

Modelos de Resposta ao Item com função de ligação t-Assimétrica

Alessandra Noeli Craveiro Pinheiro

Orientador: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos

Abril de 2007

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

P654ar

Pinheiro, Alessandra Noeli Craveiro.

Modelos de resposta ao item com função de ligação t -
assimétrica / Alessandra Noeli Craveiro Pinheiro. -- São
Carlos : UFSCar, 2007.

101 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2007.

1. Teoria de resposta ao item. 2. Distribuição normal
assimétrica. 3. Distribuição t – assimétrica. 4. Inferência
bayesiana. 5. Pseudo fator de Bayes. I. Título.

CDD: 519.542 (20^a)

Agradecimentos

Agradeço inicialmente a Deus, pois sem ele nada é possível.

À minha mãe pelo apoio emocional, carinho e confiança.

Aos amigos que estiveram do meu lado contribuindo com críticas e sugestões, incentivo, apoio moral e paciência.

Ao meu orientador Professor Josemar Rodrigues, pelos conselhos, acompanhamento e revisão do estudo.

Aos membros da banca examinadora pelas correções e sugestões para a dissertação.

Aos meus professores de graduação e de pós-graduação que contribuíram para a minha formação acadêmica.

A CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo auxílio concedido.

Resumo

A Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos que representam a probabilidade de um indivíduo dar uma resposta certa a um item (questão) como função dos parâmetros do item e da habilidade do indivíduo. O objetivo de nossa pesquisa é apresentar os modelos propostos na TRI normal assimétrica e desenvolver modelos alternativos mais flexíveis. Com esta finalidade em mente, introduzimos a distribuição t-assimétrica (Azzalini e Capitanio 1999) e obtemos resultados similares aos obtidos por Bazán (2005).

Algumas aplicações utilizando métodos bayesianos são consideradas.

Palavras-chave: Teoria de Resposta ao Item, distribuição normal assimétrica, distribuição t-assimétrica, análise Bayesiana, Pseudo Fator de Bayes, *Conditional Predictive Ordinate*.

Abstract

The Item Response Theory (IRT) is a set of mathematical models representing the probability of an individual to take a correct response of an item and its ability. The purpose of our research is to show the models formulated in the IRT under the skew-normal distributions and to develop flexible alternative models. With this goal in mind we introduced the t-skew distributions (Azzalini et al. 1999) and results similar to Bazan's results are obtained.

Some applications using Bayesian methods are also considered.

Keywords: Item Response Theory; skew-normal distributions; t-skew distributions; Bayesian analysis; Pseudo Bayes factor.

Sumário

Lista de Figuras	iv
Lista de Tabelas	v
Lista de Siglas	vii
1 Introdução	1
2 Modelos Simétricos na Teoria de Resposta ao Item	3
2.1 O modelo de resposta ao item dicotômico	4
2.1.1 Características do modelo	4
2.1.2 Interpretação e Representação Gráfica	5
2.2 Estimação	8
2.3 Inferência Bayesiana no modelo probito normal	9
2.4 MCMC usando amostrador de Gibbs ao considerar um esquema de dados aumentados (DAGS)	10
2.5 Estruturas Hierárquicas	13
3 Teoria de Resposta ao Item Normal Assimétrica (TRI-NA)	14
3.1 O modelo de Chen <i>et al</i>	15
3.2 Modelo Probito assimétrico generalizado	19

3.2.1	O modelo	20
3.3	Verossimilhança	20
3.4	Verossimilhança aumentada para o modelo probito assimétrico generalizado	21
3.5	Análise Bayesiana	22
3.5.1	Especificação de prioris	22
3.6	Esquema de dados aumentados para o amostrador de Gibbs	23
3.7	A família TRI-NA	24
3.7.1	Modelos	24
3.7.2	Probito Assimétrico-Normal	25
3.7.3	Probito Normal-Assimétrico	27
3.7.4	Probito Assimétrico Normal Assimétrico	27
3.7.5	Algoritmo de dados aumentados na TRI-NA	27
3.7.6	Especificação de prioris	29
3.7.7	Distribuições condicionais completas	30
3.7.8	Implementação MCMC	32
4	Teoria de Resposta ao Item t-Assimétrica	34
4.1	Distribuição t-assimétrica	35
4.2	Verossimilhança aumentada para o modelo probito t-assimétrico na TRI . .	39
4.3	Estrutura Hierárquica	42
4.4	Discussão	42
5	Aplicações	43
5.1	Seleção de Modelos	43
5.1.1	DIC (<i>Deviance Information Criterion</i>)	43
5.1.2	Pseudo Fator de Bayes	44

5.1.3	Soma dos quadrados dos resíduos	45
5.2	Aplicação 1	45
5.2.1	Aplicação ao modelo PN	47
5.3	Aplicação 2	53
5.4	Aplicação 3	58
5.4.1	Aplicação aos modelos probito t-assimétricos	62
6	Considerações Finais	68
A	A distribuição Normal Assimétrica	70
A.1	Propriedades Matemáticas	70
A.2	Obtendo a Distribuição Normal Assimétrica	78
A.3	Momentos	80
B	Distribuições esféricas e elípticas	83
B.1	As distribuições Esféricas Multivariadas	83
B.2	As distribuições Elípticas Multivariadas	84
B.3	A distribuição Elíptica Assimétrica	86
B.3.1	Propriedades	87
B.3.2	Obtenção da distribuição t-Assimétrica	87
C	Programas	88
C.1	Programa em R para o modelo PN	88
C.2	Programa t-assimétrico	90
C.3	Programa para a família TRI-NA	92
	Referências Bibliográficas	98

Lista de Figuras

2.1	Curva característica do item-CCI.	6
2.2	Interpretação da CCI.	7
3.1	CCI probito assimétricos para $a=1$, $b=0$, e diferentes valores do parâmetro de assimetria $\lambda = -2, -1, 0, 1, 2$	26
4.1	Densidades para diferentes g.l da distribuição t comparados com a densidade da distribuição normal padronizada.	35
5.1	Histograma dos escores.	46
5.2	Boxplot dos escores.	47
5.3	Teste de convergência para os parâmetros $a[12]$ e $b[12]$	49
5.4	Densidade, histórico e autocorrelação da cadeia para os parâmetros do item 12.	50
5.5	CCI para os itens 9,11 e 12.	52
5.6	Boxplot dos escores.	53
5.7	Gráfico de Gelman e Rubin, densidade e histórico para o item 5.	56
5.8	Boxplot para o parâmetro de assimetria δ	57
5.9	CCI para os modelos PN e PAN nos itens mais significativos do teste. . . .	58
5.10	Boxplot para o parâmetro de assimetria δ	60
5.11	Densidade tA comparada com a densidade NA.	63

5.12 Tendência do DIC.	64
5.13 Boxplot para o parâmetro de assimetria δ	66
A.1 NA(3)	72
A.2 NA(-3)	72
A.3 NA(5,4,3).	77

Lista de Tabelas

5.1	<i>Estatísticas dos escores</i>	46
5.2	<i>Priori para os parâmetros dos itens</i>	48
5.3	<i>Estatísticas do modelo Probit Normal</i>	51
5.4	<i>Estatísticas dos escores</i>	53
5.5	<i>Comparando modelos usando Priori 1 e $U \sim N(0, 1)$</i>	54
5.6	<i>Estatísticas do modelo PAN com priori 1</i>	55
5.7	<i>Estatísticas do parâmetro de assimetria δ no modelo PAN</i>	55
5.8	<i>Prioris utilizadas</i>	58
5.9	<i>Comparando modelos usando a priori 1</i>	59
5.10	<i>Comparando modelos usando a priori 2</i>	59
5.11	<i>Estatísticas para o parâmetro δ no modelo PANA</i>	60
5.12	<i>Estatísticas do modelo PANA com priori 1</i>	61
5.13	<i>Comparando modelos com a priori 2 e $U \sim N(0, 1)$</i>	64
5.14	<i>Comparando modelos com a priori 2 e $U \sim NA(0, 1, k)$</i>	65
5.15	<i>Modelos com o uso da priori 1 e $U \sim N(0, 1)$</i>	65
5.16	<i>Modelos com o uso da priori 1 e $U \sim NA(0, 1, k)$</i>	65
5.17	<i>Estatísticas do modelo PtA-N com priori 2</i>	67

Lista de Siglas

CCI	Curva Característica do Item;
CPO	<i>Conditional Predictive Ordinate</i> ;
DAGS	<i>Data augmented Gibbs sampling</i> ;
DIC	<i>Deviance Information Criterion</i> ;
f.d.a	Função Distribuição Acumulada;
f.d.p	Função Densidade de Probabilidade;
G	Gama;
GI	Gama Inversa;
g.l	graus de liberdade;
L1	Função de ligação dada por $\Phi_{NA}(\lambda)$;
L2	Função de ligação dada por $\Phi_{NA}(0, 1 + \lambda^2, \lambda)$;
LN	Log Normal;
MCMC	<i>Markov Chain Monte Carlo</i> ;
NA	Normal Assimétrica;
PAN	Probito Assimétrico Normal;
PANA	Probito Assimétrico Normal Assimétrico;
PN	Probito Normal;
PtA-N	Probito t-Assimétrico Normal;
PtA-NA	Probito t-Assimétrico Normal Assimétrico;
PFB	Pseudo Fator de Bayes;
SQRL	Soma dos Quadrados Residuais Latentes;
tA	t-Assimétrica;
TRI	Teoria de Resposta ao Item;
TRI-NA	Teoria de Resposta ao Item Normal Assimétrica;

Capítulo 1

Introdução

Atualmente é crescente o interesse em avaliar algumas variáveis que não são observadas diretamente, conhecidas como variáveis latentes, como por exemplo: ansiedade, depressão e habilidade de um aluno em determinada matéria. Uma ferramenta muito útil na resolução desse tipo de problema é a TRI (Teoria de Resposta ao Item). Segundo Hambleton e Swaminathan (1985) a TRI é um conjunto de modelos matemáticos usados para fazer previsões, estimativas ou inferências sobre as habilidades (ou competências) medidas em um teste. Esta técnica vem sendo fortemente empregada na área educacional, pois propõe modelos que representam a relação entre a probabilidade de uma resposta certa a um item e a habilidade de um aluno.

Muitos dos modelos propostos na TRI consideram que tanto a habilidade quanto a CCI (Curva Característica do Item) possuem distribuições simétricas. A CCI representa a probabilidade de uma resposta certa em função dos parâmetros dos itens. Porém, ao se trabalhar com conjuntos de dados que envolvem a conduta humana, supor que a distribuição dos dados é simétrica não é muito coerente. Ao supormos simetria na distribuição dos dados, estamos considerando que todos os alunos possuem habilidades ou competências parecidas, ou seja, em um teste não há variabilidade entre as respostas dos alunos. Mas sabemos que, geralmente, dentro de uma sala de aula isso pode não ocorrer.

O objetivo desta dissertação é desenvolver um modelo na TRI que além de detectar a assimetria presente na distribuição dos dados é capaz de captar a presença de pontos ex-

tremos. Este modelo é uma “extensão” dos modelos assimétricos já existentes, ver Bazán (2005), e envolve a distribuição t-assimétrica que é menos sensível a pontos discrepantes que a distribuição normal assimétrica.

No Capítulo 2 exibimos o modelo Probit Normal (PN) que embora seja um modelo simétrico, é de grande importância para a realização deste trabalho, pois foi a base para o estudo dos modelos assimétricos.

No Capítulo 3 estudamos o trabalho desenvolvido por Chen *et al* (1999) usando variáveis latentes auxiliares na construção do modelo probito assimétrico generalizado. Este modelo é motivado pelo uso de variáveis latentes (Albert e Chib, 1993), onde a variável latente segue um modelo com estrutura de efeito aleatório. Fazemos também, um estudo da família normal assimétrica na TRI dada por Bazán (2005), que trabalha sob a perspectiva bayesiana.

No Capítulo 4 apresentamos inicialmente, as principais características da distribuição t-assimétrica, como por exemplo a sua representação estocástica, e em seguida apresentamos resultados similares aos de Bazán (2005), mas utilizando a distribuição t-assimétrica. Estes resultados são de grande importância para a extensão dos modelos probito assimétricos para os modelos probito t-assimétricos na TRI.

No Capítulo 5 fazemos uma aplicação da metodologia descrita neste trabalho a dois conjuntos de dados distintos, visando verificar o comportamento dos modelos estudados. Devido à grande quantidade de modelos que trabalhamos, usamos alguns métodos de seleção de modelos como o DIC (*Deviance information Criterion*), SQRL (Soma dos quadrados residuais latentes) e o PFB (Pseudo Fator de Bayes).

No Capítulo 6 apresentamos uma conclusão sobre o trabalho desenvolvido e algumas sugestões para trabalhos futuros.

São apresentados três Apêndices: o Apêndice A mostra as principais características da distribuição Normal Assimétrica, o Apêndice B traz algumas informações sobre as distribuições esféricas e elípticas, essências para a construção da distribuição t-assimétrica e o Apêndice C apresenta os conjuntos de dados e programas utilizados neste trabalho, programas estes desenvolvidos nos softwares livres R (<<http://www.r-project.org/>>) e Winbugs (<<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>>).

Capítulo 2

Modelos Simétricos na Teoria de Resposta ao Item

Segundo Andrade *et al* (2000), o interesse na aplicação de técnicas derivadas da Teoria de Resposta ao Item vem aumentando, principalmente na área educacional, devido ao fato de podermos obter com estas técnicas modelos para os traços latentes, que são características do indivíduo que não podem ser observadas diretamente, como por exemplo, a inteligência, habilidade em executar uma tarefa, ansiedade, dentre outros.

O que a metodologia da Teoria de Resposta ao Item (TRI) sugere são algumas formas de representar a relação existente entre a probabilidade de um aluno responder corretamente a um item e suas habilidades ou traços latentes na área de conhecimento avaliada.

Os primeiros modelos de TRI surgiram na década de 50, e eram modelos em que se considerava que uma única habilidade, de um único grupo, estava sendo medida por um teste onde os itens eram corrigidos de maneira dicotômica. Estes modelos foram primeiramente desenvolvidos baseado na distribuição acumulada da normal e, depois, foram descritos para uma forma matemática mais conveniente, a logística, pois a função distribuição acumulada desta é uma função explícita dos parâmetros dos itens e da habilidade, diferentemente da função distribuição acumulada da normal.

