de Informação de Akaike (EAIC), ver Akaike (1973), definido por

$$\widehat{EAIC} = \overline{D(\boldsymbol{\vartheta})} + 2p, \tag{6.3.5}$$

e ao Esperado do Critério de Informação (Schwarz) Bayesiano (EBIC) definido por

$$\widehat{EBIC} = \overline{D(\boldsymbol{\vartheta})} + plog(N_D), \tag{6.3.6}$$

em que p é o numero de parâmetros e N_D é o número de observações, que no nosso caso, corresponde ao número total de respostas, ou seja, $N_D = N \times J$. Os resultados podem ser vistos na Tabela 6.4. Estes indicam que em geral os modelos assimétricos se ajustaram melhor os dados que suas versões simétricas, segundo todos os critérios de comparação. Os critérios $\widehat{DIC}$ e $\widehat{EAIC}$ selecionam como melhor modelo o 2-MP3NAC, porem o critério $\widehat{EBIC}$ escolhe o 1-MP3NAC.

Tabela 6.4: Critérios para comparação de modelos

Modelo	$\widehat{DIC}$	$\widehat{EAIC}$	$\widehat{EBIC}$
1-MP3NS	119547,9	123787,6	134868,8
1-MP3NAC	116663,8	119351,4	131398,9
2-MP3NS	123232,1	123778,5	124982,7
2-MP3NAC	118750,8	119169,1	120862,1
3-MP3NS	126993,3	154453,6	169604,9
3-MP3NAC	122597,5	150457,5	155709,6

Com o intuito de ter evidências, de um ponto de vista mais próximo da definição de dimensionalidade, realizamos uma checagem preditiva à posteriori do modelo 1-MP3NAC (Figura 6.7). Nela podemos identificar uma dimensão dominante para quase todos os grupos, porem o suposto de duas dimensões parece ser mais razoável em dois grupos. Este fato, junto com o obtido em base aos critérios de informação nós leva à conclusão que o modelo que melhor ajusta aos dados é o 2-MP3NAC.

6.4 Avaliação da qualidade do ajuste

Uma vez escolhido o modelo procedemos a avaliar à avaliação do ajuste do mesmo. Particularmente o ajuste do modelo por item, para o qual será necessário a introdução de algumas medidas de discrepância.

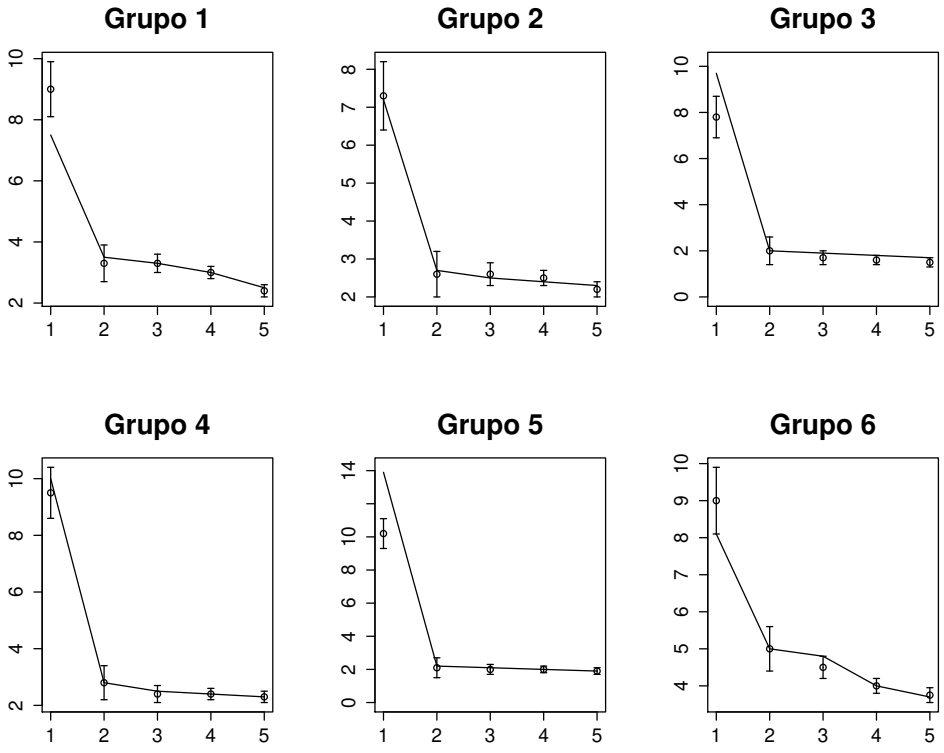


Figura 6.7: Autovalores matriz correlações tetracórica. Legenda: linha sólida, ∘ dados modelo 1-MP3NAC com seus intervalos de credibilidade

6.4.1 Medidas de diagnósticos

Uma forma de verificar a qualidade do ajuste, do ponto de vista de modelos bayesianos, é comparar a distribuição preditiva com a distribuição dos dados observados, veja Gelman et al. (2013). Azevedo, Bolfarine e Andrade (2012) adaptaram alguns diagnósticos para o caso de grupos múltiplos e uma dimensão, e posteriormente foram estendidos para modelos assimétricos de grupos múltiplos em Santos, Azevedo e Bolfarine (2013) e Santos (2012). Tais diagnósticos são baseados em medidas de discrepância escolhidas de modo que se possa avaliar alguma suposição ou o ajuste geral do modelo. Algumas notações importantes são agora definidas para o desenvolvimento da seção.

Seja $\boldsymbol{y}^{obs}$ a matriz de respostas observadas, e $\boldsymbol{y}^{rep}$ a matriz de dados replicados gerados a partir da sua distribuição preditiva. A distribuição preditiva das respostas do grupo k é representada por

$$p(\boldsymbol{y}_k^{rep}|\boldsymbol{y}_k^{obs}) = \int p(\boldsymbol{y}_k^{rep}|\boldsymbol{\vartheta}_k)p(\boldsymbol{\vartheta}_k|\boldsymbol{y}_k^{obs})d\boldsymbol{\vartheta}_k, \tag{6.4.1}$$

onde $\boldsymbol{\vartheta}_k$ denota o conjunto de parâmetros do modelo correspondentes ao grupo k. Uma opção é o valor-p bayesiano, que para uma medida de discrepância adotada, define-se como a probabilidade

de observarmos dados replicados maiores do que, ou iguais aos dados observados. Ou seja

$$p_0(\boldsymbol{y}_k^{rep}|\boldsymbol{y}_k^{obs}) = p\Big(D(\boldsymbol{y}_k^{rep}, \boldsymbol{\vartheta}_k) \geq D(\boldsymbol{y}_k^{obs}, \boldsymbol{\vartheta}_k)|\boldsymbol{y}_k^{obs}\Big), \tag{6.4.2}$$

teremos dois conjuntos de valores simulados e calculamos a proporção de vezes que as medidas calculadas usando-se os dois conjuntos replicados é maior do que quando calculada com um dos conjuntos replicados mais o observado. Onde a probabilidade é tomada sob a posteriori conjunta de $(\boldsymbol{y}_k^{rep}, \boldsymbol{\vartheta}_k)$. Na prática, a distribuição preditiva é calculada mediante dados simulados. Se tivermos uma amostra válida de tamanho T de $\boldsymbol{\vartheta}_k$, nós calculamos $\boldsymbol{y}_k^{rep}$ da distribuição preditiva para cada $\boldsymbol{\vartheta}_k$ simulado. Vamos ter, então, uma amostra de tamanho T da distribuição a posteriori conjunta de $(\boldsymbol{y}_k^{rep}, \boldsymbol{\vartheta}_k|\boldsymbol{y}_k^{obs})$. A estimativa do valor-p é dada pela proporção de valores nas T simulações tais que, $D\Big((\boldsymbol{y}_k^{rep})^{(t)}, \boldsymbol{\vartheta}_k^{(t)}\Big) \geq D\Big((\boldsymbol{y}_k^{obs})^{(t)}, \boldsymbol{\vartheta}_k^{(t)}\Big)$, $t = 1, 2, \ldots, T$.

A medida de discrepância usada, para um ítem j apresentado ao grupo k é definida como

$$D_p(\boldsymbol{y}_{ik}) = \sum_{j \in \Omega_k} \frac{(n_{jk} - E(n_{jk}))^2}{E(n_{jk})}, \tag{6.4.3}$$

em que, n_{jk} é o número de indivíduos que responderam corretamente ao item j no grupo k. A esperança e variância de n_{jk} são estimados da seguinte forma: para cada um dos parâmetros simulados (traços latentes e parâmetros dos itens) um novo conjunto de respostas é também simulado e, a partir destes, são calculados os valores de n_{jk}. Tomamos a média sob todos n_{jk}^t para toda simulação $t = 1, \ldots, T$. Grupo k, em que L é o escore máximo observado.

A Figura 6.8 mostra os valores-p bayesianos para cada item para o modelo ajustado, isto é, para o modelo 2-MP3NAC. Antes de interpretar os resultados da Figura 6.8, é importante ter em conta que segundo Sahu (2002), um bom ajuste corresponde a valores-p bayesianos entre 0,1 e 0,9. Sobre esta óptica, vemos que, de modo geral os ítens estão devidamente ajustados com exceção dos itens 10, 11, 29, 30, 44 e 45. Para os outros itens este possível problema de ajuste pode-se dever a uma má formulação do item ou da função de resposta, entre outras possibilidades.

Na Figura 6.9 é apresentada a distribuição dos escores preditos e observados para cada grupo. Podemos ver, que para quase todos os grupos os escores preditos estão bem próximos dos observados encontrontrando-se, na maior parte dos casos, dentro da região de credibilidade. É interessante ver como os grupos com maiores problemas no ajuste são o 1 e 6 (o qual precisamente é o grupo referencia). No entanto o baixo resultado do ajuste para estes dois grupos pode ser devido ao fato ao baixo numero de indivíduos neles. Em particular, o grupo 6 só aporta o 1.6% ao tamanho da amostra.

Outro gráfico de avaliação do ajuste do ítem foi o proposto por Sinharay (2006). O autor propõe um gráfico dos escores observados e preditos para cada ítem versus a correspondente proporção de respostas corretas. Os intervalos azuis representam os percentiles 5 e 95 da distribuição preditiva a posteriori, enquanto a linha vermelha representa o valor observado. Aqui apresentamos o gráfico para 6 itens, a saber: Item (1), Item (2), Item (3), Item (4), Item (29), Item (30). Esta eleição foi feita em base aos resultados do valor-p bayesiano que identificava um possível problema de ajuste para os itens (29) e (39). Os resultados se podem encontrar na Figura 6.10. Nesta figura podemos

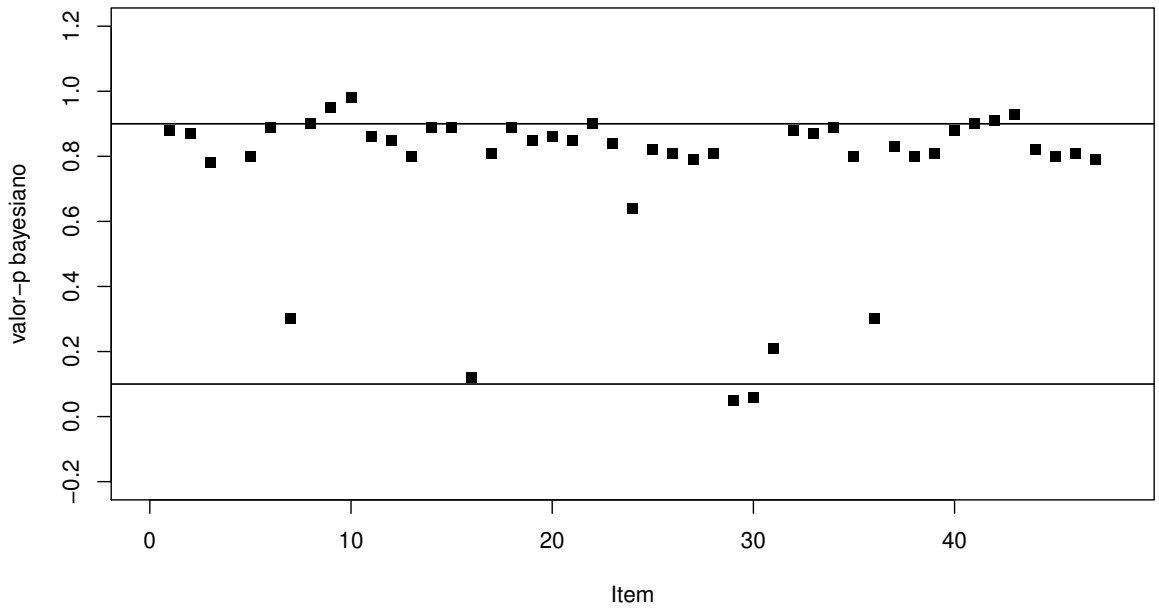


Figura 6.8: Valores-p bayesianos por ítem.

ver como os escores dos itens parecem ser bem preditos, ainda para eles, ainda que segundo o valor-p bayesiano detetou que algum problema de ajuste pareça ter ocorrido problemas de ajuste.

Depois de avaliar a capacidade preditiva a posteriori do modelo, podemos concluir que o modelo ajustado fornece um bom ajuste dos dados. Procedemos agora a apresentar os resultados da estimação para os parâmetros dos items e dos indivíduos.

Na Figura 6.11 apresentamos as estimativas dos parâmetros populacionais. Nesta podemos ver como o pressuposto de assimetria para os grupos é confirmado. Em particular para o grupo 5 ("Medicina"). Podemos ver também como as médias populacionais para todos os grupos resultaram bastante próximas, fato que concorda com as estatísticas descritivas da Tabela 6.1. Novamente o grupo que apresenta maior destaque é o 5 ("Medicina"), já que as estimativas a posteriori para o vetor de medias populacionais associado a ele é maior comparado com os outros grupos.

Nas Figuras 6.12 a 6.15 apresentamos os resultados da estimação para os parâmetros $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ e $\boldsymbol{c}$. Para os parâmetros associados à discriminação multidimensional podemos ver como resultaram ser todos positivos o qual garante que indivíduos com vetores de traços latentes altos vão ter maior probabilidade de responder corretamente os itens. Este fato concorda com o objetivo da prova. Em relação com os parâmetros de acerto casual, vemos como as estimativas para todos os 47 itens estão ao redor do $0,20$, o qual era de esperar, já que para este tipo de provas, de quatro opções sendo três erradas e uma correta, a probabilidade a priori de "chutar" a resposta é de 0.25. É interessante ver como para os itens 40, 41, 42, 44, 45 ,46, e 47 (lembrando que o item 43 foi desconsiderado das análises), as estimativas do parâmetro β são maiores com respeito aos outros

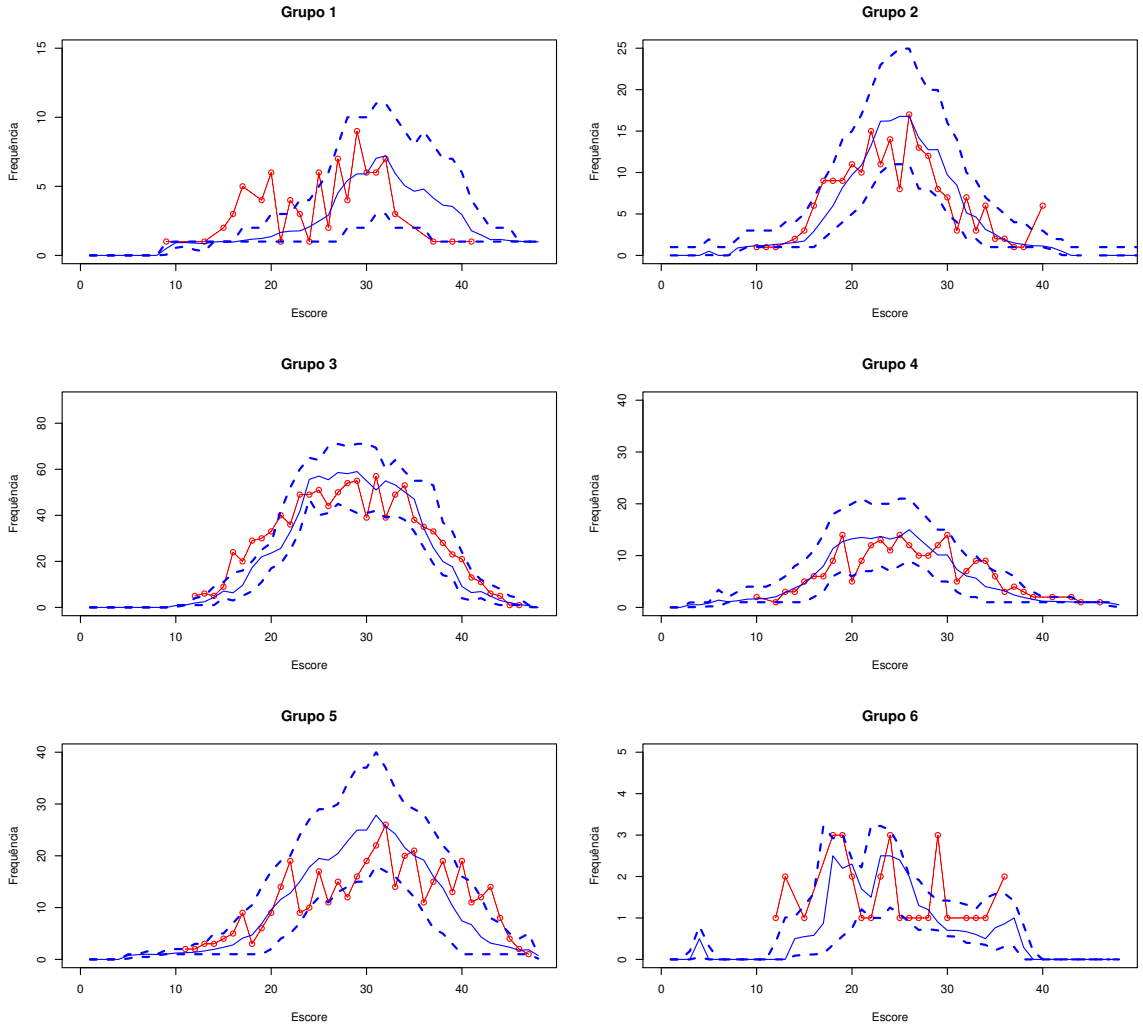
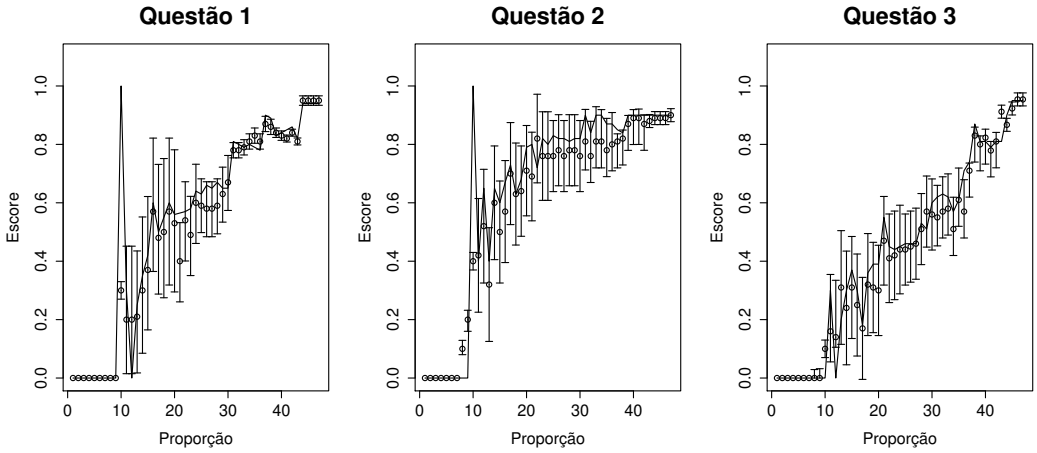
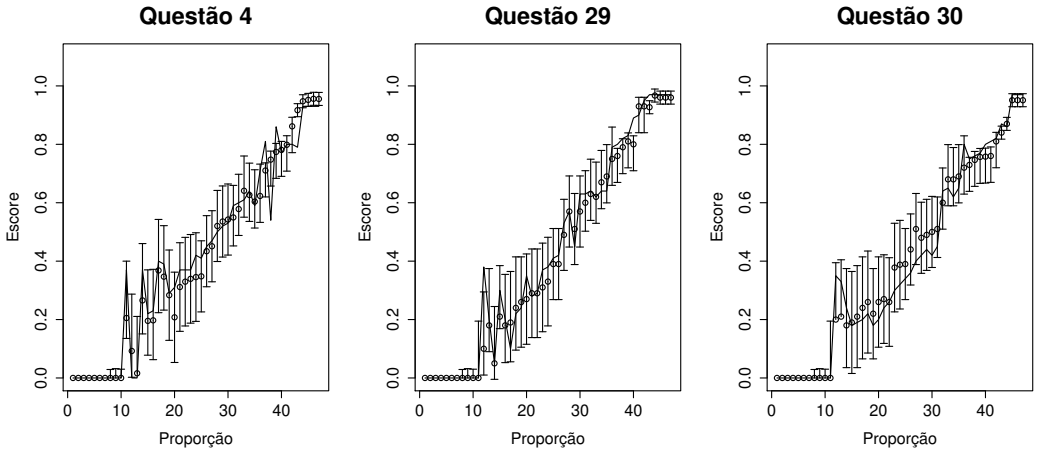