2.1 O modelo de resposta ao item dicotômico

Os vários modelos propostos na literatura dependem basicamente de quatro fatores: da natureza do item, dicotômicos ou não, do número de populações envolvidas, da quantidade de traços latentes (habilidades) que estamos medindo e do número de parâmetros dos itens. O modelo estudado neste capítulo é um modelo dicotômico com apenas um traço latente, ou seja, medindo apenas uma habilidade e com dois parâmetros para os itens.

2.1.1 Características do modelo

Mostraremos em seguida as principais características do modelo de resposta ao item dicotômico.

Seja

$$Y_{ij}|u_i, \beta_j \sim \text{Bernoulli}(p_{ij}), \quad (2.1)$$

onde y_{ij} são valores de Y_{ij} correspondentes às respostas dicotômicas do indivíduo i no item j , considere o valor 1 para indicar acerto e zero caso contrário, p_{ij} é uma função dos parâmetros β_j e u_i , em que $\beta_j = (a_j, b_j)'$, onde a_j e b_j correspondem aos parâmetros de discriminação e dificuldade associados ao item j , respectivamente, e u_i é o valor correspondente da variável latente U_i associada ao indivíduo i , descrevendo a habilidade medida no teste.

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

$$p_{ij} = P(Y_{ij} = 1|u_i, \beta_j) = F(m_{ij}), \quad (2.2)$$

com $m_{ij} = a_j u_i - b_j$, uma função linear de u_i .

A função $F(\cdot)$, em (2.2) é dada pela função distribuição acumulada de uma determinada distribuição e é conhecida como a função de resposta ao item ou curva característica do item e esta tem como característica ser uma função estritamente não decrescente de U_i .

A suposição de independência condicional é feita para a estimação dos parâmetros dos itens, ou seja, para um indivíduo i , as respostas Y_{ij} , $j = 1, \dots, k$, são condicionalmente

independentes dado os valores das variáveis latentes U_i , $i = 1, \dots, n$, isto é, o indivíduo não “aprende” ao longo da realização do teste. Além disso também é considerada a independência entre as respostas de diferentes indivíduos.

Sob as suposições apresentadas pelas equações (2.1) e (2.2) e pela suposição de independência condicional na TRI, a distribuição conjunta da matriz de respostas $Y_{n \times k}$, dado o vetor de variáveis latentes $U = (U_1, \dots, U_n)'$ e o vetor de parâmetros de item $\beta = (\beta_1, \dots, \beta_k)'$ pode ser escrita como

$$P(Y = y | U = u, \beta = \beta) = \prod_{i=1}^n \prod_{j=1}^k F(m_{ij})^{y_{ij}} (1 - F(m_{ij}))^{1-y_{ij}} \quad (2.3)$$

De acordo com a função distribuição acumulada escolhida para F , podemos obter diferentes modelos na TRI. O modelo probito é obtido se F é uma distribuição acumulada de uma normal padrão e o modelo logito é obtido se F é uma distribuição acumulada logística. Neste capítulo o modelo estudado é o modelo Probit Normal.

2.1.2 Interpretação e Representação Gráfica

De acordo com Andrade *et al* (2000), na Teoria de Resposta ao Item a habilidade pode assumir qualquer valor no intervalo $(-\infty, \infty)$. Assim é comum estabelecer-se uma origem e uma unidade de medida para a definição da escala. Esses valores podem ser escolhidos de modo a representar, respectivamente, o valor médio e o desvio padrão das habilidades do grupo em estudo. Geralmente utiliza-se a escala $(0, 1)$ e é esta a escala adotada neste trabalho, porém é necessário deixar claro que a escolha da escala é arbitrária e só depende de quem está realizando o estudo.

A fim de entendermos graficamente a TRI, será mostrado em seguida um gráfico da Curva Característica do Item (CCI) que representa, como foi dito no Capítulo 1, a probabilidade de uma resposta certa em função dos parâmetros do modelo.

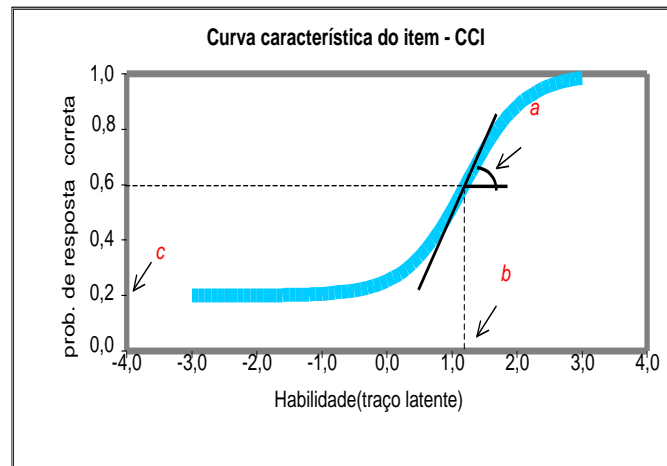


FIGURA 2.1: Curva característica do item-CCI.

A Figura 2.1 (Valle, 1999) é obtida quando consideramos o modelo logístico definido como

$$P(Y_{ij} = 1 | u_i, \gamma_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_j(u_i - b_j)}} \quad (2.4)$$

onde $\gamma_j = (a_j, b_j, c_j)$, com $i = 1, \dots, n$ e $j = 1, \dots, k$.

Existem duas propriedades técnicas da CCI, a dificuldade do item e seu poder de discriminação, que são parâmetros de nosso interesse. O modelo logístico de três parâmetros também considera a probabilidade de acerto casual. Listamos abaixo as principais características destes parâmetros:

- a_j é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto b_j .
- b_j é o parâmetro de dificuldade (ou de posição) do item j , medido na mesma escala da habilidade.
- c_j é o parâmetro do item que representa a probabilidade do indivíduo com baixa habilidade responder corretamente ao item j , ou seja, c_j representa a probabilidade de acerto casual.

Analisando a Figura 2.1 (Valle, 1999) podemos notar que quanto maior a habilidade maior a probabilidade do indivíduo acertar o item, ou seja, a CCI é uma função não

decrecente de U_i , além disso esta relação não é linear, tem a forma de S. A inclinação da curva é caracterizada pelo parâmetro a que é proporcional à derivada da tangente da curva no ponto de inflexão. Assim, somente serão esperados itens com valores positivos de a , pois caso contrário, teríamos que a probabilidade de responder corretamente o item diminuiria com a habilidade, o que não é esperado para este modelo. Baixos valores de a indicam que alunos com habilidades bastante diferentes tem aproximadamente a mesma probabilidade de responder corretamente ao item. O deslocamento é representado pelos parâmetros b e c , onde c representa a probabilidade de acerto casual, ou seja, a probabilidade de acertar quando se “chuta” uma questão.

Apresentamos abaixo dois gráficos que mostram as CCI's que obtemos quando se mantém fixo o valor de a e varia-se o valor de b e vice versa.

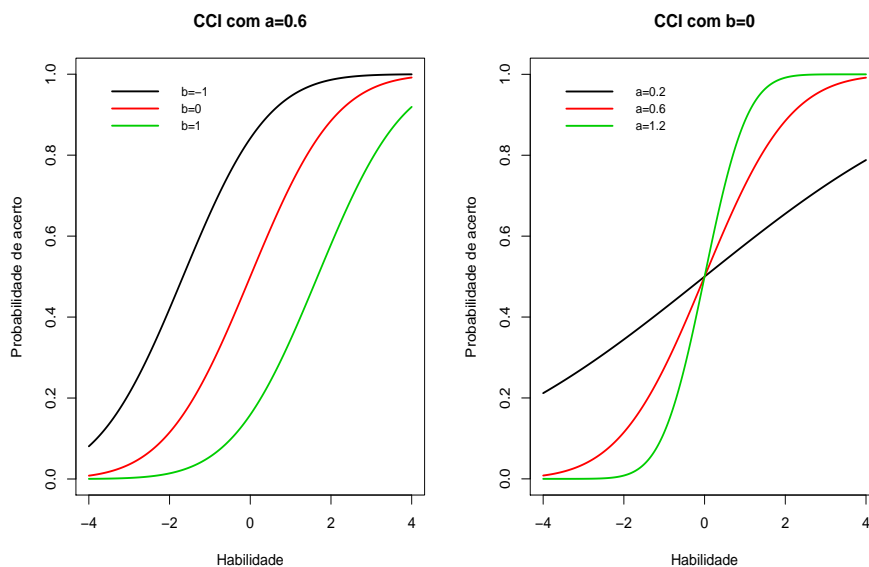


FIGURA 2.2: Interpretação da CCI.

Na Figura 2.2, no gráfico da esquerda temos 3 CCI's com mesmo nível de discriminação ($a = 0.6$) e diferentes valores para a dificuldade $b = (-1, 0, 1)$. Analisando esta figura, podemos dizer que a curva de cor preta representa um item fácil, pois para uma baixa habilidade, temos altas probabilidades de acerto, diferentemente da curva de cor verde que exige uma alta habilidade para a probabilidade de acertar uma questão. Assim poderíamos dizer que o item que corresponde à curva de cor verde é o mais difícil e o de cor vermelho tem um nível médio de dificuldade.

Ao olharmos para o gráfico da direita, podemos analisar qual a importância do parâmetro de discriminação. Neste gráfico, consideramos $b = 0$ e $a = (0.2, 0.6, 1.2)$. Ao olharmos a curva de cor verde percebemos que esta é menos inclinada em relação às outras, o que indica um maior poder de discriminação. Quando encontramos um item com esta qualidade, fica mais fácil de identificar se os alunos possuem um conhecimento específico para responder a este item ou não. Já a curva colorida de vermelho tem um médio poder discriminatório, enquanto que a curva colorida de preto tem um poder discriminatório baixo.

2.2 Estimação

Como já vimos a TRI relaciona a probabilidade de um aluno responder corretamente a um item e sua habilidade. Geralmente só conhecemos as respostas dos alunos aos itens e estamos interessados em estimar os parâmetros dos itens e suas habilidades. Em relação às habilidades é mais comum o interesse em estimar a habilidade média dos alunos envolvidos no estudo. Vários métodos de estimação são propostos na literatura, alguns clássicos e outros bayesianos.

Os métodos clássicos se utilizam da estimação por máxima verossimilhança, porém estes apresentam grandes problemas dependendo do padrão de resposta que temos, por exemplo, se temos um indivíduo que errou todos os itens, não conseguiremos estimar os parâmetros de interesse por que os estimadores de máxima verossimilhança não estão definidos neste caso. Para detalhes ver, Valle (1999). Este problema não ocorrerá nos métodos Bayesianos, que está definido para qualquer padrão de resposta.

Dentre os métodos Bayesianos temos aqueles que fazem uso da metodologia MCMC (*Markov Chain Monte Carlo*) para obter a estimação dos parâmetros de interesse e existem métodos que não necessitam desta metodologia, isto ocorre por exemplo, quando trabalhamos com famílias de prioris conjugadas. Neste trabalho, utilizaremos o método Bayesiano com a aplicação de MCMC.

A idéia da metodologia MCMC é obter uma amostra da distribuição de interesse através da simulação de passeios aleatórios. Este processo é muito utilizado em problemas onde a distribuição de interesse é multivariada. Para maiores detalhes consultar os

trabalhos de Hastings (1970) e Paulino *et al* (2003).

2.3 Inferência Bayesiana no modelo probito normal

Como foi dito na Subseção 2.1.1, obteremos o modelo probito normal quando $F(\cdot)$ denotar a função distribuição acumulada da normal padronizada.

Para obter as estimativas dos parâmetros de interesse utilizando Inferência Bayesiana é necessário inicialmente obtermos a função de verossimilhança e as prioris para os parâmetros do modelo. Consideramos $\pi(\cdot)$ como a densidade a priori. De acordo com as suposições apresentadas para a estimação dos parâmetros da TRI, temos que as habilidades dos indivíduos são independentes dos itens, assim

$$\pi(\mathbf{u}, \boldsymbol{\beta}) = \pi_1(\mathbf{u})\pi_2(\boldsymbol{\beta}).$$

Além disso, também podemos considerar que as habilidades u_i são independentes para cada indivíduo, bem como os parâmetros dos itens são independentes para cada item

$$\pi(\mathbf{u}, \boldsymbol{\beta}) = \prod_{i=1}^n \pi_1(u_i) \prod_{j=1}^k \pi_2(\beta_j),$$

onde $\pi_2(\cdot) = \pi_{21}(a_j)\pi_{22}(b_j)$. Seguindo propostas usualmente consideradas, veja Patz e Junker (1999), consideramos $\pi_{22}(\cdot)$ como a densidade da $N(0, \sigma_b^2)$. Pelo fato de $a_j > 0$, muitos autores consideram $\pi_{21}(\cdot)$ como a densidade de uma normal truncada à esquerda, ou como a densidade da $LN(\mu_a, \sigma_a^2)$. Consideramos $\pi_{21}(\cdot)$ como a densidade da normal truncada à esquerda. Logo $\pi_2(\cdot)$ é a densidade de uma $N_2(\mu_\beta, \Sigma_\beta)$, onde $\mu_\beta = (\mu_a, 0)'$ e

$$\Sigma_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$$

Consideramos para $\pi_1(\cdot)$ inicialmente $\pi_1 \sim N(0, 1)$.

A função de verossimilhança para o modelo probito normal é dada por:

$$L(\mathbf{u}, \boldsymbol{\beta} | \mathbf{y}) = \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \quad (2.5)$$

em que Φ é a função distribuição acumulada da normal padrão e $m_{ij} = a_j u_i - b_j$.

2.4 MCMC usando amostrador de Gibbs ao considerar um esquema de dados aumentados (DAGS)

Utilizamos dados aumentados para evitar trabalhar com a verossimilhança de Bernoulli e obter um modelo mais fácil do ponto de vista computacional.

Nesta seção mostramos que com o uso de uma variável latente adicional, temos um modelo semelhante ao descrito na Seção 2.1.1. Este procedimento foi proposto por Albert (1992).

Proposição 2.1 *O modelo TRI probito normal, envolvendo k ítems e n indivíduos, com $y_{ij} \sim \text{Bernoulli}(p_{ij})$ e $p_{ij} = \Phi(m_{ij})$, na qual $m_{ij} = a_j u_j - b_j$, é equivalentemente definido considerando que*

$$Z_{ij} = m_{ij} + e_{ij}, \quad (2.6)$$

$$e_{ij} \sim N(0, 1), \quad (2.7)$$

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0. \end{cases} \quad (2.8)$$

Prova:

$$p_{ij} = P(Y_{ij} = 1 | u_i, \beta_j) = P(Z_{ij} > 0 | u_i, \beta_j),$$

onde $Z_{ij} = m_{ij} + \epsilon_{ij}$ e $\epsilon_{ij} \sim N(0, 1)$.

Assim, $Z_{ij} | u_i, \beta_j \sim N(m_{ij}, 1)$.

Com este resultado podemos escrever,

$$\begin{aligned} p_{ij} &= P(Z_{ij} > 0 | u_i, \beta_j) = P(Z_{ij} - m_{ij} > 0 - m_{ij}) \\ &= P(m_{ij} - Z_{ij} \leq m_{ij}) = \Phi(m_{ij}), \text{ pois } Z_{ij} | u_i, \beta_j \sim N(m_{ij}, 1) \\ &\Rightarrow p_{ij} = p(Z_{ij} > 0 | u_i, \beta_j) = \Phi(m_{ij}). \end{aligned}$$

■

Provando que temos um modelo equivalente ao modelo probito normal ao considerarmos uma estrutura linear para a variável latente auxiliar Z_{ij} .

Como já conhecemos a distribuição de Z_{ij} dada a habilidade e o vetor de parâmetros dos itens podemos obter a função de verossimilhança dos dados aumentados, considerando $D = (\mathbf{y}, \mathbf{z})$, como é descrito por Albert e Chib, (1993).

$$L(\mathbf{u}, \boldsymbol{\beta}|D) = \prod_{i=1}^n \prod_{j=1}^k \phi(Z_{ij}; m_{ij}, 1) [I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)], \quad (2.9)$$

em que $\phi(x; \mu, \sigma^2)$, denota a densidade de probabilidade da normal na variável x com média e variância dadas respectivamente por μ e σ^2 e $I(\cdot)$ denota uma função indicadora que toma o valor 0 ou 1 dependendo do valor de y . Para evitar abuso de notação consideraremos de agora em diante que $I_* = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$.

De acordo com a equação (2.9) e as prioris especificadas na Seção 2.3 para (u, β) , temos que a distribuição a posteriori conjunta de β , Z e u dado Y é dada por

$$f(Z, u, \beta|Y) \propto f(Z|u, \beta, y)f(u, \beta|y)$$

ou seja, $f(Z, u, \beta|Y)$

$$\propto \exp \left\{ \frac{-1}{2} \left[\sum_{i=1}^n \sum_{j=1}^k (Z_{ij} - m_{ij})^2 + \sum_{i=1}^n u_i^2 + \frac{1}{\sigma_b^2} \sum_{j=1}^k b_j^2 + \frac{1}{\sigma_a^2} \sum_{j=1}^k (a_j - \mu_a)^2 \right] \right\} I_*. \quad (2.10)$$

Ao observarmos a equação (2.10), notamos que esta é difícil de amostrarmos diretamente, mas para determinarmos a distribuição marginal a posteriori para β e u através do algoritmo Gibbs Sampling é necessário que conheçamos somente as distribuições condicionais completas, ou seja, precisamos conhecer

$$f(u|Z, \beta, Y), \quad f(\beta|u, Z, Y) \quad \text{e} \quad f(Z|u, \beta, Y)$$

Proposição 2.2 *De acordo com a função de verossimilhança para dados aumentados dada pela equação (2.9) e as distribuições a priori dadas na Seção 2.3 as distribuições condicionais completas para o amostrador de Gibbs para o modelo TRI probito normal são dadas por:*

- $f(Z_{ij}|U_i, \beta_j, \mathbf{y}) \propto N(m_{ij}, 1)I_*$
onde $I_* = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$

- $f(\beta_j|U_i, Z_j, \mathbf{y}) \propto N_2(m_{\beta_j}, v_{\beta_j})$
 com $m_{\beta_j} = [W'W + \Sigma_\beta^{-1}]^{-1}[W'Z_j + \Sigma_\beta^{-1}\mu_\beta]$, $v_{\beta_j} = [W'W + \Sigma_\beta^{-1}]^{-1}$
 onde $\mu_\beta = (\mu_a, 0)'$, $\Sigma_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$ e $W = (U, -1)$, $W'_i = (U_i, -1)$
- $f(U_i|\beta, Z, \mathbf{y}) \propto N(m_{U_i}, v_{U_i})$
 com $m_{U_i} = \frac{\sum_{j=1}^k (Z_{ij} + b_j)a_j}{\sum_{j=1}^k a_j^2 + 1}$, $v_{U_i} = \frac{1}{\sum_{j=1}^k a_j^2 + 1}$

Prova: Para obter estas condicionais basta olhar para o núcleo correspondente à cada variável na equação (2.10)

$$f(\mathbf{u}|\mathbf{Z}, \boldsymbol{\beta}, \mathbf{y}) \propto \exp \left\{ \frac{-1}{2} \sum_{i=1}^n \left[\left(1 + \sum_{j=1}^k a_j^2 \right) u_i^2 - 2u_i \sum_{j=1}^k a_j (b_j + Z_{ij}) \right] \right\}$$

Completando os quadrados obtemos:

$$f(\mathbf{u}|\mathbf{Z}, \boldsymbol{\beta}, \mathbf{y}) \propto \exp \left\{ \frac{-1 \left(1 + \sum_{j=1}^k a_j^2 \right)}{2} \sum_{i=1}^n \left[u_i - \frac{\sum_{j=1}^k a_j (b_j + Z_{ij})}{1 + \sum_{j=1}^k a_j^2} \right]^2 \right\}$$

$$\Rightarrow U_i \sim N \left(\frac{\sum_{j=1}^k a_j (b_j + Z_{ij})}{1 + \sum_{j=1}^k a_j^2}, \frac{1}{1 + \sum_{j=1}^k a_j^2} \right) \quad (2.11)$$

$$f(\boldsymbol{\beta}|\mathbf{u}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ \frac{-1}{2} \left[\sum_{i=1}^n \sum_{j=1}^k -2Z_{ij}(a_j u_i - b_j) + (a_j u_i - b_j)^2 + \frac{1}{\sigma_b^2} \sum_{j=1}^k b_j^2 + \frac{1}{\sigma_a^2} \sum_{j=1}^k (a_j - \mu_a)^2 \right] \right\} \quad (2.12)$$

que é correspondente ao núcleo de k distribuições normais multivariadas para β_j .

$$f(\mathbf{Z}|\mathbf{U}, \boldsymbol{\beta}, \mathbf{y}) \propto \exp \left[\frac{-1}{2} \left(\sum_{i=1}^n \sum_{j=1}^k (Z_{ij} - m_{ij})^2 \right) \right] I_* \quad (2.13)$$

onde $I_* = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$.

■

Agora é possível implementar o amostrador de Gibbs. Se formos utilizar o R, temos um “pacote” específico para isto, o que facilita bastante. Mas podemos também utilizar o Winbugs, desde que especifiquemos corretamente as priors para os parâmetros.

2.5 Estruturas Hierárquicas

O algoritmo que considera que variáveis latentes foram adicionadas ao modelo é chamado de DAGS, de acordo com este algoritmo, temos a seguinte forma hierárquica:

$$Z_{ij}|y_{ij}, u_i, a_j, b_j \sim N(a_j u_i - b_j, 1)[I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)]$$

$$u_i \sim N(0, 1)$$

$$a_j \sim N(\mu_a, \sigma_a^2)I(a_j > 0)$$

$$b_j \sim N(0, \sigma_b^2)$$

De acordo com esta forma hierárquica podemos implementar o algoritmo DAGS utilizando o programa Winbugs para o modelo Probit Normal. No Capítulo 5 apresentamos uma aplicação onde será utilizada esta estrutura.

Capítulo 3

Teoria de Resposta ao Item Normal Assimétrica (TRI-NA)

As funções de ligação geralmente usadas para os dados de respostas binárias como logito e probito, nem sempre proporcionam um bom ajuste para um conjunto de dados. Isto ocorre quando por exemplo, a probabilidade de uma dada resposta binária se aproxima de zero muito mais rápido do que esta se aproxima de um. Neste caso, o uso de uma função de ligação simétrica é inapropriada. Maiores detalhes são dados no trabalho de Chen *et al* (1999). Assim seria interessante trabalharmos com modelos mais amplos que possuem as ligações simétricas como caso especial.

O modelo desenvolvido no artigo de Chen *et al* (1999) é motivado pelo uso de variáveis latentes abordado por Albert e Chib (1993), onde a variável latente possui um modelo com estrutura de efeito aleatório. Este modelo inclui várias ligações assimétricas e simétricas geralmente usadas como caso especial, além disso, o algoritmo MCMC para o modelo pode ser facilmente implementado.

Apresentamos neste capítulo o modelo de Chen *et al* (1999), o modelo probito assimétrico generalizado e sua abordagem para dados aumentados e os modelos propostos na TRI-NA por Bazán (2005).

3.1 O modelo de Chen *et al*

Veremos nesta seção o modelo proposto por Chen *et al* (1999) para dados com resposta dicotômica. Este modelo é motivado pelo uso de variáveis latentes abordado em Albert e Chib (1993), onde estas variáveis latentes possuem uma distribuição normal assimétrica.

Sejam $Y = (y_1, \dots, y_n)'$ um vetor com n variáveis aleatórias independentes dicotômicas, considere também $x_i = (x_{i1}, \dots, x_{ik})'$ um vetor de covariáveis, onde x_{i1} pode ser 1 (corresponde ao intercepto), $\beta = (\beta_1, \dots, \beta_k)'$ o vetor dos coeficientes de regressão e X denota a matriz $n \times k$ de traços latentes. Considerando

$$y_i = \begin{cases} 1, & p_i; \\ 0, & 1 - p_i. \end{cases}$$

No tradicional modelo de respostas dicotômicas, assume-se que:

$$p_i = F(x_i' \beta) \quad (3.1)$$

onde $F(\cdot)$ representa a função distribuição acumulada e a F^{-1} é chamada de ligação de um modelo linear generalizado. O modelo definido por (3.1) é chamado de modelo de regressão binomial padronizado. Quando F é a função distribuição acumulada de uma distribuição simétrica a ligação resultante é simétrica. Mas podemos obter uma ligação assimétrica tomando F assimétrica.

Seja $z = (z_1, \dots, z_n)'$ um vetor de variáveis latentes independentes. Então o modelo de ligação assimétrico proposto é formulado como

$$y_i = \begin{cases} 0 & \text{se } z_i \leq 0, \\ 1 & \text{se } z_i > 0, \end{cases} \quad (3.2)$$

onde

$$z_i = x_i' \beta + \lambda w_i + \epsilon_i, \quad w_i \sim G, \quad \epsilon_i \sim F. \quad (3.3)$$

w_i e ϵ_i são variáveis aleatórias independentes, G é a função distribuição acumulada de uma distribuição assimétrica e F é a função distribuição acumulada de uma distribuição simétrica. Em (3.3), $-\infty \leq \lambda \leq \infty$ é um parâmetro de assimetria. Os modelos definidos pelas equações (3.1) e (3.3) possuem várias propriedades interessantes. São elas:

- Quando $\lambda = 0$, o modelo de ligação assimétrico se reduz ao modelo com ligação simétrico.
- Quando $F = \Phi$, f.d.a de uma normal padronizada e G é a f.d.a de uma normal truncada com função densidade

$$g(w) = \frac{2}{\sqrt{2\pi}} \exp - \left\{ \frac{w^2}{2} \right\}, \quad w > 0 \quad (3.4)$$

então,

a) A distribuição condicional de $z_i|w_i \sim N(x'_i\beta + \lambda w_i, 1)$.

Prova:

$$z_i = x'_i\beta + \lambda w_i.$$

Notemos que $z_i|w_i$ vai depender somente da distribuição de ϵ_i , (pois w_i será um valor conhecido), como ϵ_i tem uma função distribuição de probabilidade normal padronizada, usaremos o fato de que uma combinação linear de variáveis aleatórias normais possui distribuição normal, logo, z_i tem uma distribuição normal, com

$$E(z_i|w_i) = x'_i\beta + \lambda w_i$$

$$Var(z_i|w_i) = Var(x'_i\beta + \lambda w_i + \epsilon_i) = Var(\epsilon_i) = 1.$$

Logo,

$$z_i|w_i \sim N(x'_i\beta + \lambda w_i, 1).$$

■

b) A distribuição marginal de z_i é normal assimétrica padronizada com função distribuição de probabilidades dada por

$$\pi(z_i|\lambda) = \frac{2}{\sqrt{1+\lambda^2}} \phi\left(\frac{z_i - x'_i\beta}{\sqrt{1+\lambda^2}}\right) \Phi\left(\frac{\lambda}{\sqrt{1+\lambda^2}}(z_i - x'_i\beta)\right), \quad (3.5)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ representam a densidade de probabilidade e a função distribuição acumulada de uma normal padrão, respectivamente .

Existem diferentes formas de se provar este resultado, neste trabalho utilizamos a expressão de Henze (1986) e o Lema 3.1.

Lema 3.1 Se $S \sim NA(\lambda)$, então $S^* = \sqrt{1 + \lambda^2}S \sim NA(0, 1 + \lambda^2, \lambda)$, ou seja,

$$f(s^*) = \frac{2}{\sqrt{1 + \lambda^2}} \phi\left(\frac{s^*}{\sqrt{1 + \lambda^2}}\right) \Phi\left(\frac{\lambda}{\sqrt{1 + \lambda^2}}s^*\right)$$

Prova:

$$\begin{aligned} P(S^* \leq s) &= P(\sqrt{1 + \lambda^2}S \leq s) = P\left(S \leq \frac{s}{\sqrt{1 + \lambda^2}}\right) \\ f_{s^*}(s) &= \frac{d}{ds} \left[F_s\left(\frac{s}{\sqrt{1 + \lambda^2}}\right) \right] \\ &= f_s\left(\frac{s}{\sqrt{1 + \lambda^2}}\right) \frac{1}{\sqrt{1 + \lambda^2}} \end{aligned}$$

em que $f_s = 2\phi(s)\Phi(\lambda s)$.

$$\text{Logo, } f_{s^*}(s) = \frac{2}{\sqrt{1 + \lambda^2}} \phi\left(\frac{s}{\sqrt{1 + \lambda^2}}\right) \Phi\left(\frac{\lambda}{\sqrt{1 + \lambda^2}}s\right).$$

■

Tendo em mãos o Lema 3.1, realizamos em seguida a Prova do item b).

Prova do item b):

$$z_i = x'_i\beta + \lambda w_i + \epsilon_i.$$

Considere $K_i = z_i - x'_i\beta = \lambda w_i + \epsilon_i$.

Devemos determinar qual a distribuição de K_i .