Figura 6.9: Distribuição dos escores. Legenda: linha cheia dados preditos. ∘ dados observados. linha tracejada intervalos de credibilidade de predição

itens. Dada a relação do parâmetro β com a dificuldade do item, isto concorda com o observado nas analises descritivas dos escores dos itens da Figura (6.6). Isto pode significar que de fato este grupo de itens teve maior dificuldade ou por serem os últimos estes foram contestados de forma displicente.



(a) Perguntas 1, 2, 3



(b) Perguntas 4, 29, 30

Figura 6.10: Proporções observadas e replicadas de respostas corretas por grupo de escore.

Na Figura 6.16, apresentamos as estimativas para os parâmetros de discriminação multidimensional. Podemos ver como para os primeiros itens, a discriminação multidimensional é menor com respeito á dos últimos itens. Na Figura 6.17, apresentamos as estimativas para os parâmetros de dificuldade multidimensional. Podemos ver como as dificuldades não seguir nenhum padrão em particular.

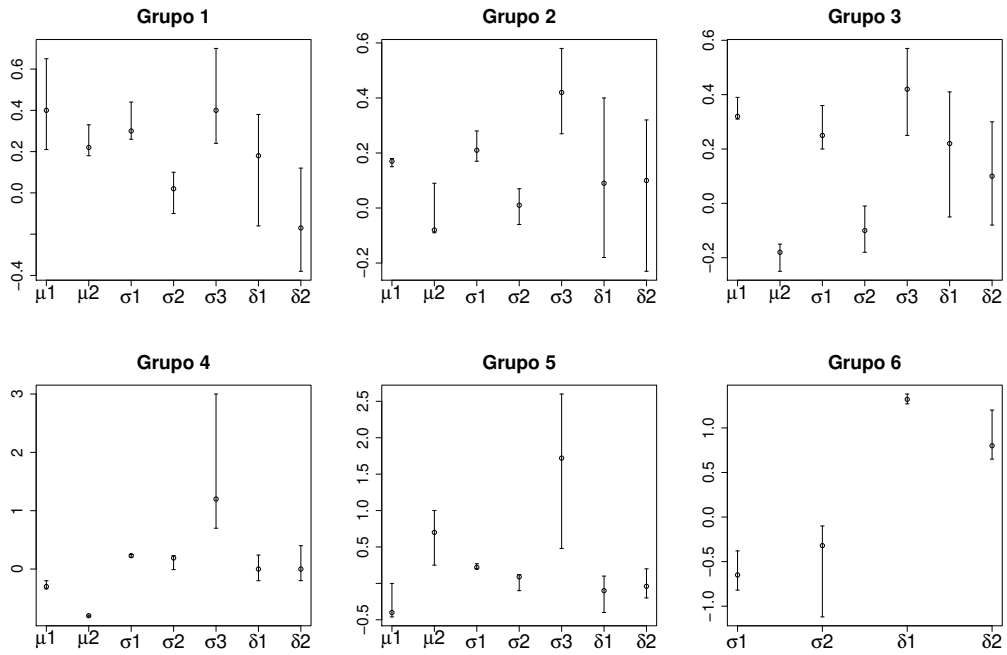


Figura 6.11: Estimativas dos parâmetros populacionais.

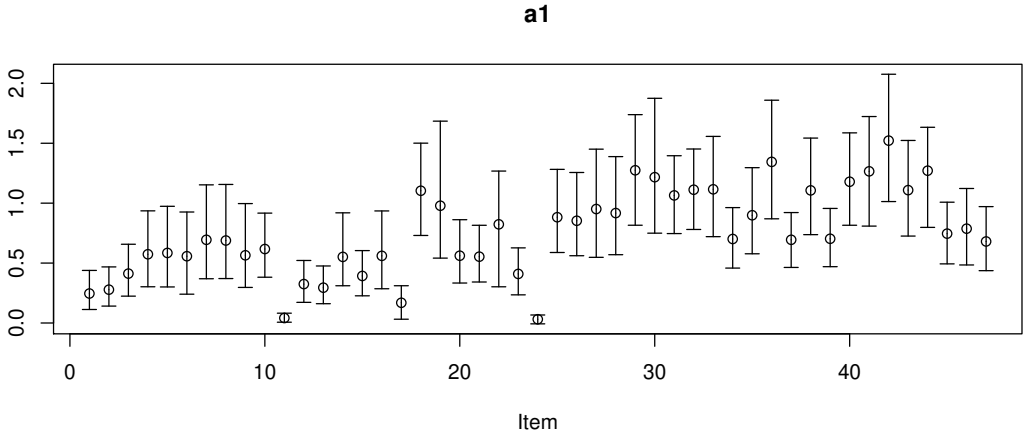


Figura 6.12: Estimativas dos parâmetros α_1.

Nas seguintes figuras apresentamos as distribuições dos traços latentes estimados para cada grupo. Primeiro para o modelo assimétrico de duas dimensões e posteriormente para o modelo simétrico de duas dimensões. Com as suas respectivas curvas teóricas. Podemos ver que o modelo assimétrico caracteriza melhor a distribuição dos traços latentes.

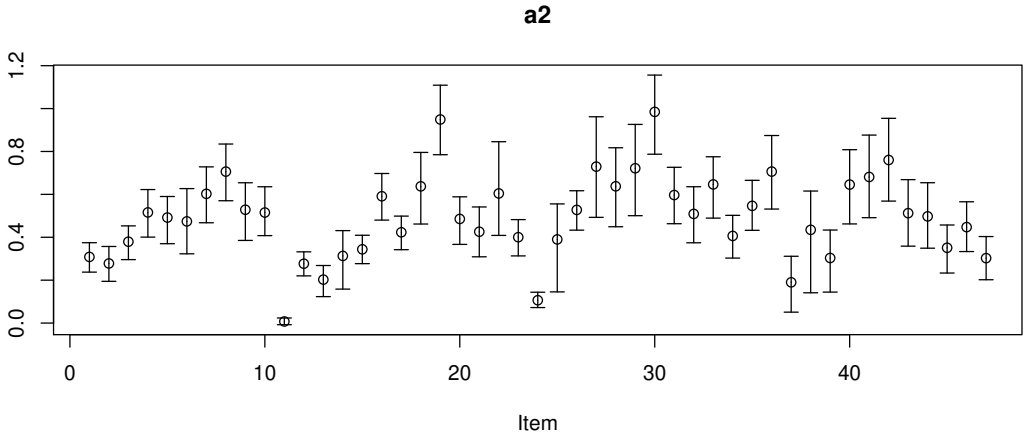


Figura 6.13: Estimativas dos parâmetros α_2.

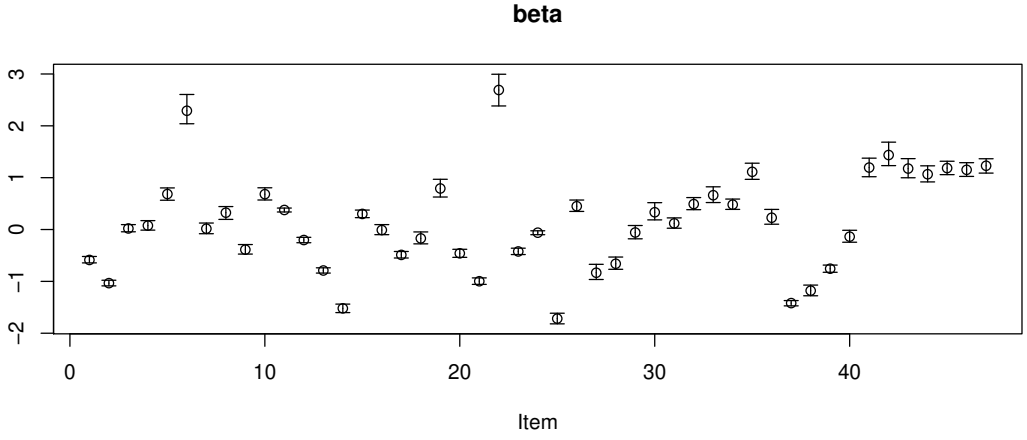


Figura 6.14: Estimativas dos parâmetros β.

Na Tabela 6.5, apresentamos para cada item, a habilidade principal que ele tenta medir, junto com as estimativas dos parâmetros α_1 e α_2. Podemos ver como os valores mais altos de α_1 parecem estar relacionados ás perguntas de "Matemática", "Física" e "Química". Rotacionado os fatores, obtemos a Tabela 6.6, nela podemos obtemos uma interpretação semelhante aquela da Tabela 6.5.

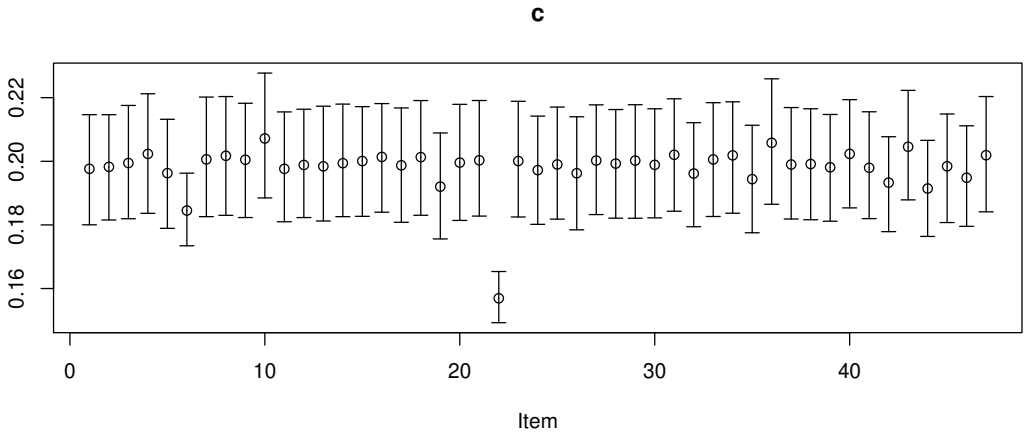


Figura 6.15: Estimativas dos parâmetros c.

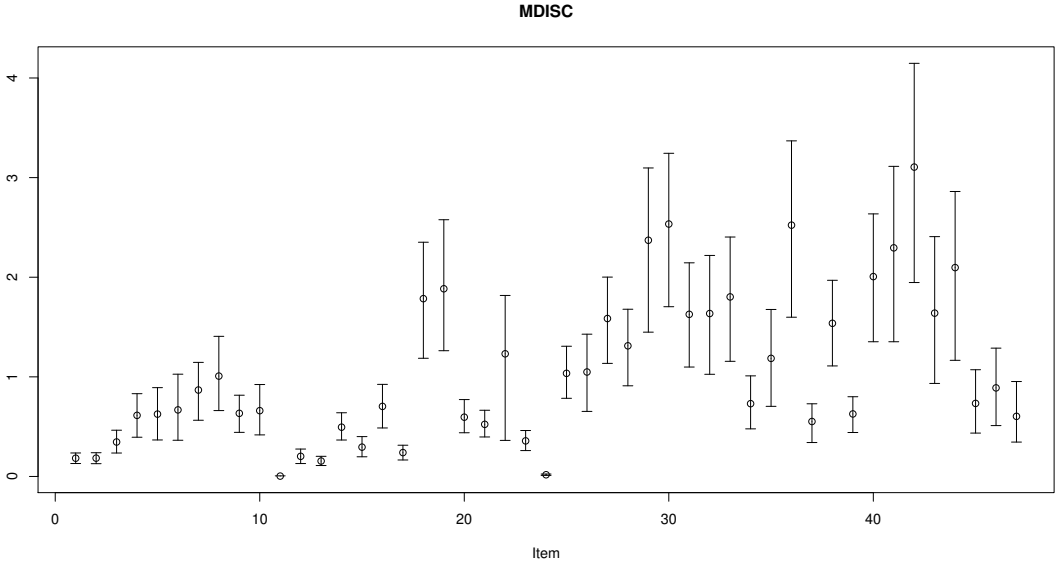


Figura 6.16: Estimativas dos parâmetros de discriminação multidimensional.

6.5 Conclusões e comentários

Na análise dos dados do COMVEST do ano 2013 agrupando os cursos aos quais os participantes se inscreveram como primeira opção e segundo o critério da DAC, o ajuste do modelo de grupos múltiplos assimétrico foi melhor do que o ajuste do modelo de grupos múltiplos usual, também se detetaram duas dimensões principais no teste. O modelo de grupos múltiplos assimétrico multidimensional, detectou assimetria em todos os seis grupos estudados. A metodologia de checagem preditiva a posteriori foi utilizada para verificar o ajuste dos parâmetros dos itens e traços latentes. O valor-p bayesiano mostrou ser uma boa ferramenta de diagnóstico para verificação de ajuste dos

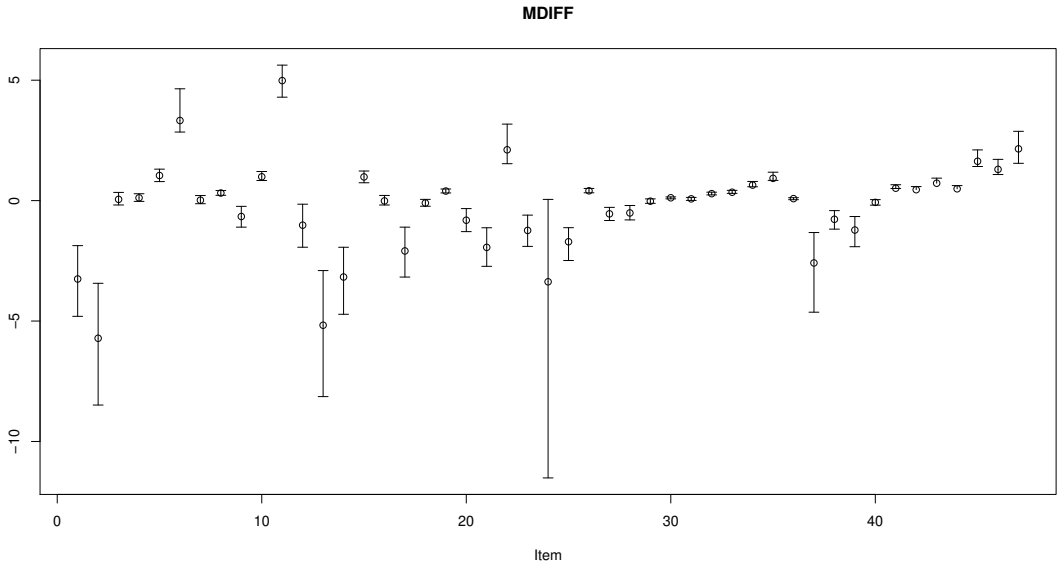


Figura 6.17: Estimativas dos parâmetros de dificuldade multidimensional.