Pela Proposição A.3 (Henze, 1986) temos que:

$$\begin{aligned} K_i^* &= \frac{\lambda}{\sqrt{1 + \lambda^2}}w_i + \frac{1}{\sqrt{1 + \lambda^2}}\epsilon_i \Rightarrow K_i^* \sim NA(\lambda) \\ K_i &= \lambda w_i + \epsilon_i \Rightarrow K_i = \sqrt{1 + \lambda^2} \frac{\lambda w_i + \epsilon_i}{\sqrt{1 + \lambda^2}}. \end{aligned}$$

Logo,

$$K_i = \sqrt{1 + \lambda^2}K_i^*.$$

Usando o lema anterior temos que $K_i \sim NA(0, 1 + \lambda^2, \lambda)$, ou seja,

$$f_{K_i}(k_i) = \frac{2}{\sqrt{1 + \lambda^2}} \phi\left(\frac{k_i}{\sqrt{1 + \lambda^2}}\right) \Phi\left(\frac{\lambda}{\sqrt{1 + \lambda^2}}k_i\right)$$

onde $K_i = z_i - x'_i\beta$. E portanto $z_i|\lambda \sim NA(x'_i\beta, 1 + \lambda^2, \lambda)$.

■

Dois casos especiais dos modelos definidos pelas equações (3.1) e (3.3) são: o modelo probito obtido quando é considerado $F(t) = \Phi(t)$ e o modelo logito que é obtido quando consideramos $F(t) = \frac{e^t}{1 + e^t}$.

De acordo com a Proposição 3.1 podemos obter as probabilidades para $y_i = 0$ e $y_i = 1$.

Proposição 3.1 *Considere a seguinte representação para y_i*

$$y_i = \begin{cases} 0, & \text{se } z_i \leq 0, \\ 1, & \text{se } z_i > 0, \end{cases}$$

onde

$$z_i = x'_i \beta + \lambda w_i + \epsilon_i, \quad w_i \sim G, \quad \epsilon_i \sim F,$$

w_i e ϵ_i são variáveis aleatórias independentes, G é a função distribuição acumulada de uma distribuição assimétrica e F é a função distribuição acumulada de uma distribuição simétrica. A probabilidade do indivíduo i acertar a questão é dada por:

$$p_i = P(y_i = 1) = \int_{-\infty}^{\infty} F(x'_i \beta + \lambda w_i) g(w_i) dw_i \quad (3.6)$$

e

$$1 - p_i = P(y_i = 0) = \int_{-\infty}^{\infty} [1 - F(x'_i \beta + \lambda w_i)] g(w_i) dw_i. \quad (3.7)$$

Prova:

Para provar o resultado dado pela equação (3.6), basta notar que

1. $E(Y) = 1.P(y_i = 1) + 0.P(y_i = 0) = P(z_i > 0)$;
2. $z_i|w_i$ possui uma distribuição simétrica;
3. $E(Y) = E\{E(Y|W)\}$ (propriedade da esperança condicional).

$$\begin{aligned} E(Y|W) &= P(z_i > 0|w_i) \\ &= P(z_i - (x'_i \beta + \lambda w_i) > -(x'_i \beta + \lambda w_i)) = P(\epsilon_i > -(x'_i \beta + \lambda w_i)) \\ &= P(-\epsilon_i \leq (x'_i \beta + \lambda w_i)) \\ &= F(x'_i \beta + \lambda w_i), \text{ pois } \epsilon_i \text{ possui uma distribuição simétrica.} \end{aligned}$$

Assim,

$$\begin{aligned} E\{E(Y|W)\} &= P(y_i = 1) = \int_{-\infty}^{\infty} F(x'_i\beta + \lambda w_i)g(w_i)dw \quad e \\ P(y_i = 0) &= 1 - P(y_i = 1) = 1 - \int_{-\infty}^{\infty} F(x'_i\beta + \lambda w_i)g(w_i)dw \\ &= \int_{-\infty}^{\infty} [1 - F(x'_i\beta + \lambda w_i)]g(w_i)dw_i, \end{aligned}$$

onde $g(w_i)$ é a densidade de w_i .

■

3.2 Modelo Probito assimétrico generalizado

Chen *et al* (1999), apresentam uma versão de verossimilhança aumentada considerada a partir da equação (3.6).

$$L(\beta, \lambda|X, y, w) = \prod_{i=1}^n \int_{-\infty}^{\infty} [F(x'_i\beta + \lambda w_i)]^{y_i} [1 - F(x'_i\beta + \lambda w_i)]^{1-y_i} g(w_i). \quad (3.8)$$

Porém se lembrarmos que $z_i \sim NA(\theta)$, em que $\theta = (x'_i\beta, 1 + \lambda^2, \lambda)$, podemos obter uma expressão analítica para p_i . Assim, teremos uma ligação probito assimétrica como a que é dada no Lema 3.1.

Podemos também considerar a função distribuição acumulada da normal assimétrica padrão dada por Azzalini (1985), para a ligação probito assimétrica. Neste caso, p_i seria dado pela expressão abaixo

$$p_i = F_\lambda(x'_i\beta) = 2\Phi_2 \left(\left(\begin{array}{c} x'_i\beta \\ 0 \end{array} \right); \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & -\delta_j \\ -\delta_j & 1 \end{array} \right) \right),$$

$$\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}.$$

Em que $F_\lambda(\cdot)$ é a função distribuição acumulada da normal assimétrica, e como pode ser visto no Apêndice A esta é representada pela f.d.a da normal bivariada, com coeficiente de correlação δ_j , $|\delta_j| \leq 1$ e vetor de médias nulo. (Demonstração feita em A.2).

Neste caso não necessitamos utilizar a verossimilhança para dados aumentados pois a função de verossimilhança segue a verossimilhança usual no modelo de Bernoulli.

Denotamos a ligação que possui a f.d.a da normal assimétrica padrão por $L1$ e a ligação que possui a f.d.a da distribuição normal assimétrica como a que é dada pela pelo Lema 3.1 por $L2$.

A formulação dos modelos que seguem as ligações $L1$ e $L2$ possuem algumas diferenças. Na ligação $L1$ não é necessário considerar uma estrutura linear latente, basta considerar uma distribuição assimétrica para F . Embora essas formulações sejam diferentes elas são casos particulares de um modelo mais geral.

A ligação probito assimétrica generalizada para a modelagem de resposta binária considerada por Bazán (2005), tem como casos particulares as ligações $L1$, $L2$ e a ligação probito.

3.2.1 O modelo

O modelo probito assimétrico generalizado é obtido considerando que a distribuição da F em (3.1) é dada pela distribuição normal assimétrica com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$, em que μ é o parâmetro de posição, σ^2 é o parâmetro de escala e λ é o parâmetro de assimetria, ou seja,

$$p_i = F_\theta(x'_i\beta), \quad i = 1, \dots, n \quad (3.9)$$

Modificando o vetor de parâmetros da distribuição normal assimétrica obtemos alguns modelos como casos particulares, por exemplo:

- Se $\theta = (0, 1, 0)$, temos a ligação probito;
- Se $\theta = (0, 1 + \lambda^2, -\lambda)$, temos a ligação $L2$
- Se $\theta = (0, 1, \lambda)$, temos a ligação $L1$ (padrão de Azzalini).

3.3 Verossimilhança

A função de verossimilhança para modelos probitos assimétricos generalizados é dada por

$$L(\beta, \theta|y, X) = \prod_{i=1}^n [F_\theta(x'_i\beta)]^{y_i} [1 - F_\theta(x'_i\beta)]^{1-y_i}. \quad (3.10)$$

Como foi dito na Seção 3.2, $F_{\theta}(x'_i\beta)$, segue uma distribuição normal bivariada e para se calcular esta distribuição existem alguns algoritmos computacionais disponíveis. Assim conseguimos através destes, obter a função de verossimilhança para os dados.

3.4 Verossimilhança aumentada para o modelo probito assimétrico generalizado

Nest seção apresentamos uma versão de verossimilhança aumentada para os dados semelhante aquela dada no modelo probito normal por Albert e Chib (1993).

Consideramos as variáveis latentes auxiliares de modo que:

$$y_i = \begin{cases} 1, & Z_i > 0, \\ 0, & Z_i \leq 0, \end{cases} \quad (3.11)$$

onde $Z_i \sim NA(\theta)$, com $\theta = (\mu + x'_i\beta, \sigma^2, -\lambda)$. Notemos que para o caso em que $\theta = (0, 1, 0)$, temos o modelo dado por Albert e Chib (1993). Optamos por inserir a variável latente Z_i para evitar o uso da verossimilhança do tipo Bernoulli, e esta estrutura linear latente, produz um modelo equivalente ao probito assimétrico generalizado. Então a função de verossimilhança para dados aumentados para o modelo probito assimétrico generalizado é dada por

$$L(\beta, \theta | Z, y) = \prod_{i=1}^n f_{\theta}(z_i) [I(Z_i > 0)I(y_i = 1) + I(Z_i \leq 0)I(y_i = 0)]. \quad (3.12)$$

Pode-se obter uma verossimilhança alternativa para o modelo probito assimétrico generalizado. Para isto será considerado a seguinte representação linear latente para Z_i .

$$Z_i = x'_i\beta + e_i, \quad e_i \sim NA(\theta^*), \quad i = 1, \dots, n \quad (3.13)$$

onde $\theta^* = (\mu, \sigma^2, -\lambda)$, os e_i são denotados de acordo com a representação estocástica de Henze (1986) dada na Proposição A.3.

$$e_i = \mu - \sigma[\delta V_i + \sqrt{1 - \delta^2} W_i], \quad i = 1, \dots, n \quad (3.14)$$

em que $V_i \sim N(0, 1)I(v_i > 0)$, e $W_i \sim N(0, 1)$.

Observe que

$$E(e_i|V_i) = \mu - \sigma\delta V_i, \quad \text{Var}(e_i|V_i) = (1 - \delta^2)\sigma^2,$$

$$e_i|V_i \sim N(\mu - \sigma\delta V_i, (1 - \delta^2)\sigma^2).$$

De acordo com a representação dos erros dada na equação (3.14), temos que $Z_i = x'_i\beta + \mu - \sigma[\delta V_i + \sqrt{1 - \delta^2}W_i]$, assim, considerando $Z_i^* = Z_i|V_i$, podemos dizer que $Z_i^* \sim N(x'_i\beta + \mu - \sigma\delta v_i, (1 - \delta^2)\sigma^2)$.

A função de verossimilhança aumentada alternativa para o modelo probito assimétrico generalizado é dada por

$$L(\beta, \theta|Z, y) = \prod_{i=1}^n \phi(z_i^*; x'_i\beta + \mu - \sigma\delta v_i, (1 - \delta^2)\sigma^2) \phi(v_i) I(v_i > 0) I_{**}, \quad (3.15)$$

onde $I_{**} = I(Z_i^* > 0)I(y_i = 1) + I(Z_i^* \leq 0)I(y_i = 0)$.

3.5 Análise Bayesiana

3.5.1 Especificação de prioris

Precisamos especificar uma distribuição a priori para β e para θ , consideramos estes parâmetros independentes a priori:

$$\pi(\beta, \theta) = \pi_1(\beta)\pi_2(\theta). \quad (3.16)$$

Podemos considerar para $\pi_1(\beta)$ as prioris consideradas para o modelo probito normal (ver Seção 2.3), além de prioris impróprias ($\pi_1(\beta) = 1$). Para θ , temos:

$$\pi_2(\theta) = \pi_{21}(\mu)\pi_{22}(\sigma^2)\pi_{23}(\lambda). \quad (3.17)$$

Assim como Bazán (2005), consideramos para $\pi_{21}(\mu)$ uma distribuição normal, isto é, $\mu \sim N(\mu_0, \sigma_0^2)$ e para $\pi_{22}(\sigma^2)$ uma distribuição qui-quadrado inversa, ou seja, $\sigma^{-2} \sim \chi^2(w, k)$.

Para o parâmetro de assimetria λ , podemos considerar a parametrização $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. De acordo com esta parametrização $\delta \in [-1, 1]$ o que nos sugere utilizar uma distribuição a

priori uniforme para δ , ou seja, $\delta \sim U[-1, 1]$, a qual nos induz ao espaço paramétrico de λ uma distribuição t-Student com os parâmetros de posição, escala e graus de liberdade dados respectivamente por 0, 0.5 e 2 ($\lambda \sim t(0, 0.5, 2)$). Para maiores detalhes ver Rodríguez (2005).

3.6 Esquema de dados aumentados para o amostrador de Gibbs

Usando a verossimilhança dada pela equação (3.10) e a distribuição a priori dada em (3.16) pode-se implementar o procedimento de estimação bayesiana utilizando a verossimilhança de Bernoulli, porém, a fim de facilitar a implementação via MCMC, usamos a verossimilhança aumentada alternativa especificada na Seção 3.4, que considera o uso da forma de Henze (1986) para representar a distribuição normal assimétrica.

$$Z_i^* | y_i, \boldsymbol{\beta}, \mu, \sigma^2, \delta \sim N(x_i' \boldsymbol{\beta} + \mu - \sigma \delta v_i, (1 - \delta^2) \sigma^2) I_{**},$$

$$V_i \sim N(0, 1) I(v_i > 0),$$

$$\boldsymbol{\beta} \sim \pi_1(\cdot),$$

$$\mu \sim \pi_{21}(\cdot),$$

$$\sigma^2 \sim \pi_{22}(\cdot),$$

$$\delta \sim U[-1, 1].$$

Se ao invés da parametrização δ , considerarmos a parametrização λ , teremos

$$Z_i^* | y_i, \boldsymbol{\beta}, \mu, \sigma^2, \lambda \sim N\left(\mu + x_i' \boldsymbol{\beta} - \frac{\lambda}{\sqrt{1 + \lambda^2}} \sigma v_i, \frac{1}{\sqrt{1 + \lambda^2}} \sigma^2\right) I_{**},$$

$$V_i \sim N(0, 1) I(v_i > 0),$$

$$\boldsymbol{\beta} \sim \pi_1(\cdot),$$

$$\mu \sim \pi_{21}(\cdot),$$

$$\sigma^2 \sim \pi_{22}(\cdot),$$

$$\lambda \sim t(0, 0.5, 2),$$

onde $I_{**} = I(Z_i^* > 0)I(y_i = 1) + I(Z_i^* \leq 0)I(y_i = 0)$, β é especificado de acordo com a sugestão dada na Seção 2.3, e $t(\mu, \sigma^2, g)$, denota a densidade da distribuição t-Student com parâmetros de locação, escala e graus de liberdade dados respectivamente por μ , σ^2 , g .

De acordo com estas representações hierárquicas podemos implementar o modelo probito assimétrico generalizado usando o software Winbugs.