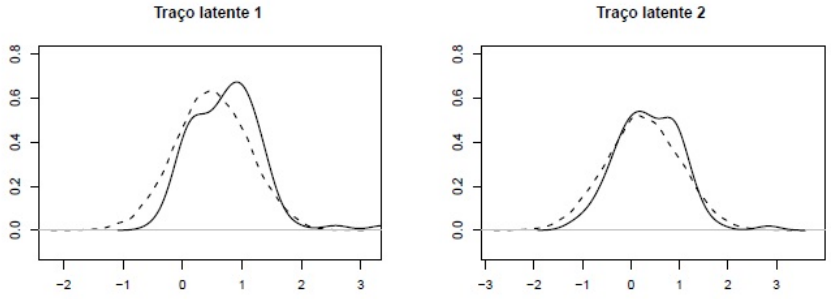


Figura 6.18: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 1-Artes. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

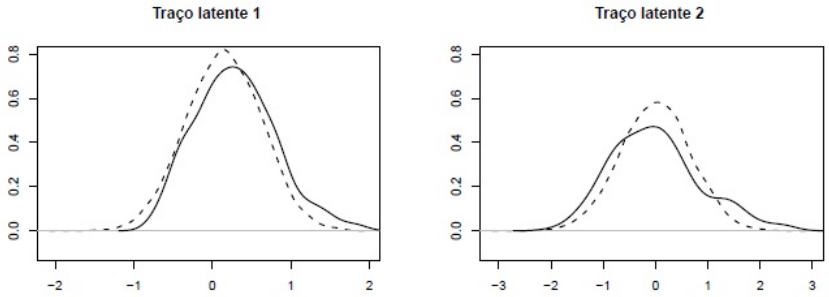


Figura 6.19: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 2-Biológicas e saúde. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

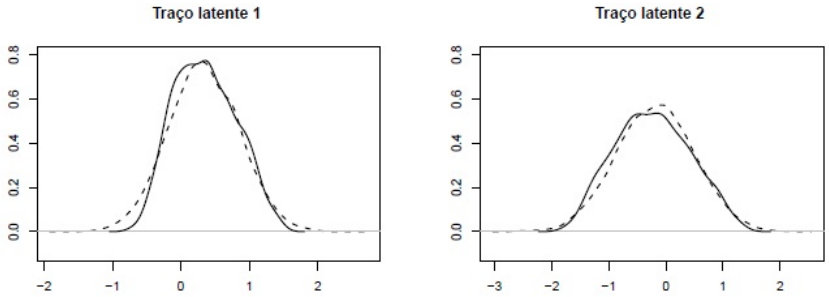


Figura 6.20: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 3-Exatas e da terra. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

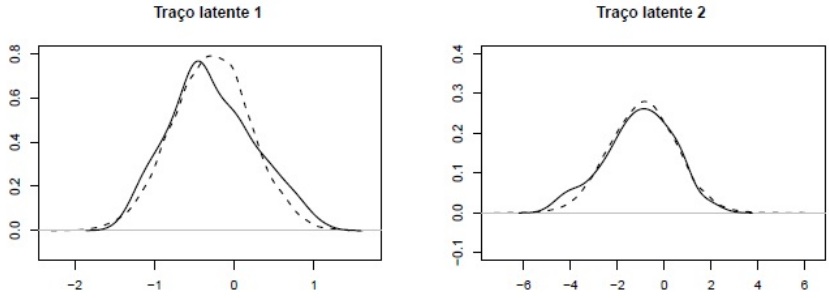


Figura 6.21: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 4-Humanas. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

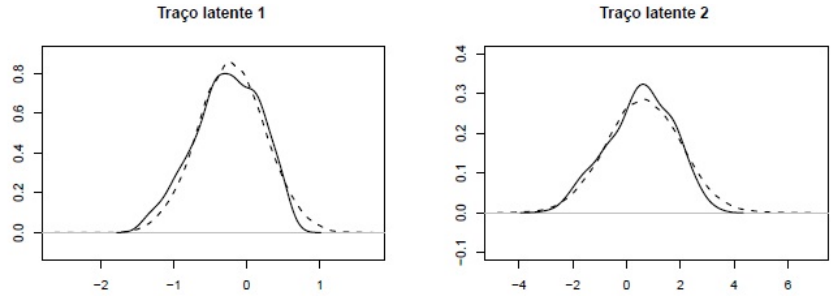


Figura 6.22: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 5-Medicina. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

itens.

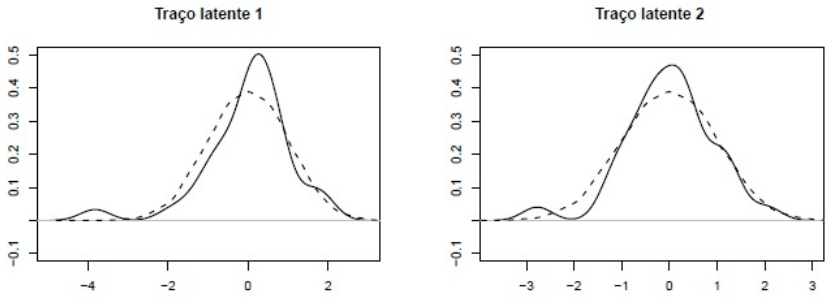


Figura 6.23: Distribuições dos traços latentes com curvas teóricas segundo o modelo assimétrico, Grupo 6-Tecnológicas. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

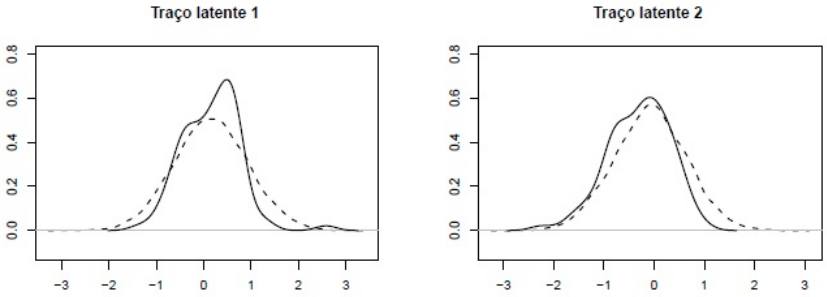


Figura 6.24: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 1-Artes. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

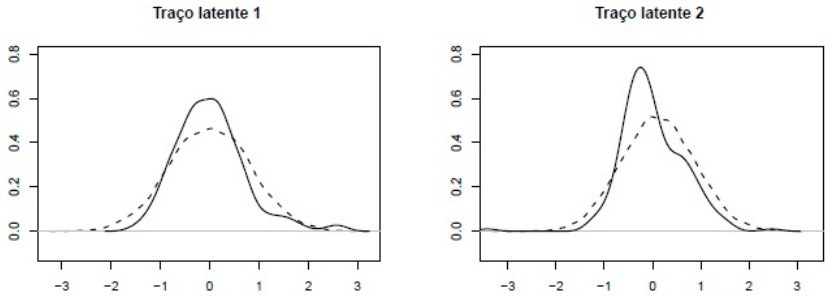


Figura 6.25: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 2-Biológicas e saúde. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

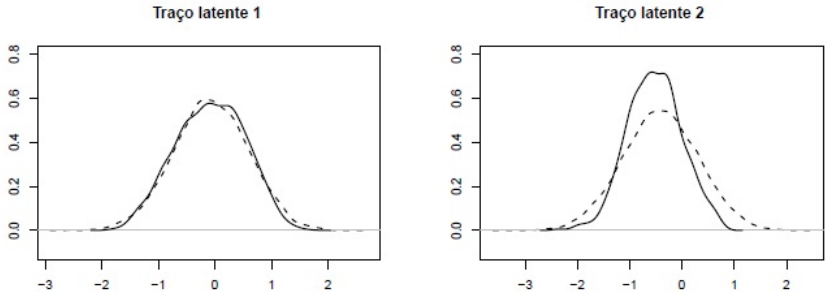


Figura 6.26: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 3-Exatas e da terra. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

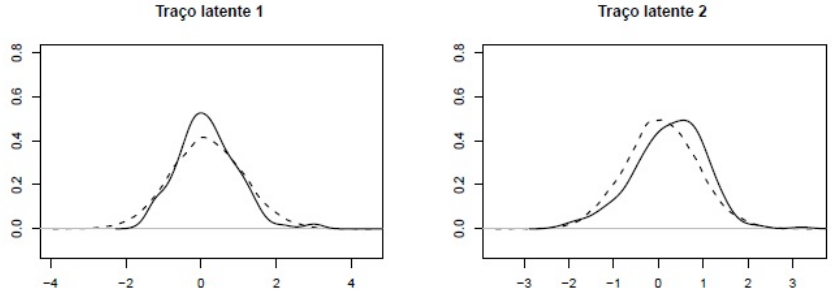


Figura 6.27: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 4-Humanas. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

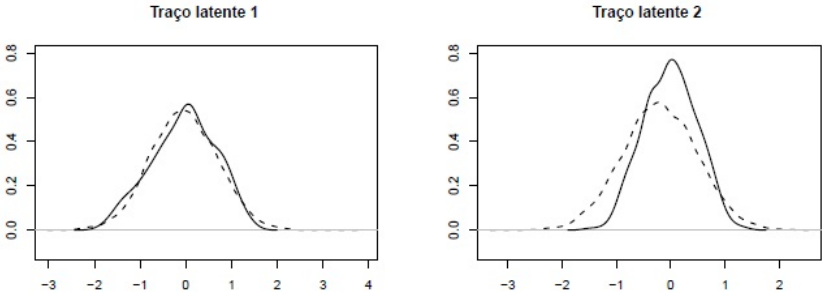


Figura 6.28: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 5-Medicina. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

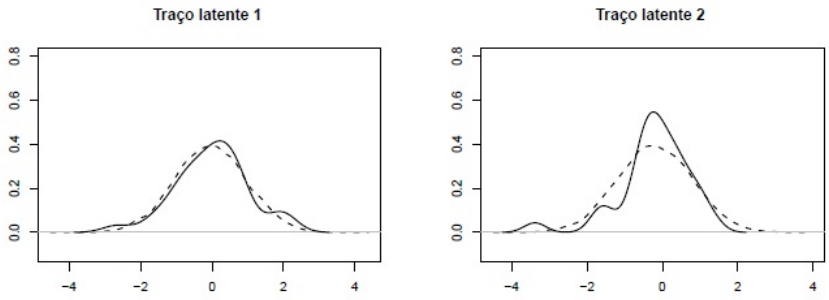


Figura 6.29: Distribuições dos traços latentes com curvas teóricas segundo o modelo simétrico, Grupo 6-Tecnológicas. Legenda: (linha sólida curva estimada, linha tracejada curva teórica).

Tabela 6.5: Habilidades das questões do teste

Item	Habilidade	α_{j1}	α_{j2}	Item	Habilidade	α_{j1}	α_{j2}
Q1	Int. De textos	0,226	0,327	Q25	Química	0,827	0,461
Q2	Int. De textos	0,261	0,294	Q26	Química	0,788	0,550
Q3	Int. De textos	0,380	0,394	Q27	Química	0,880	0,745
Q4	Int. De textos	0,526	0,520	Q28	Química	0,837	0,651
Q5	Historia	0,519	0,518	Q29	Química	1,240	0,722
Q6	Historia	0,626	0,468	Q30	Química	1,112	0,990
Q7	Int. De textos	0,609	0,611	Q31	Física	1,016	0,597
Q8	Int. De textos	0,621	0,722	Q32	Física	1,085	0,515
Q9	Int. De textos	0,489	0,547	Q33	Física	1,078	0,657
Q10	Geografia	0,590	0,496	Q34	Física	0,680	0,396
Q11	Int. De textos	0,054	0,012	Q35	Física	0,872	0,555
Q12	Economia	0,311	0,285	Q36	Física	1,335	0,722
Q13	Geologia	0,277	0,221	Q37	Matemática	0,682	0,231
Q14	Geologia	0,501	0,366	Q38	Matemática	1,045	0,506
Q15	Geologia	0,366	0,349	Q39	Matemática	0,664	0,325
Q16	Demografia	0,506	0,580	Q40	Matemática	1,133	0,658
Q17	Int. De textos	0,195	0,429	Q41	Matemática	1,265	0,680
Q18	Matemática	1,060	0,649	Q42	Matemática	1,519	0,772
Q19	Biologia	0,857	0,936	Q44	Matemática	1,100	0,506
Q20	Biologia	0,518	0,509	Q45	Matemática	1,291	0,505
Q21	Biologia	0,523	0,427	Q46	Matemática	0,727	0,340
Q22	Biologia	0,952	0,607	Q47	Matemática	0,780	0,431
Q23	Biologia	0,388	0,403	Q48	Matemática	0,662	0,298
Q24	Biologia	0,044	0,107				

Tabela 6.6: Fatores rotacionados das questões do teste

Item	Habilidade	α^r_{j1}	α^r_{j2}	Item	Habilidade	α^r_{j1}	α^r_{j2}
Q1	Int. De textos	0,231	0,331	Q25	Química	0,881	0,44
Q2	Int. De textos	0,236	0,311	Q26	Química	0,857	0,539
Q3	Int. De textos	0,382	0,417	Q27	Química	0,886	0,78
Q4	Int. De textos	0,527	0,524	Q28	Química	0,878	0,74
Q5	Historia	0,568	0,506	Q29	Química	0,866	0,789
Q6	Historia	0,509	0,483	Q30	Química	1,177	1,003
Q7	Int. De textos	0,603	0,627	Q31	Física	1,164	0,602
Q8	Int. De textos	0,588	0,727	Q32	Física	0,978	0,524
Q9	Int. De textos	0,567	0,526	Q33	Física	1,022	0,646
Q10	Geografia	0,589	0,497	Q34	Física	0,632	0,428
Q11	Int. De textos	0	0	Q35	Física	0,677	0,531
Q12	Economia	0,435	0,34	Q36	Física	1,201	0,684
Q13	Geologia	0,406	0,318	Q37	Matemática	0,591	0,226
Q14	Geologia	0,524	0,374	Q38	Matemática	0,952	0,423
Q15	Geologia	0,463	0,391	Q39	Matemática	0,585	0,346
Q16	Demografia	0,52	0,584	Q40	Matemática	1,172	0,653
Q17	Int. De textos	0,404	0	Q41	Matemática	1,251	0,676
Q18	Matemática	1,093	0,629	Q42	Matemática	1,438	0,755
Q19	Biologia	0,908	0,971	Q44	Matemática	1,137	0,541
Q20	Biologia	0,52	0,454	Q45	Matemática	1,278	0,517
Q21	Biologia	0,501	0,433	Q46	Matemática	0,589	0,266
Q22	Biologia	0,754	0,604	Q47	Matemática	0,636	0,329
Q23	Biologia	0,442	0,42	Q48	Matemática	0,621	0,228
Q24	Biologia	0,101	0				

Capítulo 7

Comentários e Perspectivas para Novas Pesquisas

7.1 Introdução

Foi muito o que no tempo estabelecido para o desenvolvimento deste projeto se conseguiu concretar, porém longe de considerar-lo como finalizado, os autores querem enfatizar que ainda se tem muitas extensões as quais serão consideradas em posteriores etapas. No entanto, espera-se que a presente dissertação possa ser de ajuda para o leitor interessado no tema. Descrevem-se em seguida alguns comentários do que foi desenvolvido e sugestões para futuras pesquisas relacionadas aos desenvolvimentos aqui descritos.

7.2 Comentários

No presente trabalho, foram estudados e implementados uma classe de modelos multidimensionais da TRI para grupos múltiplos, considerando normais multivariadas assimétricas centradas na modelagem das distribuições dos traços latentes. Como foi possível observar, os modelos aqui desenvolvidos podem ser aplicados em um grande número de situações reais, uma vez que, é comum observar, em estudos com grupos múltiplos, comportamentos de assimetria entre as distribuições dos traços latentes. Embora, a distribuição normal multivariada assimétrica tenha sido amplamente estudada nos últimos anos, seu uso em modelos da TRIM, é muito pouco, em grande parte aos problemas de identificabilidade que algumas parametrizações desta distribuição trazem consigo. Sendo uma das contribuições desta dissertação o desenvolvimento de uma parametrização que sob certas circunstâncias, tem provado garante a identificabilidade dos modelos. O conjunto de dados analisado no Capítulo 6 é um caso onde em principio se poderia considerar como um exemplo de um único grupo, porém análises descritivas identificaram como as distribuições dos traços latentes entre os grupos variam, também conseguiu-se identificar presença de assimetria nos traços latentes de alguns dos grupos. As técnicas de diagnostico permitiram avaliar a qualidade do ajuste do modelo de um modo global.

Alguns aspectos que não foram explorados neste trabalho, bem como possíveis extensões, são

discutidos a seguir.

7.3 Sugestões para futuros trabalhos

Na grande maioria de modelos da TRIM se trabalha com funções de enlace probito ou logito, assim que com relação à função de ligação note-se que, por si só, ela poderia constituir uma extensão, uma vez que, mesmo para modelos mais simples, como unidimensionais de um único grupo, tal função de ligação nunca foi utilizada. O trabalho de Bazan (2005), utiliza a parametrização usual. De acordo com trabalhos como Azevedo, Bolfarine e Andrade, 2012, Santos, Azevedo e Bolfarine (2013) e Azzalini e Capitanio, 1999, particularmente na TRI, a utilização da parametrização centrada é preferível à parametrização usual, por conta da primeira ser melhor em termos de: identificabilidade, estimação e interpretação de parâmetros.

Em relação à distribuição para os traços latentes, à semelhança do que foi realizado nesta dissertação, pode-se definir uma nova versão da distribuição t multivariada assimétrica centrada.

Na parte de estimação e validação de modelos consideraremos o paradigma bayesiano através de algoritmos do tipo MCMC com dados aumentados, utilizada com sucesso tanto nesta dissertação como em outros trabalhos, por exemplo: Albert (1992), Azevedo, Bolfarine e Andrade (2011), Azevedo, Bolfarine e Andrade (2012), Santos, 2012 e Santos, Azevedo e Bolfarine, 2013. Uma extensao pode ser no uso dos resultados de comparação de algoritmos MCMC aqui conduzidos, em especial, no uso do passo de aceleração de convergência.

Apesar dos algoritmos MCMC constituírem uma ferramenta poderosa na utilização da inferência bayesiana e na validação de modelos, seus longos períodos de computação fazem com que métodos de estimação alternativos sejam buscados. Uma alternativa interessante seria o uso do algoritmo EM condicional de dados aumentados proposto por Azevedo e Andrade (2013). Este método é uma variação do algoritmo EM e permite ajustar modelos complexos com razoável facilidade e com baixo custo de processamento computacional. Uma outra possibilidade seria a estimação por máxima verossimilhança marginal. Como já foi mencionado, um dos grandes inconvenientes no desenvolvimento desta dissertação, foi o alto tempo demandado pelos algoritmos MCMC, pelo qual encorajamos a integração entre R e C++ através do pacote Rcpp, veja R Core Team, 2012b.

Para finalizar ressaltamos vários aspectos da importância do realizado neste projeto:

- Utilização dos modelos multivariados, haja vista que os mesmos se constituem em um caso particular dos modelos multidimensionais.

- Para os testes unidimensionais na presença de grupos múltiplos, uma vez que os modelos unidimensionais de grupos múltiplos são casos particulares daquels estudados no presente trabalho.

- Deixamos uma base teórica sólida, para a utilização do algoritmo CADEM Azevedo e Andrade (2013), uma vez que o mesmo utiliza a estrutura das distribuições condicionais completas consideradas no algoritmo MCMC proposto.

- Desenvolvemos modelos, que podem chegar ser implementados em futuros estudos envolvendo MRI que apresentem alguma estrutura de dependência entre os traços latentes, como os modelos longitudinais, por exemplo.

Apêndice A

Detalhes dos algoritmos MCMC

A seguir é descrito o desenvolvimento do algoritmo MCMC para o modelo proposto no presente trabalho; usando a distribuição normal simétrica multivariada e a distribuição normal assimétrica multivariada centrada proposta no Capítulo 2 para a modelagem dos traços latentes.

A.1 Esquema de dados aumentados de Sahu (2002) com a proposta de aceleração de Gonzalez (2004): distribuição normal multivariada simétrica

- **Passo 1:** Simular $\boldsymbol{\Sigma}_\theta^{(t)}$ de $\boldsymbol{\Sigma}_\theta|\boldsymbol{\theta}$ considerando

$$\boldsymbol{\Sigma}_\theta|\boldsymbol{\theta} \sim Wishart - Inversa(\boldsymbol{S}^{-1}, N-1),$$

  onde $\boldsymbol{S}$ é uma matriz com entradas $S_{pq} = \sum_{i=1}^{N}(\theta_{ip} - \overline{\theta}_p)(\theta_{iq} - \overline{\theta}_q)$ e $\overline{\theta}_p$ e $\overline{\theta}_p$ são as medias dos traços latentes nas dimensões p e q respectivamente.

- **Passo 2:** Simular $\boldsymbol{U}^{(t)}$ e $\boldsymbol{Z}^{(t)}$ de $\boldsymbol{U}, \boldsymbol{Z}|(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta})$ considerando

  1. Simular $u_{ij}^{(t)}|(y_{ij}, z_{ij}^{(t-1)}, c_j^{(t-1)})$, assim: se $y_{ij} = 0$, fixar $u_{ij}^{(t)} = 0$, se $y_{ij} = 1$ e $z_{ij}^{(t-1)} < 0$ então fixarmos $u_{ij}^{(t)} = 1$, se $y_{ij} = 1$ e $z_{ij}^{(t-1)} >= 0$ então simulamos $u_{ij}^{(t)}$ de uma $Bernoulli(c_j^{(t)})$.
  2. Simular $z_{ij}^{(t)}|(y_{ij}, u_{ij}^{(t)}, c_j^{(t-1)})$, assim: se $u_{ij}^{(t)} = 0$ então $z_{ij}^{(t)} \sim HN(\eta_{ij}, 1)$. Se $u_{ij}^{(t)} = 1$ então $z_{ij}^{(t)} \sim N(\eta_{ij}, 1)$.

- **Passo 3:** Simular $\boldsymbol{\theta}^{(t)}$ de $\boldsymbol{\theta}_\theta|(\boldsymbol{z}^{(t)}, \boldsymbol{\zeta}^{(t-1)}, \boldsymbol{\Sigma}_\theta^{(t)})$ considerando $\boldsymbol{\theta}_i^0 = (\theta_{i1}^0, \ldots, \theta_{iD}^0)^t$, e a ${\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{1/2}$ a decomposição de Cholesky de ${\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{(t)}$. Defina $\boldsymbol{\theta}_i^0 = {\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{-1/2}(\boldsymbol{\theta}_i^{(t)})$, e seja agora

$$\boldsymbol{\eta}_i^{(t)} = \boldsymbol{\alpha}_{..}\boldsymbol{\Sigma}_\theta^{(t)}(\boldsymbol{\theta}_i) - \boldsymbol{\beta},$$

os traços latentes $\boldsymbol{\theta}_i^0$ tem densidade a posteriori dada por

$$p(\boldsymbol{\theta}_i|\boldsymbol{\zeta},\boldsymbol{z},\boldsymbol{y},\boldsymbol{u}) \propto \phi(\boldsymbol{\theta}_i^0;\mathbf{0},\boldsymbol{I})\prod_{j=1}^{J}\phi(z_{ij};\eta_{ij},1).$$

Isto motiva que $\boldsymbol{z}_i^{(t)} + \boldsymbol{\beta} = \boldsymbol{B}\boldsymbol{\theta}_i^0 + \boldsymbol{\xi}_i$, onde $\boldsymbol{B} = \boldsymbol{\alpha}_{\cdot\cdot}\boldsymbol{\Sigma}_\theta^{1/2}$ e os $\boldsymbol{\xi}_i$ são termos de erro independentes, os quais distribuíam-se $N(0,1)$. Segue-se, então que

$$\boldsymbol{\theta}_i^0 \sim N_D((\boldsymbol{I}+\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{\mathbf{0}};(\boldsymbol{I}+\boldsymbol{\Sigma}^{-1})^{-1}),$$

com $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{\mathbf{0}}$ é o estimador de mínimos quadrados ordinários e $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_\theta^{1/2t}\boldsymbol{\alpha}_{\cdot\cdot}^t\boldsymbol{\alpha}_{\cdot\cdot}^t\boldsymbol{\Sigma}_\theta^{1/2})^{-1}$. $\boldsymbol{\theta}_i^{(t)}$ pode ser obtido com a transformação $\boldsymbol{\Sigma}_\theta^{1/2}\boldsymbol{\theta}_i^0 = \boldsymbol{\theta}_i^{(t)}$.

- **Passo 4:** Simular $\boldsymbol{\zeta}^{(t)}$, para isto, combinamos as a priori de $\alpha_{jd} \sim N(\mu_{\alpha d},\sigma_{\alpha d})$ e $\beta_j \sim N(\mu_\beta,\sigma_\beta)$. Seja $\boldsymbol{\mu}_\zeta = (\mu_{\alpha 1},\ldots,\mu_{\alpha D},\mu_\beta)^t$ e $\boldsymbol{\Sigma}_{\boldsymbol{\zeta 0}} = diag(\sigma_{\alpha 1},\ldots,\sigma_{\alpha D},\sigma_\beta)$. Considere agora $\boldsymbol{X}$ a matriz de dimensão $N\times(D+1)$ formada pelas linhas $(\theta_{i1},\ldots,\theta_{id},\ldots,\theta_{iD},-1)$. Finalmente simule

$$\boldsymbol{\zeta}_j^{(t)}|(\boldsymbol{\theta}^{(t)},\boldsymbol{z},\boldsymbol{y}) \sim N_D(\boldsymbol{\mu}_{\zeta_j},(\boldsymbol{\Sigma}_{\zeta 0})^{-1}+\boldsymbol{X}^t\boldsymbol{X})^{-1}),$$

onde $\boldsymbol{\mu}_{\zeta_j} = (\boldsymbol{\Sigma}_{\zeta 0})^{-1}+\boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{\Sigma}_{\zeta 0}^{-1}\boldsymbol{\mu}_{\zeta_0}+\boldsymbol{X}^t\boldsymbol{z}_j)$.

- **Passo 5' (Proposta de aceleração de convergência de Gonzalez (2004)):** Fixar

$$(\overline{\alpha_{j2}}^{(t)},\ldots,\overline{\alpha_{jD}}^{(t)},\overline{\beta_j}^{(t)}) = (\alpha_{j2}^{(t)}/\alpha_{j1}^{(t)},\ldots,\alpha_{jD}^{(t)}/\alpha_{j1}^{(t)},\beta_j^{(t)}/\alpha_{j1}^{(t)}).$$

Simule um valor ν de uma $f(\nu)$ que nós tomamos como uma $U[0,5;1,5]$. Fixe agora

$$\alpha_{j1}^{(t)} = \nu\alpha_{j1}^{(t)},\ \alpha_{j2}^{(t)} = \nu\alpha_{j1}^{(t)}\overline{\alpha}_{j2}^{(t)},\ldots,\alpha_{jD}^{(t)} = \nu\alpha_{j1}^{(t)}\overline{\alpha}_{jD}^{(t)},$$

com probabilidade

$$pr = min\left\{\frac{L(\boldsymbol{y}|\nu\alpha_{j1}^{(t)},\ldots,\nu\beta^{(t)})_j\pi(\nu\alpha_{j1}^{(t)},\ldots,\nu\beta^{(t)})_j f(1/\nu)}{L(\boldsymbol{y}|\alpha_{j1}^{(t)},\ldots,\beta_j^{(t)})\pi(\alpha_{j1}^{(t)},\ldots,\beta_j^{(t)})f(\nu)}|\nu|,1\right\},$$

e fixe

$$\alpha_{j1}^{(t)} = \alpha_{j1}^{(t)},\ \alpha_{j2}^{(t)} = \alpha_{j1}^{(t)}\overline{\alpha}_{j2}^{(t)},\ldots,\alpha_{jD}^{(t)} = \alpha_1^{(t)}\overline{\alpha}_{jD}^{(t)},$$

com probabilidade $(1-pr)$, onde $L(\boldsymbol{y}|\alpha_{j1}^{(t)},\ldots,\beta_j^{(t)})$ é a verossimilhança dos dados.

- **Passo 6:** Simule $c_j^{(t)}$ a partir de uma distribuição

$$Beta(\kappa_1 + \sum_{i=1}^{N} u_{ij}; \kappa_2 + N - \sum_{i=1}^{N} u_{ij}),$$

com $\kappa_1 = 1$ e $\kappa_2 = 3$.

A.2 Algoritmo para o modelo D-MP3AC

A diferença entre o modelo que se apresenta a continuação com o anterior, radica em que agora se supõe que a distribuição a priori dos traços latentes é uma $NAC_D(\mathbf{0}; \boldsymbol{\Sigma}_\theta, \boldsymbol{\lambda})$, para isto,

- **Passo 1:** Simular H_i a partir de

$$HN(\boldsymbol{A\delta}(\boldsymbol{ACC}^t\boldsymbol{A}^t)^{-1}(\boldsymbol{\theta}-\boldsymbol{B})^t[\boldsymbol{A\delta}(\boldsymbol{ACC}^t\boldsymbol{A}^t)^{-1}(\boldsymbol{A\delta})^t+1]^{-1}; [\boldsymbol{A\delta}(\boldsymbol{ACC}^t\boldsymbol{A}^t)^{-1}(\boldsymbol{A\delta})^t+1]^{-1})$$

- **Passo 2:** Simular $\boldsymbol{\Sigma}_\theta^*$ de $\boldsymbol{\Sigma}_\theta|\boldsymbol{\theta}$ considerando

$$\boldsymbol{\Sigma}_\theta|\boldsymbol{\theta} \sim Wishart - Inversa(\boldsymbol{S}^{-1} + \boldsymbol{\Sigma}_\theta^{(t-1)}, N-1),$$

onde $\boldsymbol{S}$ é uma matriz com entradas $S_{pq} = \sum_{i=1}^{N}(\theta_{ip} - \overline{\theta}_p)(\theta_{iq} - \overline{\theta}_q)$ e $\overline{\theta}_p$ e $\overline{\theta}_p$ são as medias dos traços latentes nas dimensões p e q respectivamente.

Aceite $\boldsymbol{\Sigma_\theta}^{(t)} = \boldsymbol{\Sigma_\theta}^*$, com probabilidade

$$\pi = (\boldsymbol{\Sigma_\theta}^{(t)}, \boldsymbol{\Sigma_\theta}^*) = min\{R_{\Sigma_\theta}, 1\} \text{ em que,}$$

$$R_{\Sigma_\theta} = \frac{\prod_{i=1}^{n} L(\boldsymbol{\theta}_i^{(t-1)}|\boldsymbol{\Sigma}_\theta^*, \boldsymbol{\delta}_\theta^{(t-1)})}{\prod_{i=1}^{n} L(\boldsymbol{\theta}_i^{(t-1)}|\boldsymbol{\Sigma}_\theta^{(t-1)}, \boldsymbol{\delta}_\theta^{(t-1)})}$$

Caso contrario faça $\boldsymbol{\Sigma}_\theta^{(t)} = \boldsymbol{\Sigma}_\theta^{(t-1)}$.

- **Passo 3:** Simular $\boldsymbol{\delta}_\theta^*$ de $\boldsymbol{\delta}_\theta|(\boldsymbol{\theta}, \boldsymbol{\delta}^{(t-1)})$ considerando que

$$\delta_{d,\theta}|(\boldsymbol{\theta}, \delta_d^{(t-1)}) \sim U[\delta_{d\theta}^{(-1)} - 0,02; \delta_{d\theta}^{(-1)} + 0,02],\ \forall d = 1, 2, \ldots, D.$$

Aceite $\boldsymbol{\delta}_{d\theta}^{(t)} = \boldsymbol{\delta}_{d\theta}^*$, com probabilidade

$$\pi = (\boldsymbol{\delta_\theta}^{(t)}, \boldsymbol{\delta_\theta}^*) = min\{R_{\delta_\theta}, 1\} \text{ em que,}$$

$$R_{\delta_\theta} = \frac{\prod_{i=1}^{n} L(\boldsymbol{\theta}_i^{(t-1)}|\boldsymbol{\Sigma}_\theta^{(t)}, \boldsymbol{\delta}_\theta^*)}{\prod_{i=1}^{n} L(\boldsymbol{\theta}_i^{(t-1)})\boldsymbol{\Sigma}_\theta^{(t)}, \boldsymbol{\delta}_\theta^{(t-1)}}$$

Caso contrario faça $\boldsymbol{\delta}_\theta^{(t)} = \boldsymbol{\delta}_\theta^{(t-1)}$.

- **Passo 4:** Simular $\boldsymbol{u}^{(t)}$ e $\boldsymbol{z}^{(t)}$ de $\boldsymbol{u}, \boldsymbol{z}|(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{c}, \boldsymbol{\theta})$ considerando

  1. Simular $u_{ij}^{(t)}|(y_{ij}, z_{ij}^{(t-1)}, c_j^{(t-1)})$, assim: se $y_{ij} = 0$, fixar $u_{ij}^{(t)} = 0$, se $y_{ij} = 1$ e $z_{ij}^{(t-1)} < 0$ então fixarmos $u_{ij}^{(t)} = 1$, se $y_{ij} = 1$ e $z_{ij}^{(t-1)} >= 0$ então simulamos $u_{ij}^{(t)}$ de uma $Bernoulli(c_j^{(t)})$.
  2. Simular $z_{ij}^{(t)}|(y_{ij}, u_{ij}^{(t)}, c_j^{(t-1)})$, assim: se $u_{ij}^{(t)} = 0$ então $z_{ij}^{(t)} \sim HN(\eta_{ij}, 1)$. Se $u_{ij}^{(t)} = 1$ então $z_{ij}^{(t)} \sim N(\eta_{ij}, 1)$.

- **Passo 5:** Simular $\boldsymbol{\theta}^{(t)}$ de $\boldsymbol{\theta}_\theta|(\boldsymbol{z}^{(t)}, \boldsymbol{\zeta}^{(t-1)}, \boldsymbol{\Sigma}_\theta^{(t)})$.

  Vamos considerar as seguintes constantes

  $$\boldsymbol{\mu}_{iconst} = \boldsymbol{\Sigma}_\theta^{1/2}\boldsymbol{A}\boldsymbol{\delta}H_i + \boldsymbol{\Sigma}_\theta^{1/2}\boldsymbol{B} \text{ e } \boldsymbol{\Omega}_\theta = \boldsymbol{\Sigma}_\theta^{1/2}\boldsymbol{A}\boldsymbol{C}\boldsymbol{C}^t\boldsymbol{A}^t),$$

  e seja $\boldsymbol{L}$ a decomposição de Cholesky de $\boldsymbol{\Omega}_\theta$. Defina $\boldsymbol{\theta}_i^0 = \boldsymbol{L}^{-1}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_{iconst})$, e seja agora

  $$\boldsymbol{\eta}_i^{(t)} = \boldsymbol{\alpha}_{\cdot\cdot}\boldsymbol{L}\boldsymbol{L}^{-1}(\boldsymbol{\theta}_i) - \boldsymbol{\beta},$$

  os traços latentes $\boldsymbol{\theta}_i^0$ tem densidade a posteriori dada por

  $$p(\boldsymbol{\theta}_i|\boldsymbol{\zeta}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{u}) \propto \phi(\boldsymbol{\theta}_i^0; \boldsymbol{0}, \boldsymbol{I}) \prod_{j=1}^{J} \phi(z_{ij}; \eta_{ij}, 1).$$

  Isto motiva que $\boldsymbol{z}_i^{(t)} + \boldsymbol{\beta} - \boldsymbol{\alpha}_{\cdot\cdot}\boldsymbol{\mu}_{iconst} = \boldsymbol{B}\boldsymbol{\theta}_i^0 + \boldsymbol{\xi}_i$, onde $\boldsymbol{B} = \boldsymbol{\alpha}_{\cdot\cdot}\boldsymbol{\Sigma}_\theta^{1/2}$ e os $\boldsymbol{\xi}_i$ são termos de erro independentes, os quais distribuíam-se $N(0, 1)$. Segue-se, então que

  $$\boldsymbol{\theta}_i^0 \sim N_D((\boldsymbol{I} + \boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{\boldsymbol{0}}; (\boldsymbol{I} + \boldsymbol{\Sigma}^{-1})^{-1}),$$

  com $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{\boldsymbol{0}}$ é o estimador de mínimos quadrados ordinários e $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_\theta^{1/2t}\boldsymbol{\alpha}_{\cdot\cdot}^t\boldsymbol{\alpha}_{\cdot\cdot}^t\boldsymbol{L})^{-1}$. $\boldsymbol{\theta}_i^{(t)}$ pode ser obtido com a transformação $\boldsymbol{L}\boldsymbol{\theta}_i^0 + \boldsymbol{\mu}_{\boldsymbol{iconst}} = \boldsymbol{\theta}_i^{(t)}$.

- **Passo 6:** Simular $\boldsymbol{\zeta}^{(t)}$, para isto, combinamos as a priori de $\alpha_{jd} \sim N(\mu_{\alpha d}, \sigma_{\alpha d})$ e $\beta_j \sim N(\mu_\beta, \sigma_\beta)$. Seja $\boldsymbol{\mu}_\zeta = (\mu_{\alpha 1}, \ldots, \mu_{\alpha D}, \mu_\beta)^t$ e $\boldsymbol{\Sigma}_{\boldsymbol{\zeta 0}} = diag(\sigma_{\alpha 1}, \ldots, \sigma_{\alpha D}, \sigma_\beta)$. Considere agora $\boldsymbol{X}$ a matriz de dimensão $N \times (D+1)$ formada pelas linhas $(\theta_{i1}, \ldots, \theta_{id}, \ldots, \theta_{iD}, -1)$. Finalmente simule

$$\boldsymbol{\zeta}_j^{(t)}|(\boldsymbol{\theta}^{(t),\boldsymbol{z},\boldsymbol{y}}) \sim N_D(\boldsymbol{\mu}_{\zeta_j}, (\boldsymbol{\Sigma}_{\zeta 0})^{-1} + \boldsymbol{X}^t\boldsymbol{X})^{-1}),$$

onde $\boldsymbol{\mu}_{\zeta_j} = (\boldsymbol{\Sigma}_{\zeta 0})^{-1} + \boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{\Sigma}_{\zeta 0}^{-1}\boldsymbol{\mu}_{\zeta_0} + \boldsymbol{X}^t\boldsymbol{z}_j)$.

- **Passo 6' (Proposta de aceleração de convergência de Gonzalez (2004)):** Fixar

$$(\overline{\alpha_{j2}}^{(t)}, \ldots, \overline{\alpha_{jD}}^{(t)}, \overline{\beta_j}^{(t)}) = (\alpha_{j2}^{(t)}/\alpha_{j1}^{(t)}, \ldots, \alpha_{jD}^{(t)}/\alpha_{j1}^{(t)}, \beta_j^{(t)}/\alpha_{j1}^{(t)})$$

. Simule um valor ν de uma $f(\nu)$ que nós tomamos como uma $U[0,5; 1,5]$. Fixe agora

$$\alpha_{j1}^{(t)} = \nu\alpha_{j1}^{(t)},\ \alpha_{j2}^{(t)} = \nu\alpha_{j1}^{(t)}\overline{\alpha}_{j2}^{(t)}, \ldots, \alpha_{jD}^{(t)} = \nu\alpha_{j1}^{(t)}\overline{\alpha}_{jD}^{(t)},$$

com probabilidade

$$pr = min\Big\{\frac{L(\boldsymbol{y}|\nu\alpha_{j1}^{(t)}, \ldots, \nu\beta^{(t)})_j\pi(\nu\alpha_{j1}^{(t)}, \ldots, \nu\beta^{(t)})_j f(1/\nu)}{L(\boldsymbol{y}|\alpha_{j1}^{(t)}, \ldots, \beta_j^{(t)})\pi(\alpha_{j1}^{(t)}, \ldots, \beta_j^{(t)})f(\nu)}|\nu|, 1\Big\},$$

e fixe

$$\alpha_{j1}^{(t)} = \alpha_{j1}^{(t)},\ \alpha_{j2}^{(t)} = \alpha_{j1}^{(t)}\overline{\alpha}_{j2}^{(t)}, \ldots, \alpha_{jD}^{(t)} = \alpha_1^{(t)}\overline{\alpha}_{jD}^{(t)},$$

com probabilidade $(1 - pr)$, onde $L(\boldsymbol{y}|\alpha_{j1}^{(t)}, \ldots, \beta_j^{(t)})$ é a verossimilhança dos dados.