3.7 A família TRI-NA

Como caso particular do modelo probito assimétrico generalizado temos a família de modelos TRI-NA (Teoria de Resposta ao Item Normal Assimétrica) dada por Bazán (2005). Neste caso para não se ter um modelo super parametrizado, assume-se que $\mu = 0$ e $\sigma^2 = 1$, ou seja, $F_\lambda(\cdot)$ é a ligação $L1$.

O algoritmo de dados aumentados é semelhante ao dado na Seção 3.6, e é descrito na Subseção 3.7.5. Na Subseção 3.7.6, daremos a especificação de prioris para este caso. Em seguida mostramos os modelos possíveis dentro da TRI-NA.

3.7.1 Modelos

A família TRI-NA é formada por modelos probitos que envolvem alguma assimetria, esta assimetria pode estar presente na função de ligação e/ou nas habilidades dos indivíduos.

De acordo com as características da Curva Característica do Item (CCI) e da função de distribuição do traço latente (habilidade) podemos obter diferentes modelos na TRI-NA.

A seguir detalhamos as principais características de três modelos dentro da TRI-NA, propostos por Bazán (2005).

3.7.2 Probito Assimétrico-Normal

Neste caso a probabilidade condicional p_{ij} de uma resposta correta para o item j dado o valor da habilidade u_i correspondente ao i -ésimo indivíduo é:

$$p_{ij} = F_{\theta}(m_{ij}) \quad m_{ij} = a_j u_i - b_j, \quad i = 1, \dots, n \quad j = 1, \dots, k$$

onde $F_{\theta}(\cdot)$ é a função distribuição acumulada de uma distribuição normal assimétrica de parâmetro $\theta = (0, 1, \lambda)$.

É conveniente deixar que o parâmetro de assimetria λ seja diferente para cada item, pois possibilita obter itens com CCI simétricos e assimétricos. Assim podemos escrever

$$p_{ij} = F_{\lambda_j}(m_{ij}) \quad i = 1, \dots, n \quad j = 1, \dots, k \quad (3.18)$$

De acordo com as propriedades apresentadas no Apêndice A para a distribuição normal assimétrica, temos:

$$F_{\lambda_j}(m_{ij}) = 2\Phi_2 \left(\left(\begin{array}{c} m_{ij} \\ 0 \end{array} \right); \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & -\delta_j \\ -\delta_j & 1 \end{array} \right) \right), \quad \delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}.$$

Em que $F_{\lambda_j}(\cdot)$ é representada pela função distribuição acumulada da normal bivariada, $|\delta_j| \leq 1$ é o coeficiente de correlação da normal bivariada com vetor de médias nulo.

Note que para $\lambda_j = 0$, temos $F(m_{ij}) = \Phi(m_{ij})$ que representa o modelo probito normal.

Na Figura 3.1, mostramos a diferença que existe na CCI quando consideramos que a assimetria está presente, representamos nesta figura cinco curvas, os parâmetros dos itens a e b possuem os valores $a = 1$ e $b = 0$. A curva do meio indica um modelo sem assimetria, ou seja, neste caso teremos $\lambda = 0$.

Neste modelo consideramos que a variável latente $U_i \sim N(\mu, \sigma^2)$, ou seja, as variáveis latentes associadas com os indivíduos que fazem o teste são “bem comportadas”. No entanto, como foi dito no Capítulo 1, ao se tratar de dados relacionados à conduta humana é pouco realista supor que estas habilidades se comportam de maneira similar (sem variabilidade).

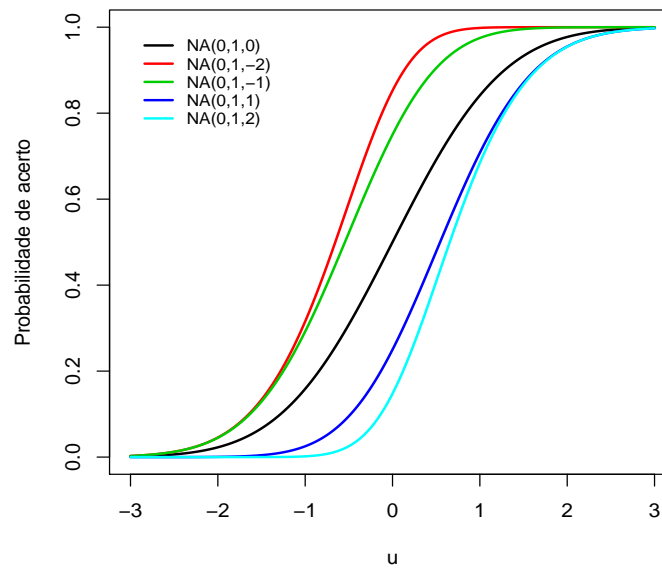


FIGURA 3.1: CCI probito assimétricos para $a=1$, $b=0$, e diferentes valores do parâmetro de assimetria $\lambda = -2, -1, 0, 1, 2$.

Interpretação dos parâmetros na TRI- NA

Ao olharmos para a Figura 3.1, notamos o comportamento de λ para diferentes valores. Quando λ assume valores negativos, ao considerarmos alunos que possuem baixa habilidade um pequeno acréscimo na habilidade aumenta consideravelmente a probabilidade de acerto. Já quando λ assume valores positivos notamos que um pequeno acréscimo na habilidade de alunos que possuem alta habilidade ocasiona um grande aumento na probabilidade de acerto, neste caso λ está bonificando alunos com alta habilidade enquanto que λ negativo bonifica alunos com baixa habilidade.

Não comentamos a interpretação dos parâmetros a_j e b_j na TRI-NA, devido ao fato desta ser totalmente semelhante à interpretação dos parâmetros no modelo PN visto na Subseção 2.1.2.

3.7.3 Probito Normal-Assimétrico

Neste modelo estaremos levando em consideração que geralmente as variáveis latentes não são bem comportadas, assim a diferença deste modelo para o modelo citado anteriormente na Subseção 3.7.2 é que neste caso, teremos $p_{ij} = F(m_{ij}) = \Phi(m_{ij})$, ou seja, neste modelo a probabilidade de um indivíduo i , $i = 1, \dots, n$ responder corretamente um item j , $j = 1, \dots, k$, dado o valor da variável latente u_i , corresponde à f.d.a de uma distribuição normal.

Além disso, a variável latente u_i tem uma distribuição normal assimétrica

$$U_i \sim NA(\mu_0, \sigma_0^2, \kappa).$$

Neste caso, o parâmetro de assimetria não muda de um indivíduo para o outro, pois, segundo Bazán (2005), esta é uma propriedade da distribuição da variável latente para a população completa da qual o indivíduo é extraído.

3.7.4 Probito Assimétrico Normal Assimétrico

Este modelo é mais geral e inclui todos os outros como casos particulares. Neste modelo, além de termos $F_\theta(m_{ij})$ onde $F_\theta(\cdot)$ é a função distribuição acumulada de uma distribuição normal assimétrica de parâmetro $\theta = (0, 1, \lambda)$ consideramos que as habilidades u_i , possuem uma distribuição normal assimétrica, ou seja

$$p_{ij} = F_{\lambda_j} = 2\Phi_2 \left(\left(\begin{pmatrix} m_{ij} \\ 0 \end{pmatrix} ; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta_j \\ -\delta_j & 1 \end{pmatrix} \right) \right) \quad \delta_j = \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}}. \quad (3.19)$$

$$U_i \sim NA(\mu_0, \sigma_0^2, \kappa). \quad (3.20)$$

Com este modelo encerramos a apresentação da família TRI-NA apresentada por Bazán (2005). Em seguida, mostramos a teoria necessária para implementar estes modelos.

3.7.5 Algoritmo de dados aumentados na TRI-NA

O algoritmo mostrado a seguir é similar ao apresentado neste trabalho para o modelo probito normal e é um caso particular do algoritmo dado na Seção 3.4.

Proposição 3.2 *O modelo TRI probito assimétrico, envolvendo k itens e n indivíduos, com $y_{ij} \sim \text{Bernoulli}(p_{ij})$ e $p_{ij} = F_\lambda(m_{ij})$ na qual $m_{ij} = a_j u_i - b_j$, é equivalentemente definido considerando que*

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0, \\ 0, & Z_{ij} \leq 0, \end{cases} \quad (3.21)$$

em que $Z_{ij} \sim NA(\theta)$, com $\theta = (m_{ij}, 1, -\lambda_j)$ $i = 1, \dots, n$ e $j = 1, \dots, k$.

Prova: Devemos mostrar que ao considerar esta estrutura de dados aumentados, temos um modelo similar ao proposto para TRI probito assimétrico.

$$p_{ij} = P(Y_{ij} = 1|u, \beta, \lambda) = P(Z_{ij} > 0|u, \beta, \lambda) = 1 - F_\lambda(z) \stackrel{*}{=} F_\lambda(-z)$$

* Pela propriedade D dada em A.1.

Assim, $p_{ij} = F_\lambda(-z)$.

■

Introduzimos a variável latente Z_{ij} para evitar de se trabalhar com a verossimilhança do tipo Bernoulli. De acordo com a proposição apresentada acima podemos escrever a função de verossimilhança aumentada para o modelo TRI probito assimétrico como

$$L(U, a, b, \lambda|D) = \prod_{i=1}^n \prod_{j=1}^k f_\theta(Z_{ij}) [I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)], \quad (3.22)$$

onde $\theta = (m_{ij}, 1, -\lambda_j)$, $D = (Z, y)$.

Podemos escrever a função de verossimilhança aumentada de uma maneira alternativa, como especificado na Seção 3.6, que facilita a implementação via Winbugs.

Proposição 3.3 *O modelo TRI probito normal assimétrico, com k itens e n indivíduos, com $y_{ij} \sim \text{Bernoulli}(p_{ij})$ e $p_{ij} = F_\lambda(m_{ij})$ na qual $m_{ij} = a_j u_i - b_j$, é equivalentemente definido considerando que*

$$y_{ij} = \begin{cases} 1, & Z_{ij}^* > 0, \\ 0, & Z_{ij}^* \leq 0, \end{cases} \quad (3.23)$$

em que $Z_{ij}^* \sim N(-\delta_j V_{ij} + m_{ij}, 1 - \delta_j^2)$, e V_{ij} possui uma distribuição normal truncada à esquerda, isto é, $V_{ij} \sim N(0, 1)I(v_i > 0)$, $j = 1, \dots, k$ e $i = 1, \dots, n$.

Prova: A probabilidade do indivíduo i acertar a questão j dada a variável V_{ij} , o vetor de parâmetros dos itens e a habilidade é dada por

$$\begin{aligned} P(y_{ij} = 1|u, \beta, \lambda, v_{ij}) &= P(Z_{ij}^* > 0|u, \beta, \lambda, v_{ij}) \\ &= P\left(\frac{Z_{ij}^* - (m_{ij} - \delta_j v_{ij})}{\sqrt{1 - \delta_j^2}} > -\frac{(m_{ij} - \delta_j v_{ij})}{\sqrt{1 - \delta_j^2}}\right), \end{aligned}$$

considere X_{ij} como a variável Z_{ij}^* padronizada, assim,

$$\begin{aligned} &= P\left(-X_{ij} \leq \left(\frac{m_{ij} - \delta_j v_{ij}}{\sqrt{1 - \delta_j^2}}\right)\right) = \Phi\left(\frac{m_{ij} - \delta_j v_{ij}}{\sqrt{1 - \delta_j^2}}\right) \\ \Rightarrow p_{ij}|v_{ij} &= \Phi\left(\frac{m_{ij} - \delta_j v_{ij}}{\sqrt{1 - \delta_j^2}}\right), \text{ logo} \\ p_{ij} &= \int_0^\infty \Phi\left(\frac{m_{ij} - \delta_j v_{ij}}{\sqrt{1 - \delta_j^2}}\right) f(v_{ij}) dv_{ij}. \end{aligned}$$

De acordo com a Proposição A.4, podemos dizer que:

$$p_{ij} = F_\lambda(m_{ij}).$$

■

3.7.6 Especificação de prioris

Precisamos especificar uma distribuição a priori para β , u e λ . Consideramos estes parâmetros independentes a priori:

$$\pi(\beta, u, \lambda) = \pi_1(\beta)\pi_2(u)\pi_3(\lambda). \quad (3.24)$$

De acordo com as propostas usualmente consideradas na literatura, assumimos independência a priori entre os itens e a habilidade dos indivíduos.

$$\pi(\beta, u, \lambda) = \prod_{j=1}^k g_{1j}(\beta_j)g_{2j}(\lambda_j) \prod_{i=1}^n g_{3i}(u_i), \quad (3.25)$$

em que $g_{1j}(\beta_j) = g_{11j}(a_j)g_{12j}(b_j)$.

Como foi proposto no Capítulo 1, consideramos que $g_{11j}(\cdot)$ possui a densidade da normal positiva, ou seja, $a_j \sim N(\mu_a, \sigma_a^2)I(a_j > 0)$, $j = 1, \dots, k$ e $g_{12j}(\cdot)$ possui a densidade da $N(0, \sigma_b^2)$, $j = 1, \dots, k$ e portanto $g_{1j}(\cdot)$ é a densidade da $N_2(\mu_\beta, \Sigma_\beta)$, com vetor de médias $\mu_\beta = (\mu_a, 0)$ e matriz de covariâncias $\Sigma_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$.

Para $g_{3i}(u_i)$ consideramos a densidade da $NA(\theta)$, $i = 1, \dots, n$ com $\theta = (0, 1, k)$ e $g_{2j}(\cdot)$ com densidade $NA(w)$, com w considerado fixo.

3.7.7 Distribuições condicionais completas

De acordo com as prioris especificadas na Subseção 3.7.6 e com as versões de verossimilhança aumentadas da Subseção 3.7.5, podemos obter as distribuições condicionais completas necessárias para implementar os modelos probito assimétricos.