- **Passo 7:** Simule $c_j^{(t)}$ a partir de uma distribuição

$$Beta(\kappa_1 + \sum_{i=1}^{N} u_{ij}; \kappa_2 + N - \sum_{i=1}^{N} u_{ij}),$$

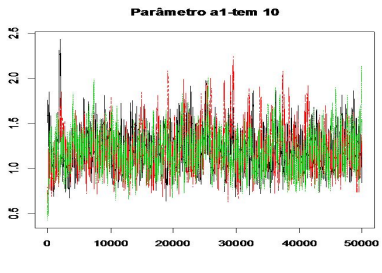
com $\kappa_1 = 1$ e $\kappa_2 = 3$.

Apêndice B

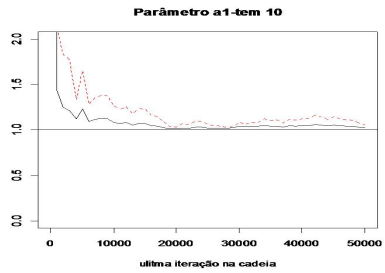
Gráficos

B.1 Estudo de simulação capítulo 4

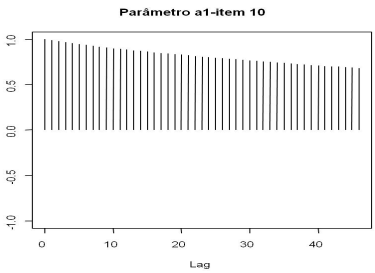
No presente apêndice apresentamos as figuras referentes ao estudo de simulação feito no Capítulo 4. Só lembrando, dito estudo era baseado em uma simulação de um teste de 2000 indivíduos 30 itens, e 2 dimensões. Nós supomos que os traços latentes distribuíam-se segundo uma normal bivariada simétrica. Apresenta-se as figuras, com o objetivo de complementar os resultados do Capítulo 4, especialmente no referente à convergência dos algoritmos.



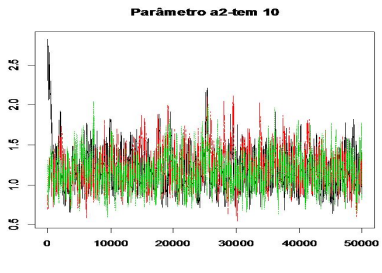
(a) Mistura de cadeias parâmetro a1-item 10



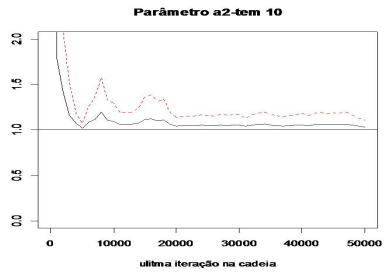
(b) Gelman e Rubin diagnostico parâmetro a1-item 10



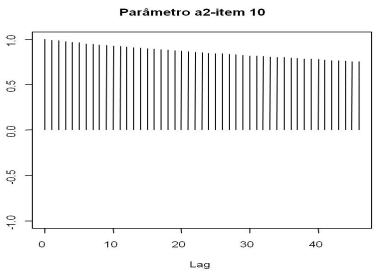
(c) Autocorrelação parâmetro a1-item 10



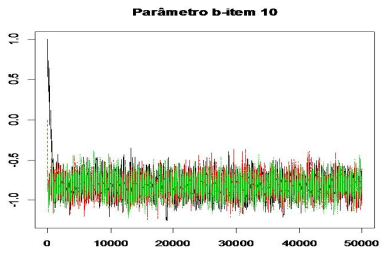
(d) Mistura cadeias parâmetro a2-item 10



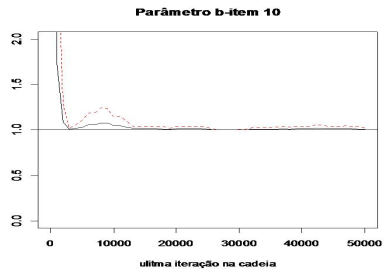
(e) Gelman e Rubin diagnostico parâmetro a2-item 10



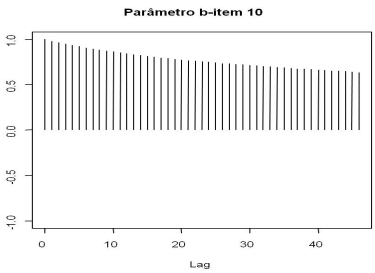
(f) Autocorrelação parâmetro a2-item 10



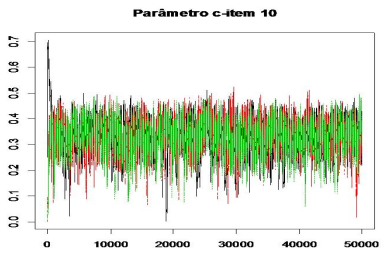
(g) Mistura cadeias parâmetro b-item 10



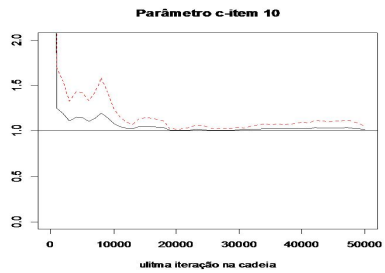
(h) Gelman e Rubin diagnostico parâmetro b-item 10



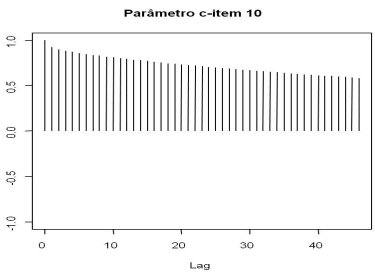
(i) Autocorrelação parâmetro b-item 10



(j) Mistura cadeias parâmetro c-item 10

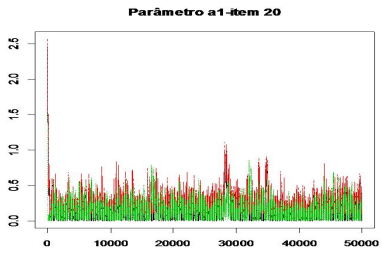


(k) Gelman e Rubin diagnostico parâmetro c-item 10

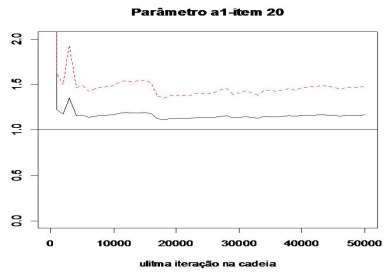


(l) Autocorrelação parâmetro c-item 10

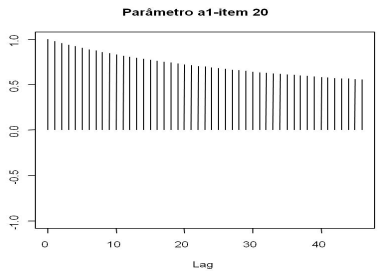
Figura B.1: Beguin-Glass item 10



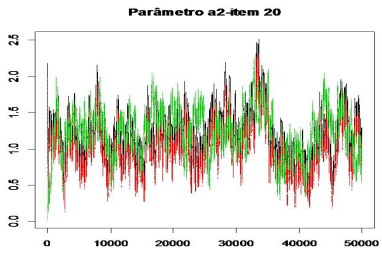
(a) Mistura cadeias parâmetro a1-item 20



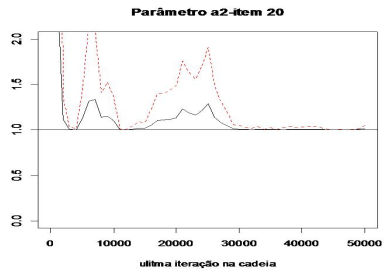
(b) Gelman e Rubin diagnostico parâmetro a1-item 20



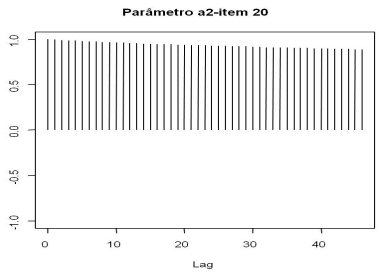
(c) Autocorrelação par. a1-item 20



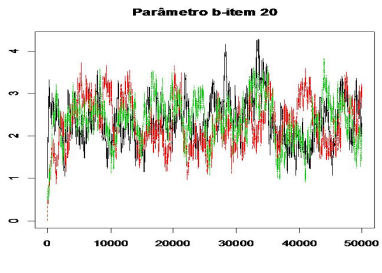
(d) Mistura cadeias parâmetro a2-item 20



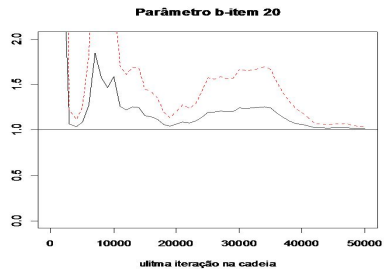
(e) Gelman e Rubin diagnostico parâmetro a2-item 20



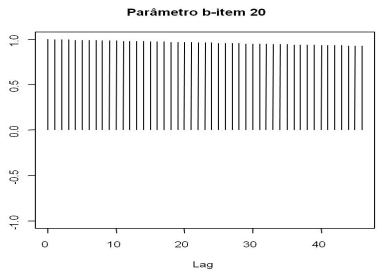
(f) Autocorrelação par. a2-item 20



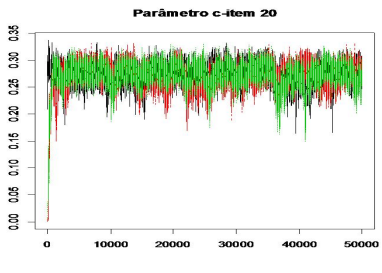
(g) Mistura cadeias parâmetro b-item 20



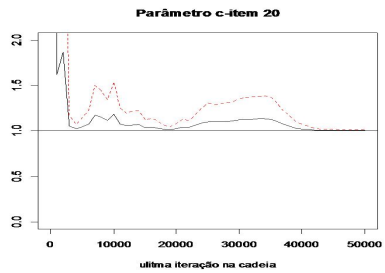
(h) Gelman e Rubin diagnostico parâmetro b-item 20



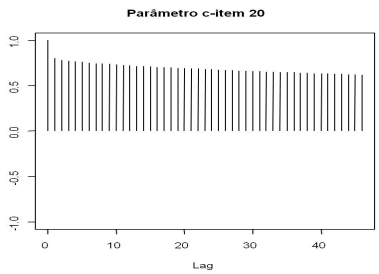
(i) Autocorrelação par. b-item 20



(j) Mistura cadeias parâmetro c-item 20

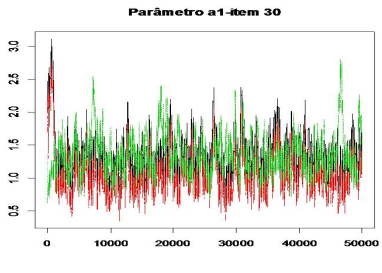


(k) Gelman e Rubin diagnostico parâmetro c-item 20

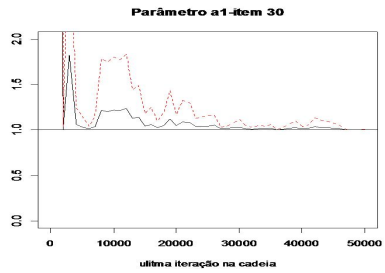


(l) Autocorrelação par. c-item 20

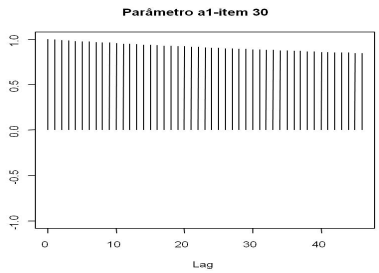
Figura B.2: Beguin-Glass item 20



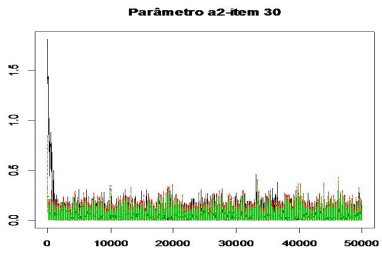
(a) Mistura cadeias parâmetro a1-item 30



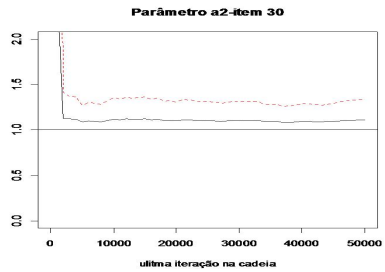
(b) Gelman e Rubin diagnostico parâmetro a1-item 30



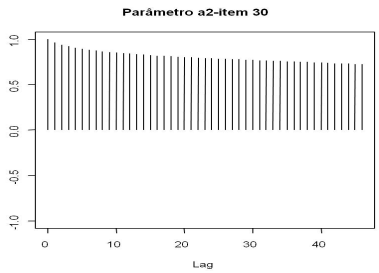
(c) Autocorrelação par. a1-item 30



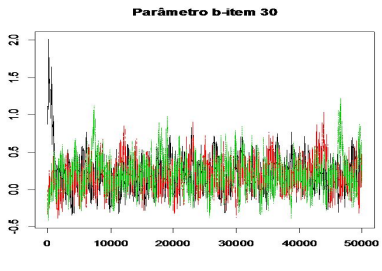
(d) Mistura cadeias parâmetro a2-item 30



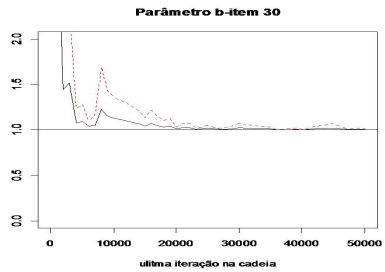
(e) Gelman e Rubin diagnostico parâmetro a2-item 30



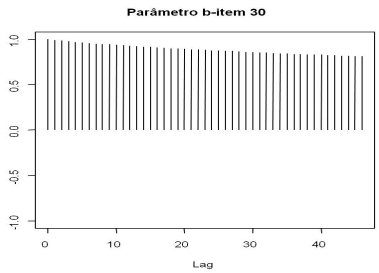
(f) Autocorrelação par. a2-item 30



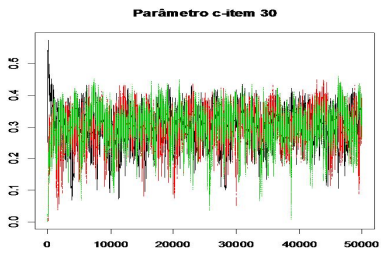
(g) Mistura cadeias parâmetro b-item 30



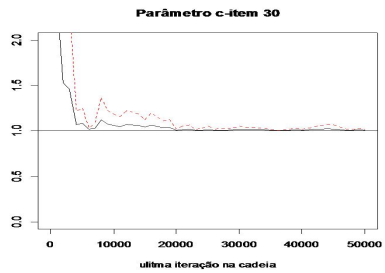
(h) Gelman e Rubin diagnostico parâmetro b-item 30



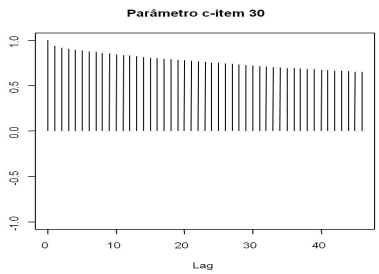
(i) Autocorrelação par. b-item 30



(j) Mistura cadeias parâmetro c-item 30

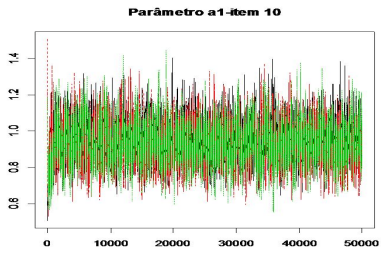


(k) Gelman e Rubin diagnostico parâmetro c-item 30

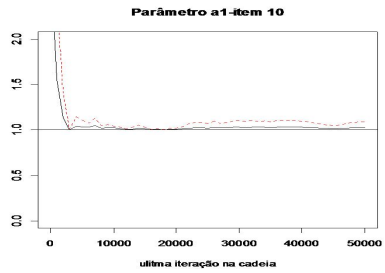


(l) Autocorrelação par. c-item 30

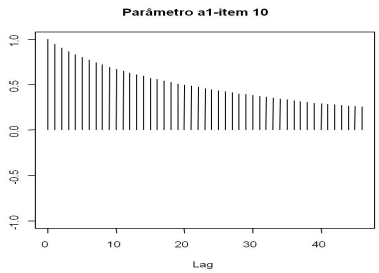
Figura B.3: Beguin-Glass item 30



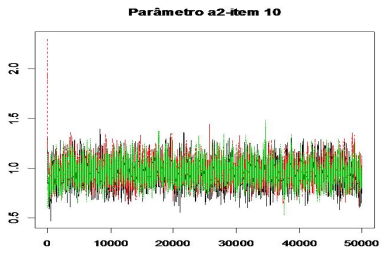
(a) Mistura cadeias parâmetro a1-item 10



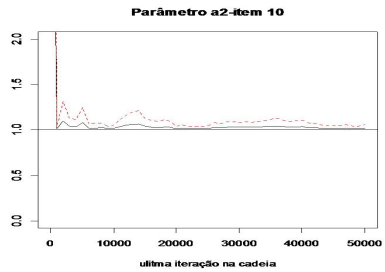
(b) Gelman e Rubin diagnostico parâmetro a1-item 10



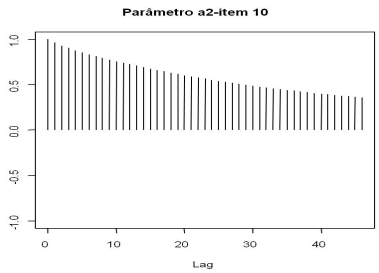
(c) Autocorrelação par. a1-item 10



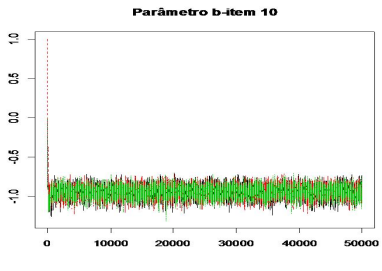
(d) Mistura cadeias parâmetro a2-item 10