Proposição 3.4 *De acordo com a função de verossimilhança aumentada dada pela equação (3.12) e com as distribuições a priori dadas na Subseção 3.7.6. Na família TRI-NA, as distribuições condicionais completas para as variáveis dos modelos probito assimétricos são dadas por:*

1. $\pi(Z_{ij}|U_i, \beta_j, \lambda_j, y) \propto \phi_{NA}(Z_{ij}; m_{ij}, 1, -\lambda_j)I$,
em que $I = [I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)]$
2. $\pi(U_i|Z_{ij}, \beta_j, \lambda_j, y) \propto \phi_{NA}(U_i, m_{U_i}, v_{U_i})\psi(U_i)$
onde $\psi(U_i) = \prod_{j=1}^k \Phi(-\lambda_j Z_{ij} + b_j - U_i a_j)\Phi(kU_i)$ com
 $m_{U_i} = \frac{\sum_{j=1}^k a_j(Z_{ij} + b_j)}{1 + \sum_{j=1}^k a_j^2}$, $v_{U_i} = \frac{1}{1 + \sum_{j=1}^k a_j^2}$
3. $\pi(\beta_j|U, Z_j, \lambda_j, y) \propto \phi_2(\beta_j; m_{\beta_j}, v_{\beta_j})\psi(\beta_j)$
onde $\psi(\beta_j) = \prod_{i=1}^n \Phi(\lambda_j Z_{ij} - \mathbf{W}'_i \beta_j)$
com $m_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1}[\mathbf{W}'\mathbf{Z}_j + \Sigma_\beta^{-1}\mu_\beta]$, $v_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1}$
em que $\mu_\beta = \begin{pmatrix} \mu_a \\ 0 \end{pmatrix}$, $\Sigma_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$ e $\mathbf{W} = (u, -1)$ com $\mathbf{W}'_i = (\mathbf{U}_i, -1)$
4. $\phi(\lambda_j|U, Z_j, \beta_j, y) \propto \phi_2(\lambda_j)\psi(\lambda_j)$ em que
 $\psi(\lambda_j) = \prod_{i=1}^n \Phi(-\lambda_j(Z_{ij} - m_{ij}))\Phi(w\lambda_j)$

Prova: De acordo com a equação (3.22) e as prioris especificadas na Subseção 3.7.6, temos que a distribuição conjunta a posteriori é dada por

$$\begin{aligned}
f(U, \beta, y, \lambda, Z) &= f(Z|U, \beta, \lambda, y) f(U|\beta, \lambda, y) f(\beta|\lambda, y) f(\lambda|y) \\
&= \prod_{i=1}^n \prod_{j=1}^k \phi_{NA}(Z_{ij}; m_{ij}, 1, \lambda_j) I \times \\
&\quad \prod_{i=1}^n \phi_{NA}(U_i, k) \prod_{j=1}^k \phi_2(\beta_j, \mu_\beta, \Sigma_\beta) \phi_{NA}(\lambda_j, w) \quad (3.26)
\end{aligned}$$

onde $I = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$

Para provarmos os itens basta olhar a distribuição conjunta dos dados considerando apenas o núcleo, o que nos fornece para o item 1

$$\pi(Z_{ij}|U_i, \beta_j, \lambda_j, y) \propto \phi_{NA}(Z_{ij}; m_{ij}, 1, -\lambda_j) I$$

No item 2 teremos

$$\pi(U_i|Z_{ij}, \beta_j, \lambda_j, y) \propto \prod_{i=1}^n \prod_{j=1}^k \phi_{NA}(Z_{ij}; m_{ij}, 1, \lambda_j) I \prod_{i=1}^n \phi_{NA}(U_i, k)$$

Substituindo ϕ_{NA} por $2\phi(\cdot)\Phi(\cdot)$, temos

$$\pi(U_i|Z_{ij}, \beta_j, \lambda_j, y) \propto 2\phi(Z_{ij} - m_{ij})\Phi(-\lambda(Z_{ij} - m_{ij}))2\phi(U_i)\Phi(kU_i)$$

em que $\phi(Z_{ij} - m_{ij})\phi(U_i)$ pode ser escrito como

$$\phi(Z_{ij} - m_{ij})\phi(U_i) \propto \prod_{i=1}^n \prod_{j=1}^k \frac{1}{2\pi} \exp \left\{ \frac{-1}{2} ((Z_{ij} - m_{ij})^2 + (U_i)^2) \right\}$$

Substituindo $m_{ij} = a_j u_i - b_j$ e desenvolvendo os quadrados, obtemos

$$\begin{aligned}
&\prod_{i=1}^n \prod_{j=1}^k \exp \left\{ \frac{-1}{2} (Z_{ij} - 2Z_{ij}(a_j u_i + b_j) + a_j^2 u_i^2 - 2a_j u_i + b_j^2 + u_i^2) \right\} \\
&\propto \exp \left\{ \frac{-1}{2} \sum_{i=1}^n \left[\left(U_i^2 (1 + \sum_{j=1}^k a_j^2) - 2u_i \sum_{j=1}^k a_j (b_j + Z_{ij}) \right) \right] \right\} \\
&= \exp \left\{ \frac{-1(1 + \sum_{j=1}^k a_j^2)}{2} \sum_{i=1}^n \left[\left(U_i^2 - 2u_i \frac{\sum_{j=1}^k a_j (b_j + Z_{ij})}{(1 + \sum_{j=1}^k a_j^2)} \right) \right] \right\}
\end{aligned}$$

Completando os quadrados, temos

$$\propto \exp \left\{ \frac{-1}{2 \frac{1}{(1 + \sum_{j=1}^k a_j^2)}} \sum_{i=1}^n \left[\left(U_i - \frac{\sum_{j=1}^k a_j (b_j + Z_{ij})}{(1 + \sum_{j=1}^k a_j^2)} \right)^2 \right] \right\}$$

Assim $\phi(Z_{ij} - m_{ij})\phi(U_i) \propto \phi(U_i, m_{U_i}, v_{U_i})$ em que

$$m_{U_i} = \frac{\sum_{j=1}^k a_j (b_j + Z_{ij})}{(1 + \sum_{j=1}^k a_j^2)} \quad v_{U_i} = \frac{1}{(1 + \sum_{j=1}^k a_j^2)}$$

Logo,

$$\begin{aligned}\pi(U_i | Z_{ij}, \beta_j, \lambda_j, y) &\propto \Phi(-\lambda(Z_{ij} - m_{ij}))\Phi(kU_i)\phi(U_i, m_{U_i}, v_{U_i}) \\ &\propto \phi(U_i, m_{U_i}, v_{U_i})\psi(U_i)\end{aligned}$$

O item 3 segue de resultados de Modelos de Regressão, Neter *et al* (1983). Para o item 4, temos:

$$\begin{aligned}\phi(\lambda_j | U, Z_j, \beta_j, y) &\propto \prod_{i=1}^n \prod_{j=1}^k \phi_{NA}(Z_{ij}, m_{ij}, 1 - \delta_j) \prod_{j=1}^k \phi_{NA}(\lambda_j, w) \\ &\propto \prod_{i=1}^n \prod_{j=1}^k \phi(Z_{ij}, m_{ij})\Phi(-\lambda_j(Z_{ij} - m_{ij}))\phi(\lambda_j)\Phi(w\lambda_j) \\ &\propto \phi(\lambda_j) \prod_{i=1}^n \Phi(-\lambda_j(Z_{ij} - m_{ij}))\Phi(w\lambda_j) \\ &\propto \phi(\lambda_j)\psi(\lambda_j)\end{aligned}$$

■

Note que as distribuições condicionais obtidas na proposição acima não correspondem a um procedimento direto do amostrador de Gibbs, somente a primeira condicional tem forma “fechada”, ou seja, essas condicionais não são identificáveis, necessitando assim de um algoritmo diferente do Gibbs para fazer a simulação. Quando isto ocorre pode-se usar, por exemplo, o algoritmo de Metrópolis-Hastings.

3.7.8 Implementação MCMC

Baseados na função de verossimilhança aumentada alternativa dada por (3.15), um algoritmo computacional é desenvolvido para a família TRI-NA. Este algoritmo é facilmente implementado no programa Winbugs.

Para a parametrização δ , obtemos:

$$Z_{ij}^* | u_i, a_j, b_j, \delta_j, y_{ij} \sim N(m_{ij} - \delta_j v_{ij}, 1 - \delta_j) I_{**};$$

$$V_{ij} \sim N(0, 1) I(v_{ij} > 0);$$

$$U_i \sim NA(\kappa);$$

$$a_j \sim N(\mu_a, \sigma_a^2)I(a_j > 0);$$

$$b_j \sim N(0, \sigma_b^2);$$

$$\delta_j \sim U[-1, 1].$$

Em que $I_{**} = [I(Z_{ij}^* > 0)I(y_{ij} = 1) + I(Z_{ij}^* \leq 0)I(y_{ij} = 0)]$.

Para a parametrização λ , obtemos:

$$Z_{ij}^* | u_i, a_j, b_j, \lambda_j, y_{ij} \sim N \left(m_{ij} - \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}} V_{ij}, \frac{1}{1 + \lambda_j^2} \right) I_{**};$$

$$V_{ij} \sim N(0, 1)I(v_i > 0);$$

$$U_i \sim NA(\kappa);$$

$$a_j \sim N(\mu_a, \sigma_a^2)I(a_j > 0);$$

$$b_j \sim N(0, \sigma_b^2);$$

$$\lambda_j \sim t \left(0, \frac{1}{2}, 2 \right).$$

Em que $I_{**} = [I(Z_{ij}^* > 0)I(y_{ij} = 1) + I(Z_{ij}^* \leq 0)I(y_{ij} = 0)]$.

Algumas das aplicações realizadas nesta dissertação foram feitas com base nestas estruturas hierárquicas.

Capítulo 4

Teoria de Resposta ao Item

t-Assimétrica

Sabemos que a distribuição normal não identifica pontos discrepantes. Assim quando temos um conjunto de dados onde existem alunos que acertaram todas as questões e/ou alunos que erraram tudo, sugerir que a habilidade ou a CCI tenham uma distribuição normal pode não ser uma boa alternativa. Tentando sanar este problema propomos a utilização da distribuição t-Student para a CCI, considerando que esta distribuição é muito mais robusta a pontos discrepantes que a distribuição normal, devido à suas caudas serem mais pesadas.

Assim o problema inicial que enfrentamos é a escolha dos graus de liberdade ($g.l$) da distribuição t-Student. Albert e Chib (1993), sugerem que seja utilizado a distribuição t-Student com 9 $g.l$ quando queremos que esta se pareça com a distribuição logística, porém, o nosso interesse é que a distribuição t-Student apresente caudas mais grossas para identificar os pontos discrepantes presentes nos dados. Observando a Figura 4.1 que apresenta gráficos que comparam a distribuição normal padrão com a distribuição t-Student para diferentes graus de liberdade ($g.l$) verificamos que conforme diminuimos os $g.l$ conseguimos engrossar as caudas da distribuição t-Student, e que para graus de liberdade maiores teremos um comportamento similar ao obtido pela distribuição normal padrão.

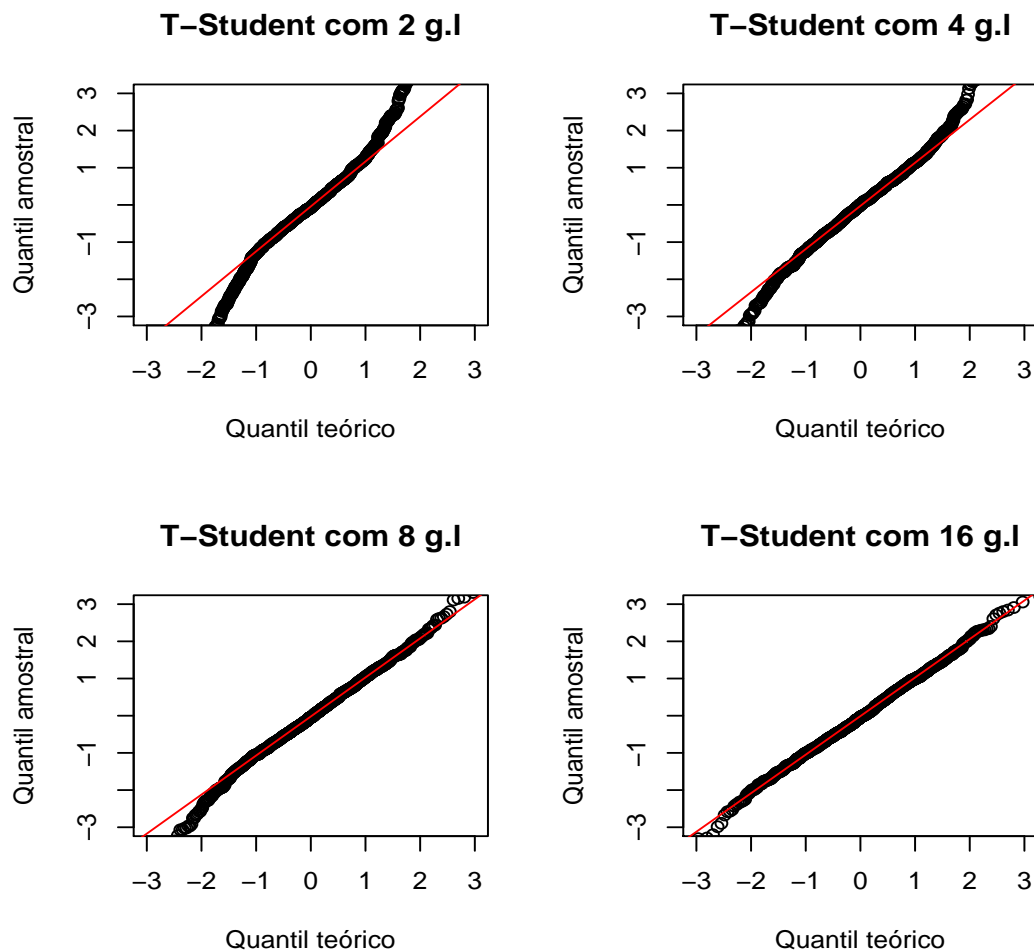


FIGURA 4.1: Densidades para diferentes g.l da distribuição t comparados com a densidade da distribuição normal padronizada.

Assim notamos que seria interessante o uso de valores baixos para os g.l da distribuição t-Student, para que esta identifique os pontos discrepantes presentes no conjunto de dados em estudo.

De acordo com o que foi visto no Capítulo 3, sabemos que o modelo probito assimétrico generalizado é obtido considerando que a distribuição da F em (3.1) é dada pela distribuição normal assimétrica com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$. O que pretendemos aqui é fazer com que esta função, que representa a CCI, tenha uma função distribuição de probabilidade acumulada da t-assimétrica e a habilidade U_i , uma distribuição normal assimétrica.

Na Seção 4.1 descrevemos algumas características da distribuição t-assimétrica para

podermos na Seção 4.2, desenvolver o modelo probito t-assimétrico. Os resultados obtidos são discutidos na Seção 4.4.

4.1 Distribuição t-assimétrica

A distribuição t-assimétrica é uma subclasse importante da família de distribuições elípticas assimétricas, como pode ser visto no Apêndice B. A densidade de uma variável Z com distribuição t-assimétrica é dada pela equação (B.5).

Para provar alguns resultados desta seção usamos o Lema 4.1 dado por Azzalini e Capitanio (2003).

Lema 4.1 *Se $V \sim G(\psi, \xi)$, então para qualquer $a, b, \in \mathbb{R}$*

$$E_V \left\{ \Phi(a\sqrt{V} + b) \right\} = P(T' \leq a\sqrt{\psi/\xi}) \quad (4.1)$$

em que T' possui uma distribuição de probabilidades t-Student com parâmetro de não centralidade $-b$ e 2ψ graus de liberdade.

A demonstração deste Lema pode ser encontrada em Azzalini e Capitanio (2003).

Uma importante característica da distribuição t-Student, é que esta pode ser representada como a mistura de normais. Para maiores detalhes consultar Johnson e Kotz (1970). Este resultado também é válido para obtermos a distribuição t-assimétrica, mas ao invés de utilizarmos uma distribuição normal, usamos a distribuição normal assimétrica.

Teorema 4.1 (Teorema da representação estocástica da t-assimétrica) *Considere as variáveis aleatórias V e Z independentes, tal que $V \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ e $Z \sim NA(\lambda)$. Então a variável $Y = \frac{Z}{\sqrt{V}}$ possui uma distribuição t-assimétrica com parâmetro de assimetria λ e ν graus de liberdade.*

$$Y = \frac{Z}{\sqrt{V}} \sim tA(0, 1, \lambda, \nu). \quad (4.2)$$

Prova:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(V^{-\frac{1}{2}}Z \leq y) = \int_0^\infty P\left(Z \leq v^{\frac{1}{2}}y \mid V = v\right) f_V(v) dv \\
&= \int_0^\infty F_Z(v^{\frac{1}{2}}y) f_V(v) dv \\
f_Y(y) &= \int_0^\infty \frac{d}{dy} F_Z(v^{\frac{1}{2}}y) f_V(v) dv \\
&= \int_0^\infty f_Z(v^{\frac{1}{2}}y) v^{\frac{1}{2}} f_V(v) dv \\
&= \int_0^\infty 2\phi(v^{\frac{1}{2}}y) \Phi(\lambda v^{\frac{1}{2}}y) v^{\frac{1}{2}} f_V(v) dv \\
&= 2 \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}vy^2\right\} v^{\frac{1}{2}} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} v^{\frac{\nu}{2}-1} \exp\left\{-\frac{\nu}{2}v\right\} \Phi(\lambda v^{\frac{1}{2}}y) dv \\
&= 2 \int_0^\infty \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}} v^{\frac{\nu+1}{2}-1} \exp\left\{-\left(\frac{\nu+y^2}{2}\right)v\right\} \Phi(\lambda v^{\frac{1}{2}}y) dv \\
&= 2 \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{\nu+y^2}{2}\right)^{-\frac{\nu+1}{2}}} \int_0^\infty \frac{\left(\frac{\nu+y^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} v^{\frac{\nu+1}{2}-1} \exp\left\{-\left(\frac{\nu+y^2}{2}\right)v\right\} \Phi(\lambda y v^{\frac{1}{2}}) dv.
\end{aligned}$$

De acordo com o Lema 4.1, obtemos: $a = \lambda y$, $b = 0$, $\Psi = \frac{\nu+1}{2}$, $\xi = \frac{\nu+y^2}{2}$, $g.l = \nu + 1$.

$$\begin{aligned}
\Rightarrow f_Y(y) &= 2 \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{\nu+y^2}{2}\right)^{-\frac{\nu+1}{2}}} P\left\{T' \leq \lambda y \sqrt{\frac{\nu+1}{\nu+y^2}}\right\} \\
&\propto 2 \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}} \left(\frac{\nu+y^2}{\nu}\right)^{-\frac{\nu+1}{2}} T_{\nu+1}\left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}}\right) \\
&\propto 2t_\nu T_{\nu+1}\left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}}\right).
\end{aligned}$$

■

Usando o Teorema da representação estocástica da t-assimétrica (4.1) e o Lema 4.1, podemos obter a f.d.a da t-assimétrica, e esta é dada pelo Teorema 4.2.

Teorema 4.2 A função de distribuição acumulada (f.d.a) da variável aleatória X , em que $X \sim tA(0, 1, \lambda, \nu)$, é dada por:

$$\int_0^\infty T_\nu\left(\frac{x}{\sqrt{1-\delta^2}}\right) f_W(w) dw, \quad (4.3)$$

em que T é a f.d.a de uma distribuição t-Student com ν g.l e parâmetro de não centralidade $-b = \frac{\delta w}{\sqrt{1 - \delta^2}}$.

Prova:

Considere $X = \frac{Z}{\sqrt{V}}$, como no Teorema 4.1,

$$\begin{aligned} P(X \leq x) &= P(V^{-\frac{1}{2}}Z \leq x) = \int_0^\infty P(Z \leq xv^{\frac{1}{2}}|V = v)f_V(v) dv \\ &= \int_0^\infty P(\delta W + \sqrt{1 - \delta^2}Y \leq xv^{\frac{1}{2}}|V = v)f_V(v) dv \quad \text{com } \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}} \\ &= \int_0^\infty \int_0^\infty P\left(Y \leq \frac{xv^{\frac{1}{2}} - \delta w}{\sqrt{1 - \delta^2}}|V = v, W = w\right) f_V(v) dv f_W(w) dw \\ &= \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{xv^{\frac{1}{2}} - \delta w}{\sqrt{1 - \delta^2}}\right) f_V(v) dv \right] f_W(w) dw. \end{aligned}$$

De acordo com o Lema 4.1, obtemos: $a = \frac{x}{\sqrt{1 - \delta^2}}$, $b = \frac{-\delta w}{\sqrt{1 - \delta^2}}$ e g.l = ν .

$$\begin{aligned} \Rightarrow F_X(x) &= \int_0^\infty P\left(T' \leq \frac{x}{\sqrt{1 - \delta^2}}\right) f_W(w) dw \\ &= \int_0^\infty T_\nu\left(\frac{x}{\sqrt{1 - \delta^2}}\right) f_W(w) dw, \end{aligned}$$

em que T_ν é a f.d.a de uma t-Student com ν g.l e parâmetro de não centralidade $-b = \frac{\delta w}{\sqrt{1 - \delta^2}}$.

■

Teorema 4.3 A função de distribuição acumulada (f.d.a) da variável aleatória X , em que $X \sim tA(0, 1, \lambda, \nu)$, pode ser, alternativamente, dada por:

$$F_X = 2 T_2 \left(\begin{pmatrix} 0 \\ x \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix} \right) \quad \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}, \quad (4.4)$$

em que T_2 representa a f.d.a da t-Student com vetor de médias nulo e com coeficiente de correlação $-\delta$, $|\delta| \leq 1$.

Prova:

Considere $X = \frac{Z}{\sqrt{V}}$, como no Teorema 4.1,

$$P(X \leq x) = P(V^{-\frac{1}{2}}Z \leq x).$$

De acordo com a Proposição A.5, podemos escrever $Z = Y_2|Y_1 > 0$, em que $(Y_1, Y_2) \sim N_2$ com densidades marginais padronizadas e coeficiente de correlação δ . Assim,

$$\begin{aligned} P(X \leq x) &= P(V^{-\frac{1}{2}}Y_2 \leq x|Y_1 > 0) \\ &= \frac{P(V^{-\frac{1}{2}}Y_2 \leq x, Y_1 > 0)}{P(Y_1 > 0)} \\ &= 2P \left\{ V^{-\frac{1}{2}} \begin{pmatrix} -Y_1 \\ Y_2 \end{pmatrix} \leq \begin{pmatrix} 0 \\ x \end{pmatrix} \right\} \\ &= 2P \left\{ T' \leq \begin{pmatrix} 0 \\ x \end{pmatrix} \right\}, \end{aligned}$$

em que T' possui densidade t-Student. Logo,

$$\Rightarrow F_X = 2 T_2 \left(\begin{pmatrix} 0 \\ x \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix} \right).$$

■

De acordo com estes Teoremas, percebemos uma grande similaridade da distribuição t-assimétrica com a distribuição normal assimétrica, o que facilita a obtenção do modelo probito t-assimétrico, como veremos a seguir.

4.2 Verossimilhança aumentada para o modelo probito t-assimétrico na TRI

Nesta seção, desenvolvemos o modelo probito t-assimétrico abordando variáveis latentes auxiliares como no capítulo anterior. Esta estrutura possui características similares às apresentadas pelos modelos probito assimétricos, como veremos a seguir. Analogamente ao Capítulo 3.4 consideramos as variáveis auxiliares como na equação 3.11, porém assumimos que:

$$\begin{aligned} Z_{ij} &= m_{ij} + e_{ij}, \\ e_{ij} &\sim tA(0, 1, -\lambda_j, \nu), \quad i = 1, \dots, n, j = 1, \dots, k. \end{aligned} \tag{4.5}$$

Sabemos de acordo com o Teorema 4.1 que a distribuição t-assimétrica pode ser escrita como uma mistura da Normal Assimétrica, assim podemos escrever os erros da forma:

$$e_{ij} = \sqrt{V_{ij}}X_{ij},$$

em que $V_{ij} \sim GI\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ e $X_{ij} \sim NA(-\lambda_j)$.

Reescrevendo Z_{ij} , obtemos:

$$Z_{ij} = m_{ij} + \sqrt{V_{ij}}X_{ij}, \quad Z_{ij} \sim tA(m_{ij}, 1, -\lambda_j, \nu) \quad (4.6)$$

Note que $Z_{ij}|V_{ij} = v_{ij} \sim NA(m_{ij}, v_{ij}, -\lambda_j)$. Ao considerarmos a representação de Henze (1986) apresentada na Proposição A.3 para a variável X_{ij} , facilitamos a maneira de representar a variável Z_{ij} .

$$Z_{ij} = m_{ij} - \sqrt{V_{ij}}(\delta_j R_{ij} + \sqrt{1 - \delta_j^2} W_{ij}), \quad (4.7)$$

onde $R_{ij} \sim N(0, 1)I(r_{ij} > 0)$ e $W_{ij} \sim N(0, 1)$.

Seja $Z'_{ij} = (Z_{ij}|V_{ij}, R_{ij})$, note que:

$$Z'_{ij} \sim N(m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}, v_{ij}(1 - \delta_j^2)). \quad (4.8)$$

Logo, a função de verossimilhança para dados aumentados $D = (Z, V, R, y)$ no modelo t-assimétrico, em que Z, V, R , são matrizes $n \times k$ é dada por:

$$L(U, \beta, \delta|D) = \prod_{i=1}^n \prod_{j=1}^k \phi(Z'_{ij}; m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}, v_{ij}(1 - \delta_j^2)) \times \phi(r_{ij}; 0, 1)I(r_{ij} > 0)f(v_{ij})I' \quad (4.9)$$

onde $f(v_{ij})$ é a f.d.p de uma distribuição $GI\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ e $I' = I(Z'_{ij} > 0)I(y_{ij} = 1) + I(Z'_{ij} \leq 0)I(y_{ij} = 0)$.

Considerando Z'_{ij} como em (4.8), e usando o Lema 4.1, dado por Azzalini e Capitanio (2003), podemos obter facilmente a distribuição de p_{ij} .

Proposição 4.1 *O modelo TRI proibito t-assimétrico, com k itens e n indivíduos, com $y_{ij} \sim Bernoulli(p_{ij})$ e $p_{ij} = F_{tA}(m_{ij}, 1, -\lambda_j, \nu)$ na qual $m_{ij} = a_j u_i - b_j$, é equivalentemente definido considerando que*

$$y_{ij} = \begin{cases} 1, & Z'_{ij} > 0, \\ 0, & Z'_{ij} \leq 0, \end{cases} \quad (4.10)$$

em que $Z'_{ij} \sim N(m_{ij} - \delta_j R_{ij} \sqrt{V_{ij}}, (1 - \delta_j^2) V_{ij})$, V_{ij} possui uma distribuição gama invertida, isto é, $V_{ij} \sim GI\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, e $R_{ij} \sim N(0, 1)I(r_{ij} > 0)$ $j = 1, \dots, k$ e $i = 1, \dots, n$.

Prova: A probabilidade de que o indivíduo i acerte a questão j dadas as variáveis V_{ij} , R_{ij} , o vetor de parâmetros dos itens e as habilidades é dada por

$$\begin{aligned} p_{ij}(v_{ij}, r_{ij}) &= P(y_{ij} = 1 | u, \beta, \lambda) = P(Z'_{ij} > 0 | u, \beta, \lambda) \\ &= P\left(\frac{Z'_{ij} - (m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}})}{\sqrt{(1 - \delta_j^2)v_{ij}}} > -\frac{m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}}{\sqrt{(1 - \delta_j^2)v_{ij}}}\right) \end{aligned}$$

$$\text{Considere } X_{ij} = \frac{Z'_{ij} - (m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}})}{\sqrt{(1 - \delta_j^2)v_{ij}}}$$

$$\begin{aligned} p_{ij}(v_{ij}, r_{ij}) &= P\left(-X_{ij} \leq \left(\frac{m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}}{\sqrt{(1 - \delta_j^2)v_{ij}}}\right)\right) \\ &= \Phi\left(\frac{m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}}{\sqrt{(1 - \delta_j^2)v_{ij}}}\right) \\ &= \Phi\left(\frac{m_{ij}}{\sqrt{(1 - \delta_j^2)v_{ij}}} - \frac{\delta_j r_{ij}}{\sqrt{1 - \delta_j^2}}\right), \text{ considere } V_{ij}^* = \frac{1}{V_{ij}} \\ &= \Phi\left(\frac{m_{ij} \sqrt{v_{ij}^*}}{\sqrt{1 - \delta_j^2}} - \frac{\delta_j r_{ij}}{\sqrt{1 - \delta_j^2}}\right) \end{aligned}$$

Assim a probabilidade do indivíduo i acertar a questão j é dada por

$$p_{ij} = \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{m_{ij} \sqrt{v_{ij}^*}}{\sqrt{1 - \delta_j^2}} - \frac{\delta_j r_{ij}}{\sqrt{1 - \delta_j^2}}\right) g(v_{ij}^*) dv_{ij}^* \right] f(r_{ij}) dr_{ij},$$

em que f e g representam respectivamente as f.d.p de R_{ij} e V_{ij}^* . Usando o Lema 4.1, temos:

$$\begin{aligned} p_{ij} &= \int_0^\infty P\left(T' \leq \frac{m_{ij}}{\sqrt{1 - \delta_j^2}}\right) f(r_{ij}) dr_{ij}, \quad T' \sim t\left(\frac{\delta_j r_{ij}}{\sqrt{1 - \delta_j^2}}, 1, \nu\right) \\ &\Rightarrow p_{ij} = \int_0^\infty T_\nu\left(\frac{m_{ij}}{\sqrt{1 - \delta_j^2}}\right) f(r_{ij}) dr_{ij}. \end{aligned}$$

Portanto, usando o Teorema 4.2, podemos dizer que,

$$p_{ij} = F_{tA}(m_{ij}, 1, -\lambda, \nu).$$

■

De acordo com a equação (4.2) podemos escrever p_{ij} como uma mistura da f.d.a de uma t-Student ponderada pela normal truncada. Esta representação é similar a que Bazán (2005) obteve usando a normal assimétrica, ver Proposição A.4, consequentemente as Proposições 4.1 e 3.3 são similares.