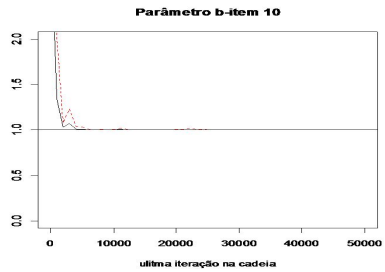
(e) Gelman e Rubin diagnostico parâmetro a2-item 10



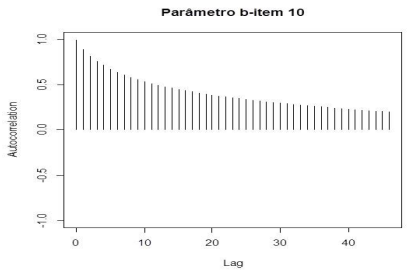
(f) Autocorrelação par. a2-item 10



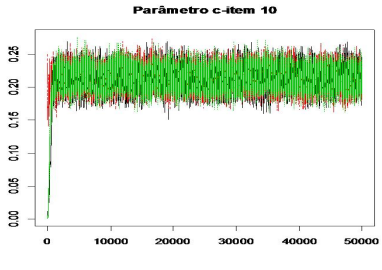
(g) Mistura cadeias parâmetro b-item 10



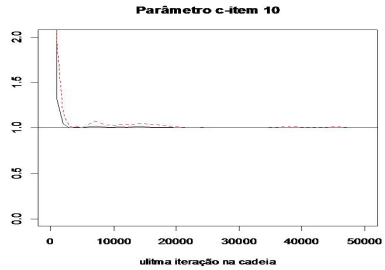
(h) Gelman e Rubin diagnostico parâmetro b-item 10



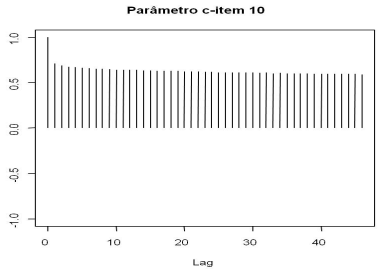
(i) Autocorrelação par. b-item 10



(j) Mistura cadeias parâmetro c-item 10

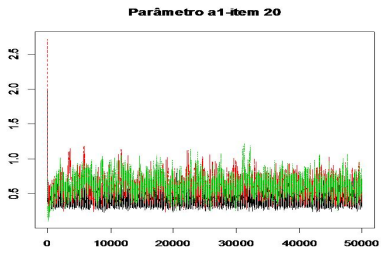


(k) Gelman e Rubin diagnostico parâmetro c-item 10

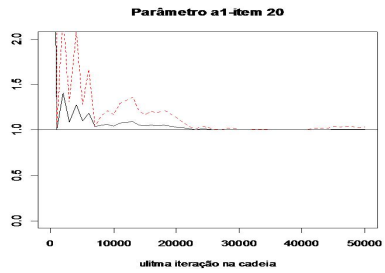


(l) Autocorrelação par. c-item 10

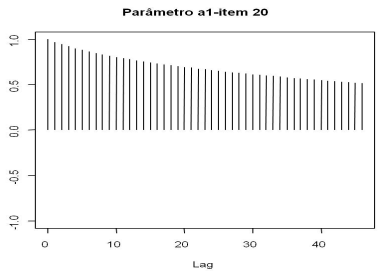
Figura B.4: Sahu item 10



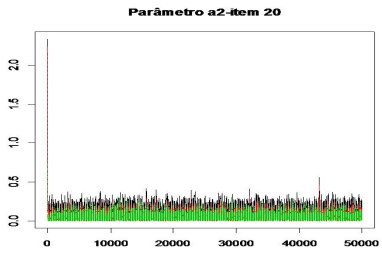
(a) Mistura cadeias parâmetro a1-item 20



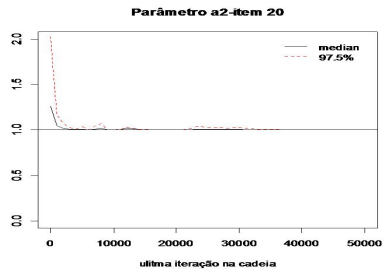
(b) Gelman e Rubin diagnostico parâmetro a1-item 20



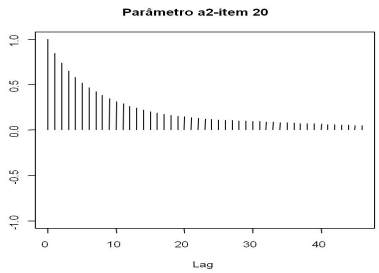
(c) Autocorrelação par. a1-item 20



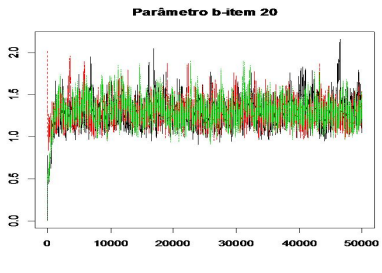
(d) Mistura cadeias parâmetro a2-item 20



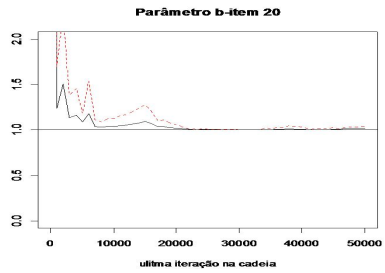
(e) Gelman e Rubin diagnostico parâmetro a2-item 20



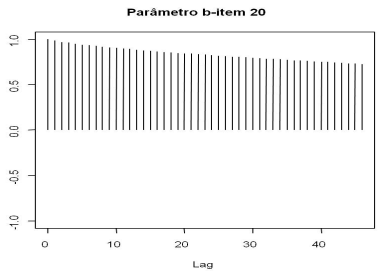
(f) Autocorrelação par. a2-item 20



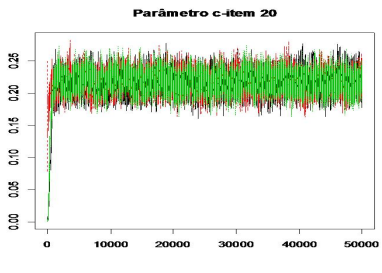
(g) Mistura cadeias parâmetro b-item 20



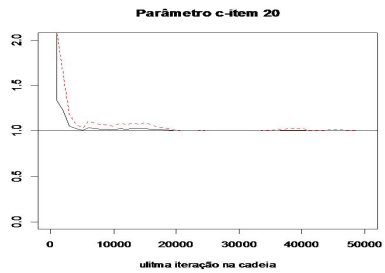
(h) Gelman e Rubin diagnostico parâmetro b-item 20



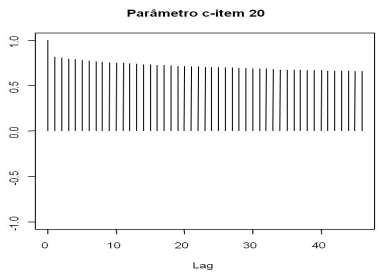
(i) Autocorrelação par. b-item 20



(j) Mistura cadeias parâmetro c-item 20

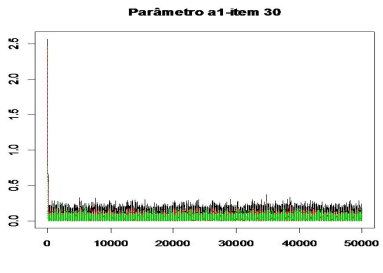


(k) Gelman e Rubin diagnostico parâmetro c-item 20

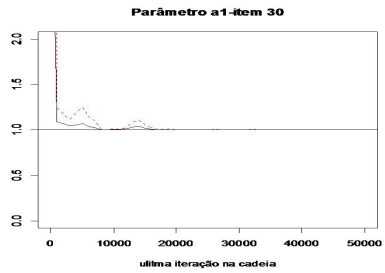


(l) Autocorrelação par. c-item 20

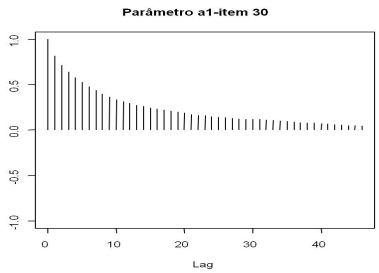
Figura B.5: Sahu item 20



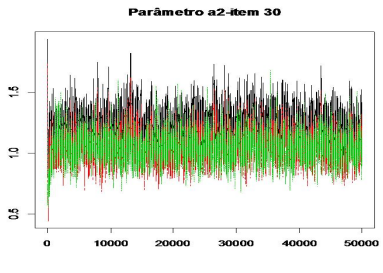
(a) Mistura cadeias parâmetro a1-item 30



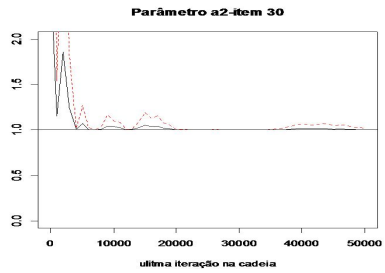
(b) Gelman e Rubin diagnostico parâmetro a1-item 30



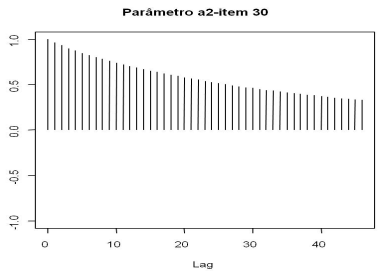
(c) Autocorrelação par. a1-item 30



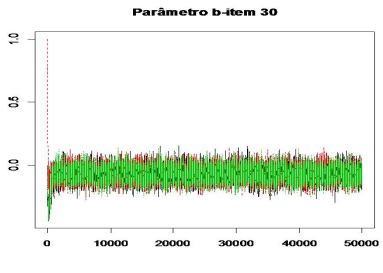
(d) Mistura cadeias parâmetro a2-item 30



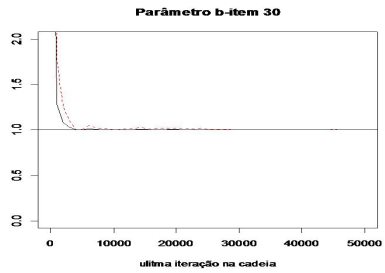
(e) Gelman e Rubin diagnostico parâmetro a2-item 30



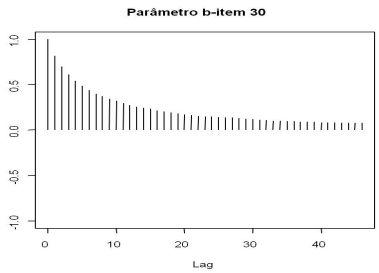
(f) Autocorrelação par. a2-item 30



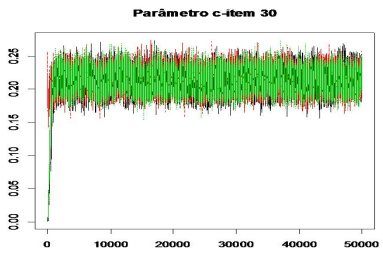
(g) Mistura cadeias parâmetro b-item 30



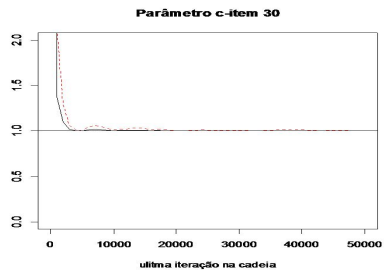
(h) Gelman e Rubin diagnostico parâmetro b-item 30



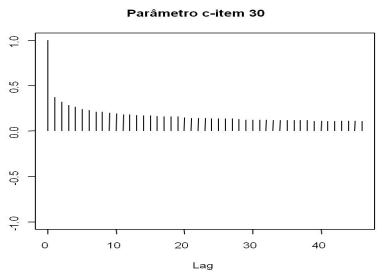
(i) Autocorrelação par. b-item 30



(j) Mistura cadeias parâmetro c-item 30

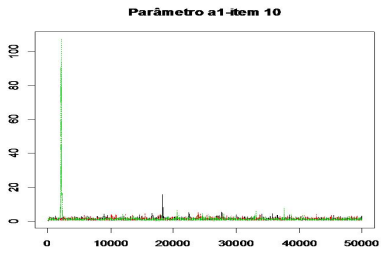


(k) Gelman e Rubin diagnostico parâmetro c-item 30

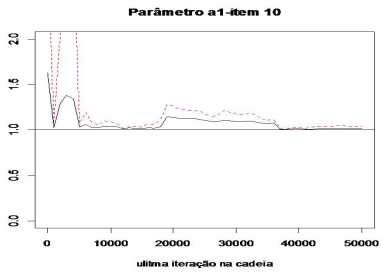


(l) Autocorrelação par. c-item 30

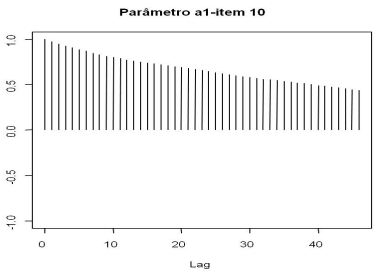
Figura B.6: Sahu item 30



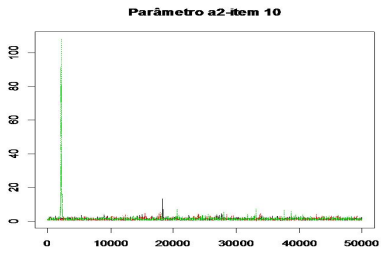
(a) Mistura cadeias parâmetro a1-item 10



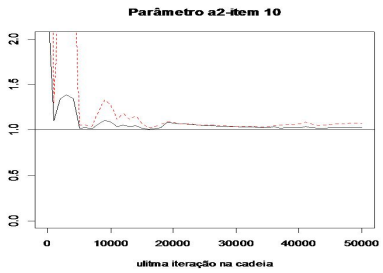
(b) Gelman e Rubin diagnostico parâmetro a1-item 10



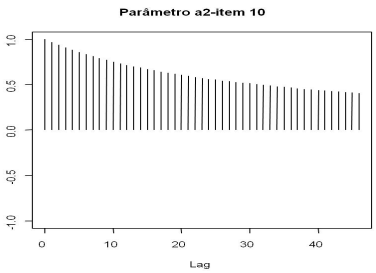
(c) Autocorrelação par. a1-item 10



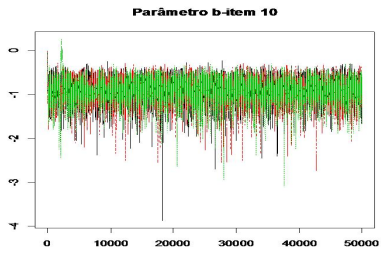
(d) Mistura cadeias parâmetro a2-item 10



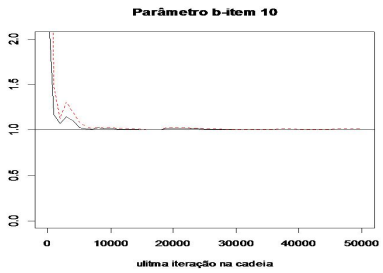
(e) Gelman e Rubin diagnostico parâmetro a2-item 10



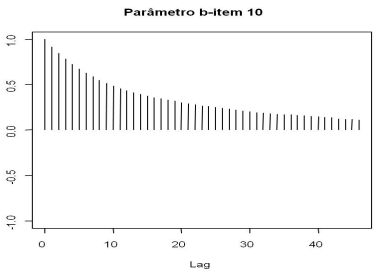
(f) Autocorrelação par. a2-item 10



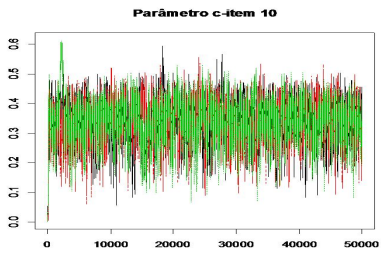
(g) Mistura cadeias parâmetro b-item 10



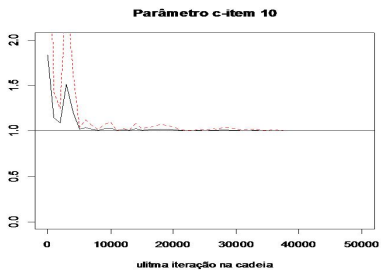
(h) Gelman e Rubin diagnostico parâmetro b-item 10



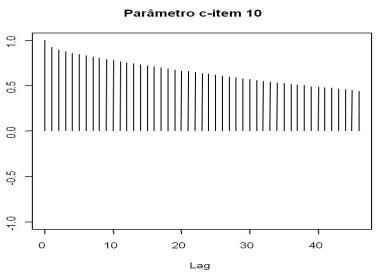
(i) Autocorrelação par. b-item 10



(j) Mistura cadeias parâmetro c-item 10

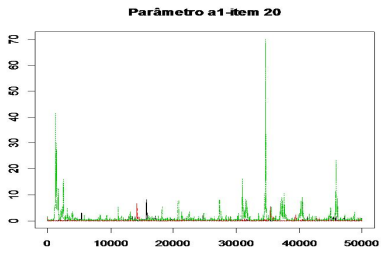


(k) Gelman e Rubin diagnostico parâmetro c-item 10

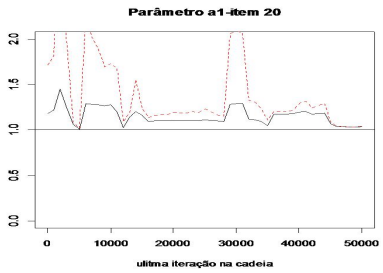


(l) Autocorrelação par. c-item 10

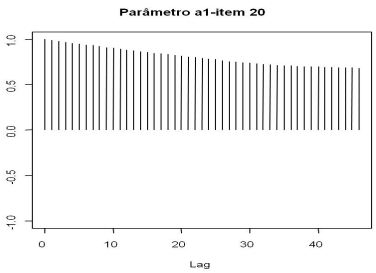
Figura B.7: Beguin-Glas com Gonzalez item 10



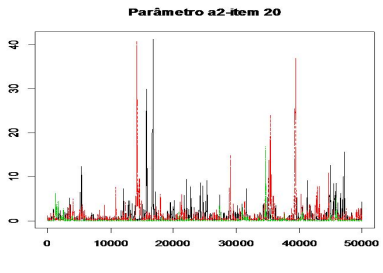
(a) Mistura cadeias parâmetro a1-item 20



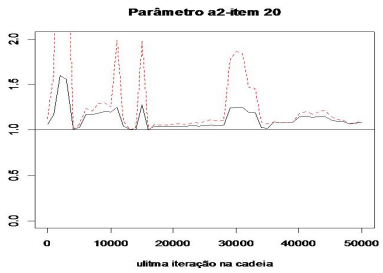
(b) Gelman e Rubin diagnostico parâmetro a1-item 20



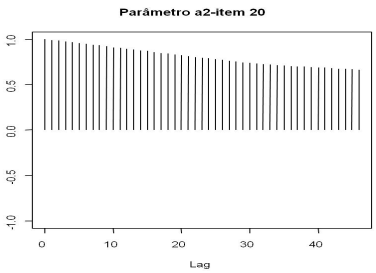
(c) Autocorrelação par. a1-item 20



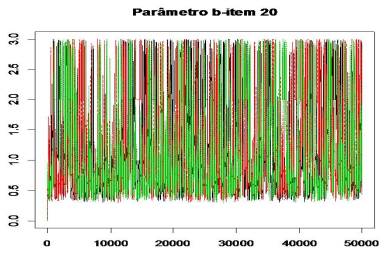
(d) Mistura cadeias parâmetro a2-item 20



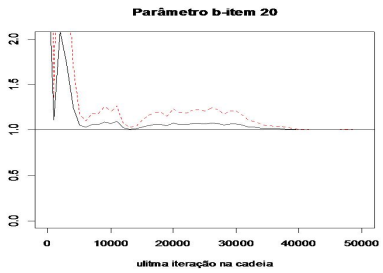
(e) Gelman e Rubin diagnostico parâmetro a2-item 20



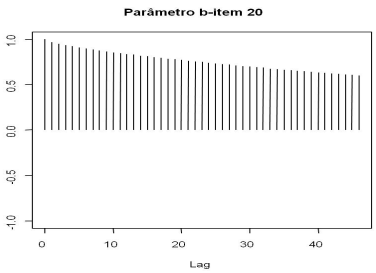
(f) Autocorrelação par. a2-item 20



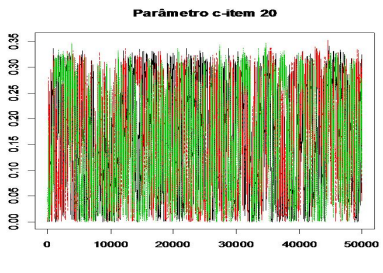
(g) Mistura cadeias parâmetro b-item 20



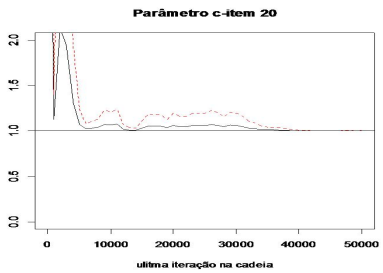
(h) Gelman e Rubin diagnostico parâmetro b-item 20



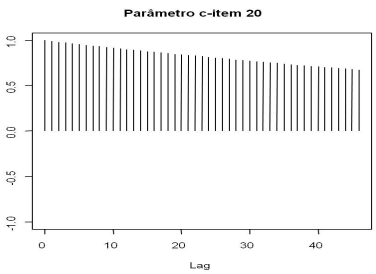
(i) Autocorrelação par. b-item 20



(j) Mistura cadeias parâmetro c-item 20

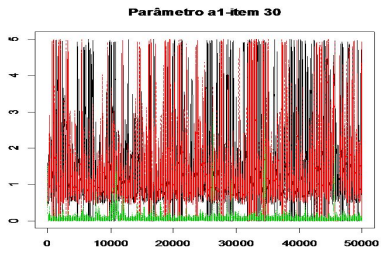


(k) Gelman e Rubin diagnostico parâmetro c-item 20

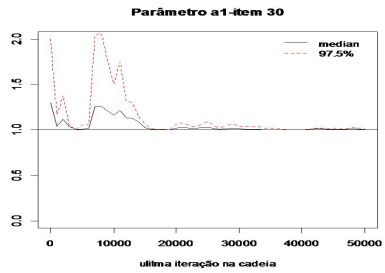


(l) Autocorrelação par. c-item 20

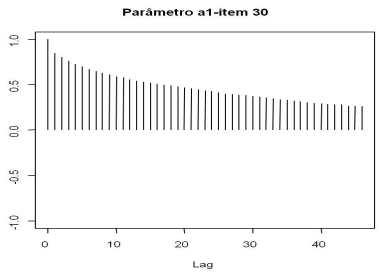
Figura B.8: Beguin-Glas com Gonzalez item 20



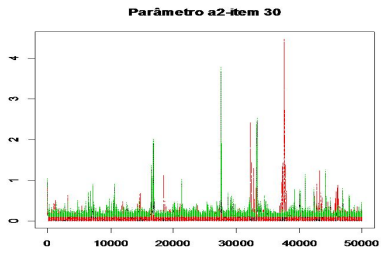
(a) Mistura cadeias parâmetro a1-item 30



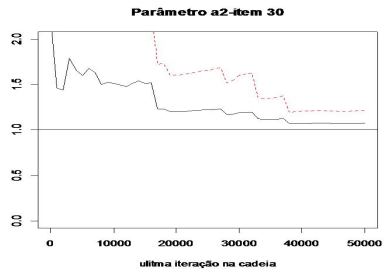
(b) Gelman e Rubin diagnostico parâmetro a1-item 30



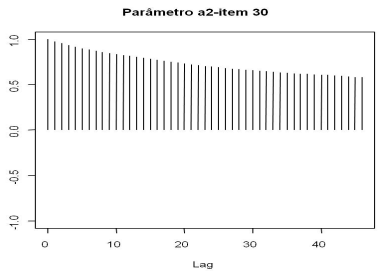
(c) Autocorrelação par. a1-item 30



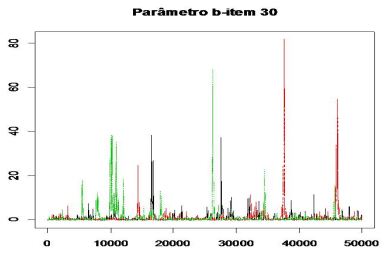
(d) Mistura cadeias parâmetro a2-item 30



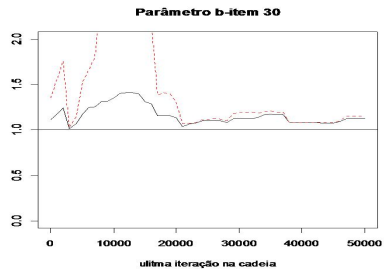
(e) Gelman e Rubin diagnostico parâmetro a2-item 30



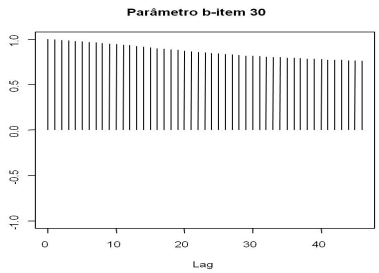
(f) Autocorrelação par. a2-item 30



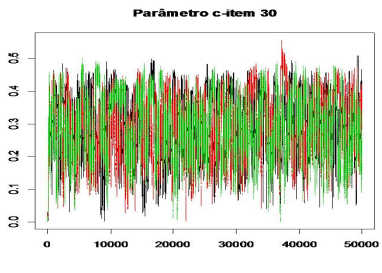
(g) Mistura cadeias parâmetro b-item 30



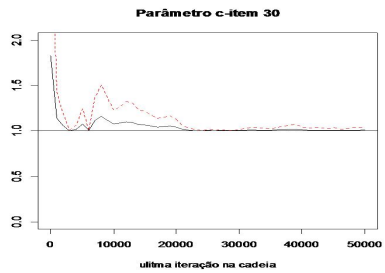
(h) Gelman e Rubin diagnostico parâmetro b-item 30



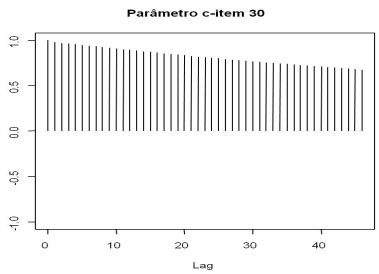
(i) Autocorrelação par. b-item 30



(j) Mistura cadeias parâmetro c-item 30

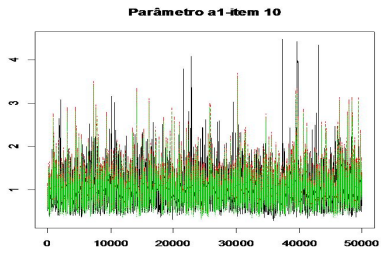


(k) Gelman e Rubin diagnostico parâmetro c-item 30

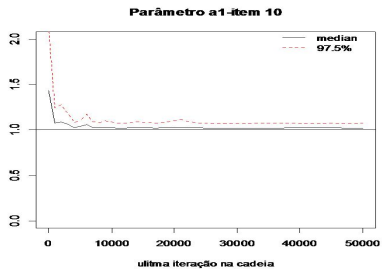


(l) Autocorrelação par. c-item 30

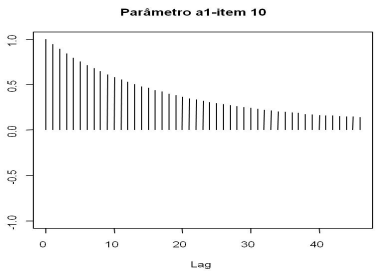
Figura B.9: Beguin-Glas com Gonzalez item 30



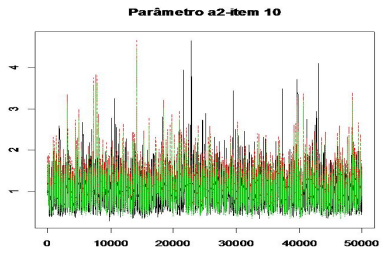
(a) Mistura cadeias parâmetro a1-item 10



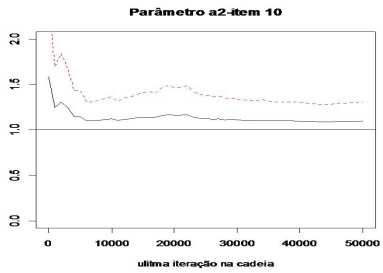
(b) Gelman e Rubin diagnostico parâmetro a1-item 10



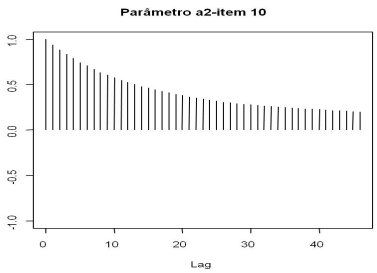
(c) Autocorrelação par. a1-item 10



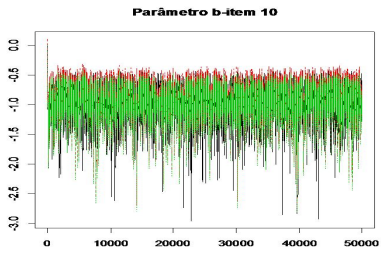
(d) Mistura cadeias parâmetro a2-item 10



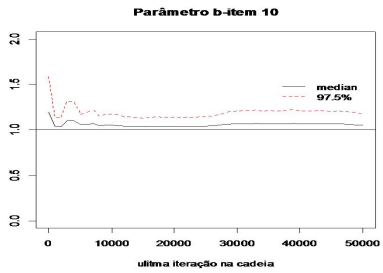
(e) Gelman e Rubin diagnostico parâmetro a2-item 10



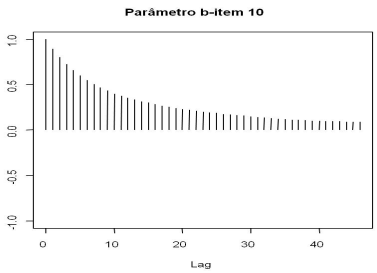
(f) Autocorrelação par. a2-item 10



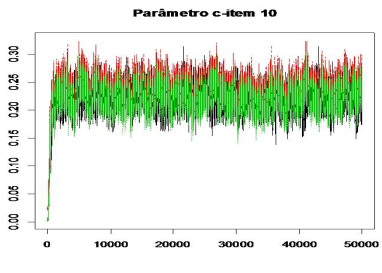
(g) Mistura cadeias parâmetro b-item 10



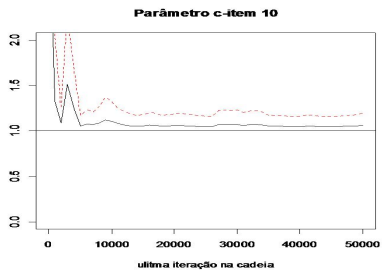
(h) Gelman e Rubin diagnostico parâmetro b-item 10



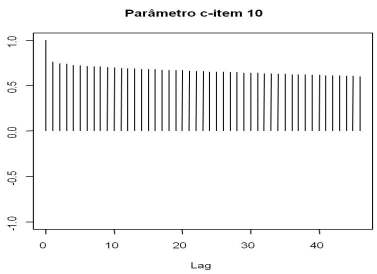
(i) Autocorrelação par. b-item 10



(j) Mistura cadeias parâmetro c-item 10

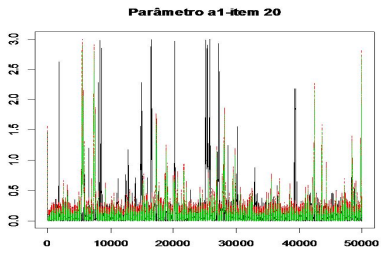


(k) Gelman e Rubin diagnostico parâmetro c-item 10

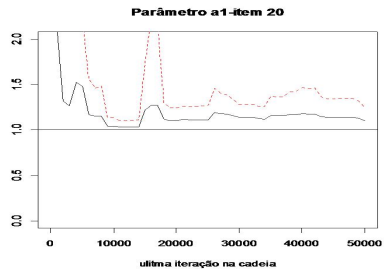


(l) Autocorrelação par. c-item 10

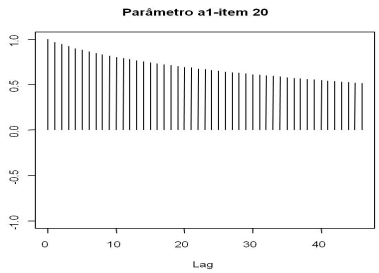
Figura B.10: Sahu com Gonzalez item 10



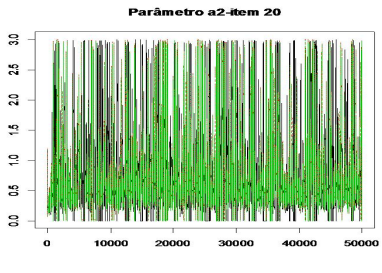
(a) Mistura cadeias parâmetro a1-item 20



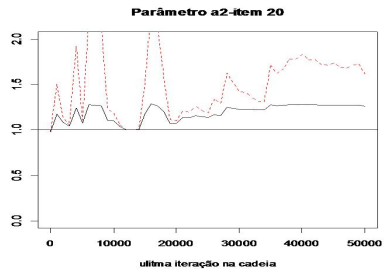
(b) Gelman e Rubin diagnostico parâmetro a1-item 20



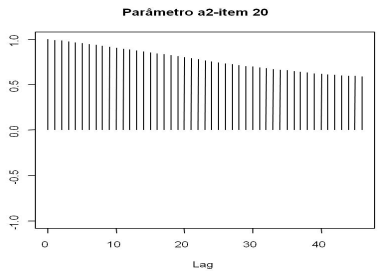
(c) Autocorrelação par. a1-item 20



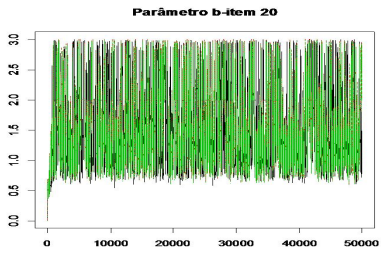
(d) Mistura cadeias parâmetro a2-item 20



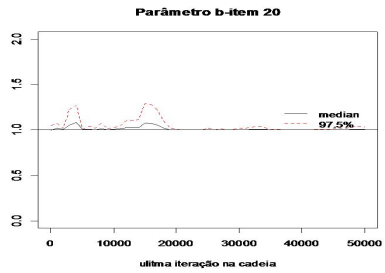
(e) Gelman e Rubin diagnostico parâmetro a2-item 20



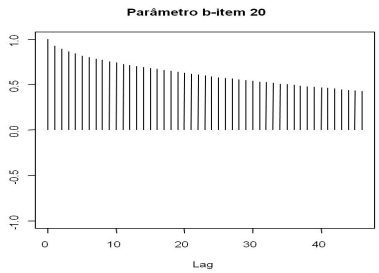
(f) Autocorrelação par. a2-item 20



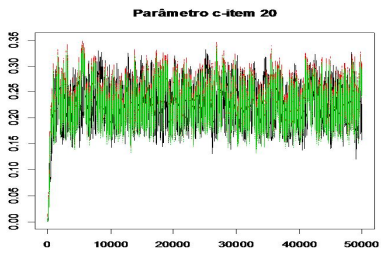
(g) Mistura cadeias parâmetro b-item 20



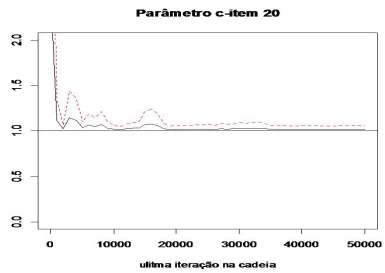
(h) Gelman e Rubin diagnostico parâmetro b-item 20



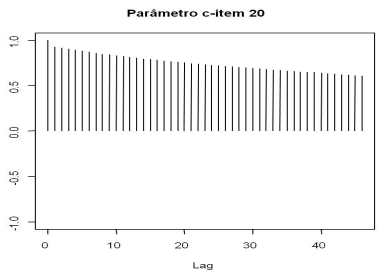
(i) Autocorrelação par. b-item 20



(j) Mistura cadeias parâmetro c-item 20

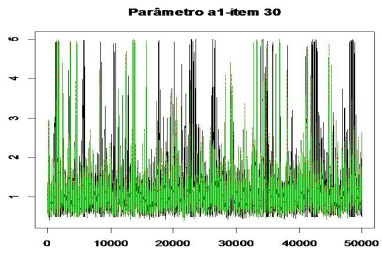


(k) Gelman e Rubin diagnostico parâmetro c-item 20

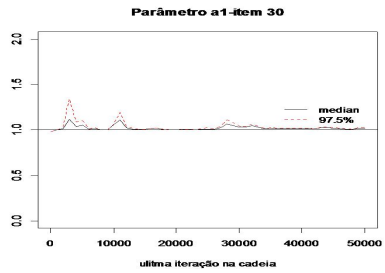


(l) Autocorrelação par. c-item 20

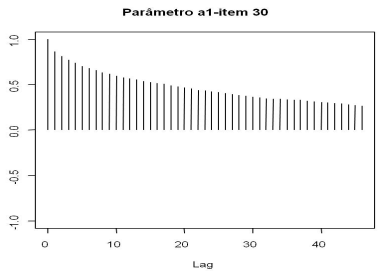
Figura B.11: Sahu com Gonzalez item 20



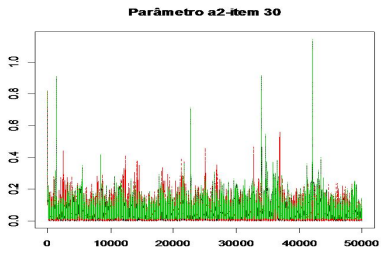
(a) Mistura cadeias parâmetro a1-item 30



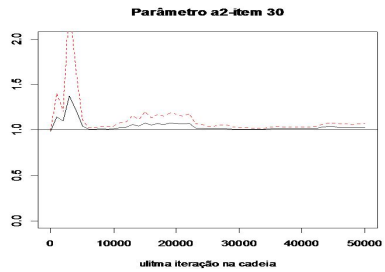
(b) Gelman e Rubin diagnostico parâmetro a1-item 30



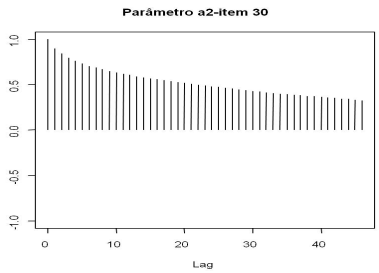
(c) Autocorrelação par. a1-item 30



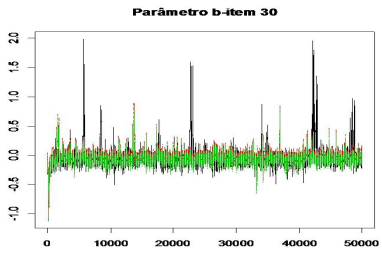
(d) Mistura cadeias parâmetro a2-item 30



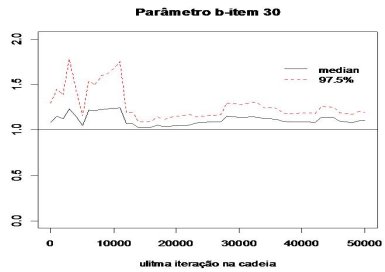
(e) Gelman e Rubin diagnostico parâmetro a2-item 30



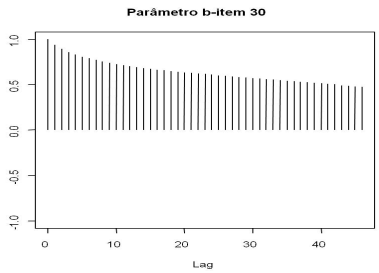
(f) Autocorrelação par. a2-item 30



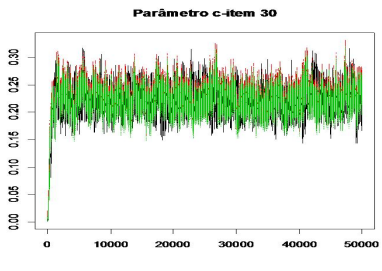
(g) Mistura cadeias parâmetro b-item 30



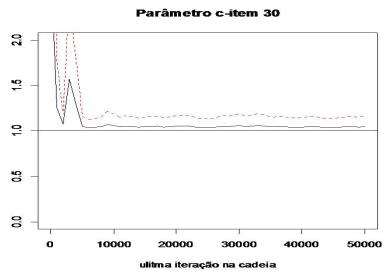
(h) Gelman e Rubin diagnostico parâmetro b-item 30



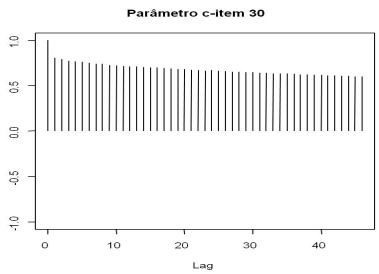
(i) Autocorrelação par. b-item 30



(j) Mistura cadeias parâmetro c-item 30



(k) Gelman e Rubin diagnostico parâmetro c-item 30



(l) Autocorrelação par. c-item 30

Figura B.12: Sahu com Gonzalez item 30

Referências Bibliográficas

Ackerman, Terry A (1994). "Using multidimensional item response theory to understand what items and tests are measuring". Em: *Applied Measurement in Education* 7.4, pp. 255–278.

Albert, J.H (1992). "Bayesian estimation of normal ogive item response curves using Gibbs sampling". Em: *Journal of Educational Statistics* 17, pp. 251–269.

Andrade, D.F, H.R. Tavares e R. C. Valle (2000). *Teoria da Resposta ao Item: Conceitos e Aplicacoes*. Sao Paulo: ABE.

Andrich, David (1978). "A rating formulation for ordered response categories". Em: *Psychometrika* 43.4, pp. 561–573.

Arellano-Valle, R. B e Adelchi Azzalini (2008). "The centred parametrization for the multivariate skew-normal distribution". Em: *Journal of Multivariate Analysis* 99.7, pp. 1362 –1382.

Arellano-Valle, R. B e M. G. Genton (2005). "On fundamental skew distributions". Em: *Journal of Multivariate Analysis* 96.1, pp. 93–116.

Arellano-Valle, R.B., H Bolfarine e V.H. Lachos (2005). "Skew-normal linear mixed models". Em: *Journal of Data Science* 3.4, pp. 415–438.

Arellano-Valle, Reinaldo B, Héctor W Gómez e Fernando A Quintana (2004). "A new class of skew-normal distributions". Em: *Communications in Statistics-Theory and Methods* 33.7, pp. 1465–1480.

Azevedo, C. L. N. e D.F. Andrade (2013). "CADEM: A conditional augmented data EM algorithm for fitting one parameter probit models". Em: *Brazilian Journal of Probability and Statistics*.

Azevedo, C. L.N., Heleno Bolfarine e Dalton F. Andrade (2012). "Parameter recovery for a skew-normal IRT model under a Bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size". Em: *Journal of Statistical Computation and Simulation* 82.11, pp. 1679–1699.

Azevedo, Caio L.N., Heleno Bolfarine e Dalton F. Andrade (2011). "Bayesian inference for a skew-normal {IRT} model under the centred parameterization". Em: *Computational Statistics and Data Analysis* 55.1, pp. 353 –365.

Azzalini, A. e A. Dalla Valle (1996). "The Multivariate Skew-Normal Distribution". Em: *Biometrika* 83.4, pp. 715–726.

Azzalini, Adelchi (1985). "A class of distributions which includes the normal ones". Em: *Scandinavian journal of statistics*, pp. 171–178.

— (2005). "The Skew-normal Distribution and Related Multivariate Families*". Em: *Scandinavian Journal of Statistics* 32.2, pp. 159–188.

Azzalini, Adelchi e Antonella Capitanio (1999). "Statistical applications of the multivariate skew normal distribution". Em: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3, pp. 579–602.

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment e Evaluation.

Bazan, J. L. (2005). "Uma Familia de Modelos de Resposta ao Item Normal Assimetrica". Tese de doutorado. Instituto de Matematica e Estatistica Universidade de Sao Paulo.

Bazan, J. L., M. D. Branco e H. Bolfarine (2006). "A Skew Item Response Model". Em: *Bayesian Analysis* 1.4, pp. 861 –892.

Beguin, A.A. e C.A.W. Glas (2001). "MCMC estimation and some model-fit analysis of multidimensional IRT models". Em: *Psychometrika* 66.4, pp. 541–561.

Birnbaum, A. (1957). *Efficient Design and Use of Tests of a Mental Ability for Varius Decision Making Problems*. Series Report 58 - 16. Randolsph Air Force Base, Texas: USAF School of Aviation Medicine.

Bock, R Darrell e Michele F Zimowski (1987). "Contributions of the biometrical approach to individual differences in personality measures". Em: *Behavioral and Brain Sciences* 10.01, pp. 17–18.

Bock, R.Darrell e Murray Aitkin (1981). "Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm". Em: *Psychometrika* 46.4, pp. 443–459.

Bolt, Daniel M e Venessa F Lall (2003). "Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo". Em: *Applied Psychological Measurement* 27.6, pp. 395–414.

Brooks, Stephen P e Andrew Gelman (1998). "General methods for monitoring convergence of iterative simulations". Em: *Journal of computational and graphical statistics* 7.4, pp. 434–455.

Camilli, Gregory (1994). "Teachers corner: Origin of the scaling constant d= 1.7 in item response theory". Em: *Journal of Educational and Behavioral Statistics* 19.3, pp. 293–295.

Courville, G. (2004). "An Empirical Comparison of Item Response Theory and Classical Test Theory and Classical Test Theory Item Person Statistics". Tese de doutorado. Texas A e M University.

Crocker, L. e J Algina (1986). *Introduction to Classical and Modern Test Theory*. Rinehart e Winston.

D. F., Andrade e Tavares H. R. (2005). "Item response theory for longitudinal data: population parameter estimation". Em: *Journal of Multivariate Analysis* 95.1, pp. 1 –22.

Darrell Bock, R. (1972). "Estimating item parameters and latent ability when responses are scored in two or more nominal categories". Em: *Psychometrika* 37.1, pp. 29–51.

Darrell Bock, R. e Marcus Lieberman (1970). "Fitting a response model forn dichotomously scored items". Em: *Psychometrika* 35.2, pp. 179–197.

Embretson, S.E. e S.P. Reise (2000). *Item Response Theory*. Multivariate Applications Series. Taylor & Francis.

Fox, J.P. *Bayesian Item Response Modeling: Theory and Applications*. Statistics for social and behavioral sciences. ISBN: 9781441907424.

Fragoso, T. M. (2010). "Modelos multidimensionais da teoria de resposta ao item". Diss. de mestrado. Departamento de Matematica Aplicada e Estatistica ICMC, USP.

Fraser, Colin e Roderick P. McDonald (1988). "NOHARM: Least Squares Item Factor Analysis". Em: *Multivariate Behavioral Research* 23.2, pp. 267–269.

Frühwirth-Schnatter, Sylvia e Saumyadipta Pyne (2010). "Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions". Em: *Biostatistics* 11.2, pp. 317–336.

Fu, Zhi-Hui, Jian Tao e Ning-Zhong Shi (2009). "Bayesian estimation in the multidimensional three-parameter logistic model". Em: *Journal of Statistical Computation and Simulation* 79.6, pp. 819–835.

Fung, Thomas e Eugene Seneta (2010). "Modelling and estimation for bivariate financial returns". Em: *International statistical review* 78.1, pp. 117–133.

Gammerman, D. e H. F. Lopes (2006). *Markov Chain Monte Carlo: Sthocastic simulation for bayesian inference*. Vol. 16. Chapman e Hall CRC.

Gelfand, Alan E e Adrian FM Smith (1990). "Sampling-based approaches to calculating marginal densities". Em: *Journal of the American statistical association* 85.410, pp. 398–409.

Gelman, Andrew et al. (2013). *Bayesian data analysis*. CRC press.

Geman, Stuart e Donald Geman (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". Em: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6.6, pp. 721–741.

Genton, M.G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*.

Gonzalez, L. R. (2004). *Data Augmentation in the Bayesian Multivariate Probit Model*. Sheffield Economic Research Paper Series. Department of Economics University of Sheffield.

Gulliksen, H. (1950). *Theory of Mental Tests*. Wiley Publications in Psychology.

Hambleton, R.K. e H. Swaminathan (1985). *Item Response Theory: Principles and Applications*. Evaluation in education and human services. Springer. ISBN: 9780898380651.

Hambleton, R.K., H. Swaminathan e H.J. Rogers (1991). *Fundamentals of Item Response Theory*. Measurement Methods for the Social Science.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". Em: *Biometrika* 57.1, pp. 97–109.

Henze, Norbert (1986). "A probabilistic representation of the skew-normal distribution". Em: *Scandinavian journal of statistics*, pp. 271–275.

Holman, Rebecca et al. (2004). "Practical methods for dealing with'not applicable'item responses in the AMC Linear Disability Score project". Em: *Health and quality of life outcomes* 2.1, p. 29.

Jiang, Y. (2005). "Estimating Parameters for Multidimensional Item Response Theory Model by MCMC Methods". Tese de doutorado. Michigan State University.

Kolen, M.J. e R.L. Brennan (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. Statistics for Social and Behavioral Sciences.

Lachos, V. H. (2004). "Modelos Lineares Mistos Assimetricos". Tese de doutorado. Instituto de Matematica e Estatistica da Universidade de Sao Paulo.

Li, Yuan H e Robert W Lissitz (2000). "An evaluation of the accuracy of multidimensional IRT linking". Em: *Applied Psychological Measurement* 24.2, pp. 115–138.

Liseo, Brunero e Nicola Loperfido (2006). "A note on reference priors for the scalar skew-normal distribution". Em: *Journal of Statistical Planning and Inference* 136.2, pp. 373 –389.

Liseo, Brunero e Antonio Parisi (2013). "Bayesian inference for the multivariate skew-normal model: A population Monte Carlo approach". Em: *Computational Statistics & Data Analysis*.

Lord, F. (1952). "A Theory of Test Scores". Em: *Psychometric Monograph* 7.

— (1953). "The Relation of Test Score to the Trait Underlying the Test." Em: *Educational and Psychological Measurement* 13, pp. 517 –548.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum Associates. ISBN: 9780898590067.

Lord, Frederic M, Melvin R Novick e Allan Birnbaum (1968). "Statistical theories of mental test scores." Em:

Masters, GeoffN. (1982). "A rasch model for partial credit scoring". Em: *Psychometrika* 47.2, pp. 149–174.

Matos, G. S. (2001). "Teoria de Resposta ao Item: Uma proposta de modelo multivariado". Diss. de mestrado. UFPE.

— (2008). "Modelos Multidimensionais na TRI com Distribuicoes assimetricass para os tracos latentes". Tese de doutorado. IME, USP.

Maydeu-Olivares, Albert (2001). "Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates". Em: *Journal of Educational and Behavioral Statistics* 26.1, pp. 51–71.

McDonald, Roderick P (2000). "A basis for multidimensional item response theory". Em: *Applied Psychological Measurement* 24.2, pp. 99–114.

McKinley, R. e M. D. Reckase (1983). *An Extension of the Two Parameter Logistic Model to the Multidimensional Latent Space*. Rel. téc. American College Testing Program.

Metropolis, Nicholas, Arianna W. Rosenbluth e Rosenbluth (dez. de 1972). "Equation of State Calculations by Fast Computing Machines". Em: *The Journal of Chemical Physics* 21.6, pp. 1087–1092.

Micceri, T. (1989). "The Unicorn, The Normal Curve, and Other Improbable Creatures". Em: *Psychhological Bulletin* 105.1, pp. 156 –166.

Mislevy, Robert J (1986). "Bayes modal estimation in item response models". Em: *Psychometrika* 51.2, pp. 177–195.

Montenegro, Alvaro M. "Multidimensional item response models". Em: *Tese Doutorado Universidad Nacional de Colombia*.

Muthen, Bengt (1978). "Contributions to factor analysis of dichotomous variables". Em: *Psychometrika* 43.4, pp. 551–560.

Nojosa, R. T. (2001). "Modelos Multidimensionais na Teoria de Resposta ao Item". Diss. de mestrado. UFPE.

— (2010). "Inferencia Bayesiana em Modelos Multidimensionais de Resposta ao Item". Tese de doutorado. IME, USP.

Patz, Richard J e Brian W Junker (1999a). "A straightforward approach to Markov chain Monte Carlo methods for item response models". Em: *Journal of educational and behavioral Statistics* 24.2, pp. 146–178.

— (1999b). "Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses". Em: *Journal of educational and behavioral statistics* 24.4, pp. 342–366.

R Core Team (2012a). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

— (2012b). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

Rasch, G. (1960). "Probabilistic Models for Some Intelligence and Attainment Tests". Em: *Copenhagen: Danish Institute for Educational Research*.

Reckase, M. (2009). *Multidimensional Item Response Theory*. Statistics for social and behavioral sciences. Springer.

Reckase, M. D. (1985). *The difficulty of Test Items That Measure More than One Ability*. Rel. téc. American College Testing Program.

— (1986). *The discrimination Power of Items that Measure More than One Dimension*. Rel. téc. American College Testing Program.

Reckase, M. D. e R. L. McKinley (1983). *The Definition of Difficulty and Discrimination for Multidimensional Item Response Theory Models*. Rel. téc. American College Testing Program.

Reckase, Mark D., Terry A. Ackerman e James E. Carlson (1988). "Building a Unidimensional Test Using Multidimensional Items". Em: *Journal of Educational Measurement* 25.3, pp. 193–203.

Richardson, M.W. (1936). "The relation between the difficulty and the differential validity of a test". Em: *Psychometrika* 1.2, pp. 33 –49.

Rivers, Douglas (2003). "Identification of multidimensional spatial voting models". Em: *Typescript. Stanford University*.

Sahu, Sujit K. (2002). "Bayesian Estimation and Model Choice in Item Response Models". Em: *Journal of Statistical Computation and Simulation* 72.3, pp. 217–232.

Sahu, Sujit K, Dipak K Dey e Marcia D Branco (2003). "A new class of multivariate skew distributions with applications to Bayesian regression models". Em: *Canadian Journal of Statistics* 31.2, pp. 129–150.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Psychometrika: Monograph supplement.

Samejima, Fumiko (1997). "Departure from normal assumptions: A promise for future psychometrics with substantive mathematical modeling". Em: *Psychometrika* 62.4, pp. 471–493.

Santos, J. R. S. (2012). "Um Modelo de Resposta ao Item para Grupos Multiplos com Distribuicoes Normais Assimetricas Centralizadas". Diss. de mestrado. Departamento de Estatistica, IMECC, UNICAMP.

Santos, J. R. S., C. L. N. Azevedo e H. Bolfarine (2013). "A multiple group Item Response Theory model with centred skew normal latent trait distributions under a Bayesian framework". Em: *Journal of Applied Statisitcs* 40.10, pp. 2129 –2149.

Seong, Tae-Je (1990). "Sensitivity of Marginal Maximum Likelihood Estimation of Item and Ability Parameters to the Characteristics of the Prior Ability Distributions". Em: *Applied Psychological Measurement* 1.3, pp. 229 –311.

Sheng, Yanyan (2008). "Markov Chain Monte Carlo estimation of normal ogive IRT models in MATLAB". Em: 25.8.

Sheng, Yanyan e Christopher K Wikle (2008). "Bayesian multidimensional IRT models with a hierarchical structure". Em: *Educational and Psychological Measurement* 68.3, pp. 413–430.

Sinharay, Sandip (2006). "Bayesian item fit analysis for unidimensional item response theory models". Em: *British Journal of Mathematical and Statistical Psychology* 59.2, pp. 429–449.

Swaminathan, Hariharan e Janice A Gifford (1985). "Bayesian estimation in the two-parameter logistic model". Em: *Psychometrika* 50.3, pp. 349–364.

— (1986). "Bayesian estimation in the three-parameter logistic model". Em: *Psychometrika* 51.4, pp. 589–601.

Tanner, Martin A. e Wing H. Wong (1987). "The Calculation of Posterior Distributions by Data Augmentation". Em: *Journal of the American Statistical Association* 82.398, pp. 528–540.

Toribio, S. G. (2006). "Bayesian Model Checking Strategies for Dichotomous Item Response Theory Models". Tese de doutorado. Bowling Green State University.

Torre, Jimmy de la e Richard J Patz (2005). "Making the most of what we have: A practical application of multidimensional item response theory in test scoring". Em: *Journal of Educational and Behavioral Statistics* 30.3, pp. 295–311.

Traub, R. (1997). "Classical Test Theory in Historical Perspective". Em: *Educational Measurement: Issues and Practice* 16.

Wilson, D.T. et al. (1991). *Testfact: test scoring, item statistics, and item factor analysis*. SSI, Scientific Software International.