4.3 Estrutura Hierárquica

Consideramos para os parâmetros do modelo t-assimétrico as mesmas prioris utilizadas na família TRI-NA, assim a estrutura hierárquica necessária para realizar o processo MCMC via Winbugs é dada abaixo.

$$Z'_{ij} | \beta_j, U_i, \delta_j, y_{ij} \sim N(m_{ij} - \delta_j r_{ij} \sqrt{v_{ij}}, v_{ij} 1 - \delta_j^2) I';$$

$$R_{ij} \sim N(0, 1) I(r_{ij} > 0);$$

$$V_{ij} \sim GI\left(\frac{\nu}{2}, \frac{\nu}{2}\right);$$

$$U_i \sim NA(\kappa);$$

$$a_j \sim N(\mu_a, \sigma_a^2) I(a_j > 0);$$

$$b_j \sim N(0, \sigma_b^2);$$

$$\delta_j \sim U[-1, 1];$$

em que $I' = [I(Z'_{ij} > 0)I(y_{ij} = 1) + I(Z'_{ij} \leq 0)I(y_{ij} = 0)]$.

Essa estrutura foi utilizada na implementação dos modelos t-assimétricos no Capítulo 5.

4.4 Discussão

Neste Capítulo é proposto uma nova ligação assimétrica para os modelos probitos com distribuição acumulada da *t*-assimétrica. O uso desta distribuição é interessante, pois além de detectar a assimetria presente nos dados, podemos detectar a presença de pontos discrepantes. O que é muito comum ocorrer na área educacional. Por exemplo, podemos ter uma sala de aula com alunos que tenham um bom rendimento num determinado teste, mas um ou dois alunos podem não ter apresentado o mesmo rendimento da maioria. Nestes casos seria interessante o uso desta nova ligação.

Uma característica importante desse novo modelo é a existência de uma similaridade entre a Proposição 4.1 e a Proposição 3.3 dada por Bazán (2005). A Proposição 4.1 poderia ser vista como uma “extensão” da Proposição 3.3 quando os graus de liberdade tendem ao infinito.

Capítulo 5

Aplicações

Estamos interessados em ilustrar os procedimentos de estimação discutidos nos capítulos anteriores, para isto realizamos aplicações em 2 conjuntos de dados distintos. Para realizarmos essas aplicações, utilizamos os softwares estatísticos livres: R e Winbugs. O R foi utilizado para a obtenção da análise descritiva dos dados e para obter as estimativas de interesse (parâmetros dos itens e habilidade média) no modelo PN. O Winbugs foi utilizado para obter as estimativas de interesse nos modelos da família TRI-NA e t-assimétrico.

Devido à grande quantidade de modelos possíveis dentro da TRI, usamos alguns métodos de seleção de modelos. A definição destes métodos pode ser encontrada na Seção 5.1.

Os programas e banco de dados utilizados se encontram no Apêndice C.

5.1 Seleção de Modelos

5.1.1 DIC (*Deviance Information Criterion*)

O DIC para a seleção de um modelo é definido por:

$$DIC = \bar{D} + \rho_D. \quad (5.1)$$

Como pode ser visto na equação (5.1), o DIC possui dois termos, um que mede a

penalidade pelo aumento da complexidade do modelo, ou seja, através do número efetivo de parâmetros ρ_D e outro que mede a bondade do ajuste e é definido pela esperança condicional da *deviance*, como é mostrado a seguir:

$$\bar{D} = E[D(\boldsymbol{\beta}, \mathbf{u})] = E[-2 \ln p(\mathbf{Y} = \mathbf{y} | \boldsymbol{\beta}, \mathbf{u})], \quad (5.2)$$

onde \mathbf{y} denota a resposta observada de \mathbf{Y} , \mathbf{u} são os parâmetros associados à habilidade dos indivíduos e $\boldsymbol{\beta}$ são os parâmetros associados aos itens.

O termo que mede o número efetivo de parâmetros, ρ_D , é definido como a diferença entre o desvio médio a posteriori (*posteriori mean of deviance*) e o desvio bayesiano dos esperados a posteriori (*deviance of posteriori mean*) obtido considerando os valores médios e esperados a posteriori dos parâmetros do modelo, como é mostrado a seguir

$$\rho_D = \bar{D} - D[E(\boldsymbol{\beta}), E(\mathbf{u})]. \quad (5.3)$$

Assim, podemos reescrever o DIC, como

$$\begin{aligned} DIC &= \bar{D} + \rho_D \\ &= 2E[D(\boldsymbol{\beta}, \mathbf{u})] - D[E(\boldsymbol{\beta}), E(\mathbf{u})]. \end{aligned} \quad (5.4)$$

O modelo que tiver o menor valor do DIC sera considerado o melhor modelo. Para um estudo mais profundo sobre este método, ver Spiegelhalter *et al.* (2002).

5.1.2 Pseudo Fator de Bayes

Um método em Inferência Bayesiana para a comparação de dois modelos é o cálculo do Fator de Bayes (FB).

$$FB_{12} = \frac{m_1(\mathbf{y})}{m_2(\mathbf{y})}, \quad (5.5)$$

onde $m_i(\mathbf{y})$ é a densidade marginal preditiva de \mathbf{y} sobre o modelo M_i , $i = 1, 2$.

$$m_i(\mathbf{y}) = \int p_i(\mathbf{y} | \theta_i) \pi(\theta_i) d\theta_i$$

Um Fator de Bayes maior que um sugere que a seleção do modelo 1, enquanto que um valor menor que um sugere a seleção do modelo 2, (Sung e Kang, 2006). Pelo fato do Fator de Bayes ser sensível à especificação de prioris $\pi(\theta_i)$, ver O'Hagan (1995), vários

autores propõe, por exemplo, o uso do Pseudo Fator de Bayes. Este foi proposto em 1979 por Geisser e Eddy e é baseado na densidade preditiva $p(y_i|y_{(-i)})$, onde $y_{(-i)} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$.

O pseudo Fator de Bayes para a seleção de dois modelos pode ser calculado com o uso da estatística CPO. Essa estatística é definida como a densidade preditiva marginal a posteriori de y_i dado y_{-i} .

$$CPO_i = p(y_i|y_{-i}) = \left\{ E_{\theta|y} \frac{1}{p(y_i|\theta)} \right\}^{-1} \quad (5.6)$$

O logaritmo da estatística CPO é conhecido como verossimilhança pseudo marginal (LPML), isto é,

$$LPML = \sum_{i=1}^n \log(CPO_i). \quad (5.7)$$

A relação entre o pseudo fator de Bayes e a verossimilhança pseudo marginal é dada pela expressão

$$PFB = \exp(LPML_1 - LPML_2) \quad (5.8)$$

onde 1 e 2 são os modelos a serem testados.

5.1.3 Soma dos quadrados dos resíduos

Assim como em Bazán (2005), usamos as somas dos quadrados residuais latentes a posteriori, isto é

$$SQRL = \sum_{i=1}^n \sum_{j=1}^k e_{ij}^* \quad (5.9)$$

em que $e_{ij}^* = e_{ij}|\beta_j, y_{ij}, z_{ij}, v_{ij}$.

Esse método de seleção é muito sensível. Caso tenhamos algum modelo que não apresente distribuição normal, que apresente pontos discrepantes, este não será um método apropriado para utilizar na escolha de modelos.

5.2 Aplicação 1

O conjunto de dados utilizado nesta seção corresponde a um teste de matemática que contém 14 itens e foi aplicado a 131 alunos de sexta série de escolas peruanas. O

programa e os dados utilizados para obtermos as estimativas de interesse foram dados por Bazán, (2005).

Antes de fazermos qualquer análise bayesiana, seria interessante fazer uma análise descritiva dos dados. Para isto faremos um resumo das estatísticas dos escores (quantidade de questões certas por aluno) do teste de matemática. Estas estatísticas são dadas na Tabela 5.1.

TABELA 5.1: *Estatísticas dos escores*

Media	10.83969	Mediana	11
Variância	3.45872	Desvio Padrão	1.859763
Mín	5	Máx	14
assimetria	-0.8042115	curtose	0.5139686

De acordo com os valores da Tabela 5.1 podemos dizer que os alunos parecem ter uma alta habilidade, pois em média eles acertaram 10,8 questões dentre 14, além disso temos a presença de assimetria à esquerda (parâmetro de assimetria negativo), o que indica a presença de escores altos. Isto pode ser observado na Figura 5.1.

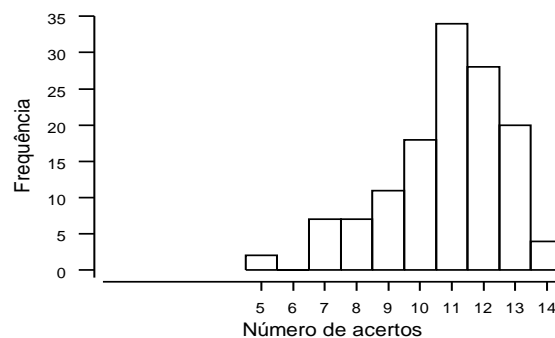


FIGURA 5.1: Histograma dos escores.

Ao olharmos para o boxplot do vetor de acertos (escores), podemos notar a existência de pontos atípicos. Isso se deve aos dois alunos que acertaram somente cinco questões, enquanto que a média de acerto foi maior que dez.

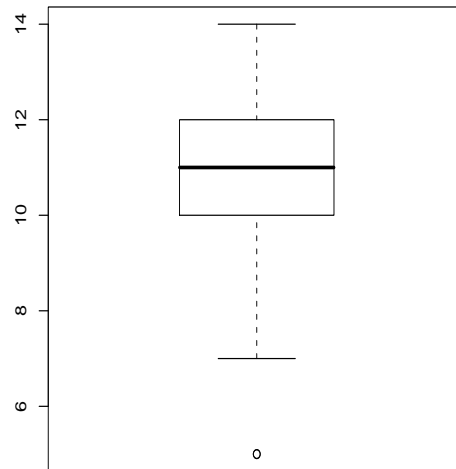


FIGURA 5.2: Boxplot dos escores.

5.2.1 Aplicação ao modelo PN

Utilizamos uma proposta bayesiana considerando o esquema DAGS (*Data augmented Gibbs sampling*) como mostrado na Seção 2.4 para o modelo probito normal, usando o conjunto de dados em questão.

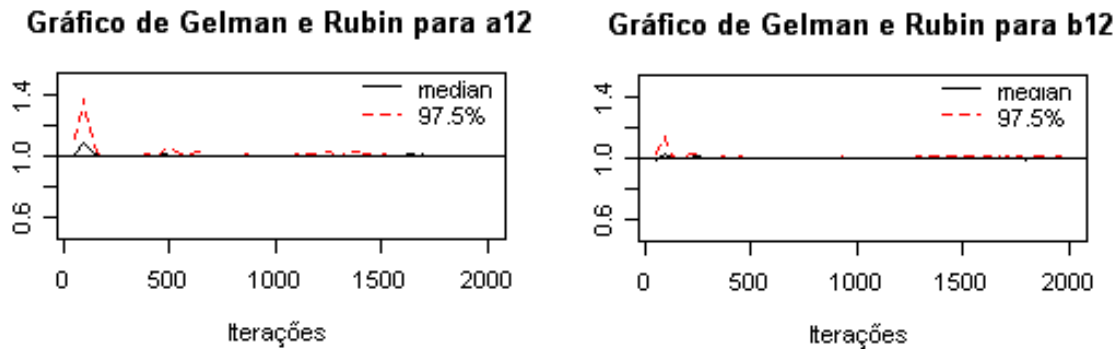
Algumas prioris para os parâmetros dos itens são citadas na literatura. Como foi dito na Subseção 2.1.2, o parâmetro de discriminação (inclinação) a_j , assume somente valores positivos pois a CCI é uma função não decrescente de u_i , assim dentre as várias prioris especificadas na literatura para os parâmetros dos itens utilizamos nesta seção a priori dada por Sahu (2002), que considera uma distribuição normal positiva para a_j e uma distribuição normal para b_j como é mostrado na Tabela 5.2. Utilizamos a notação $N(0, 1)I(a_j > 0)$ para indicar uma distribuição normal assumindo somente valores positivos com média 0 e variância 1.

É importante considerar prioris próprias para os parâmetros dos itens, pois estas garantem que as posterioris para o modelo sejam próprias, ou seja, sua função densidade de probabilidade integra 1, Albert e Ghosh (2000).

TABELA 5.2: *Priori para os parâmetros dos itens*

Priori 1	$a_j \sim N(1, 0.5)I(a_j > 0)$	$b_j \sim N(0, 2)$
----------	--------------------------------	--------------------

O procedimento de estimação bayesiana usando DAGS como proposto na Seção 2.4 foi implementado usando o software R. Utilizamos um valor de descarte (burn-in) igual a 4000, período necessário para o aquecimento das cadeias, e foram feitas 100000 iterações após o descarte com um salto (“thin”) igual a 50, tamanho necessário para diminuir a autocorrelação existente entre os valores amostrados, gerando uma amostra de tamanho 2000. Além disso, usamos duas cadeias paralelas e verificamos a convergência através visualização do gráfico de Gelman e Rubin mostrado na Figura 5.3.

FIGURA 5.3: Teste de convergência para os parâmetros a_{12} e b_{12} .

A densidade empírica e o histórico das cadeias podem ser vistos na Figura 5.4. De acordo com as densidades apresentadas, podemos dizer que estas apresentam uma única moda e em relação ao histórico notamos que as cadeias convergem a um valor estacionário. O teste de Gelman e Rubin foi efetuado e verificou-se a convergência para todos os parâmetros do teste. É importante deixar claro que foi escolhido um item arbitrário para mostrar que a distribuição a posteriori dos parâmetros converge a um valor estacionário.

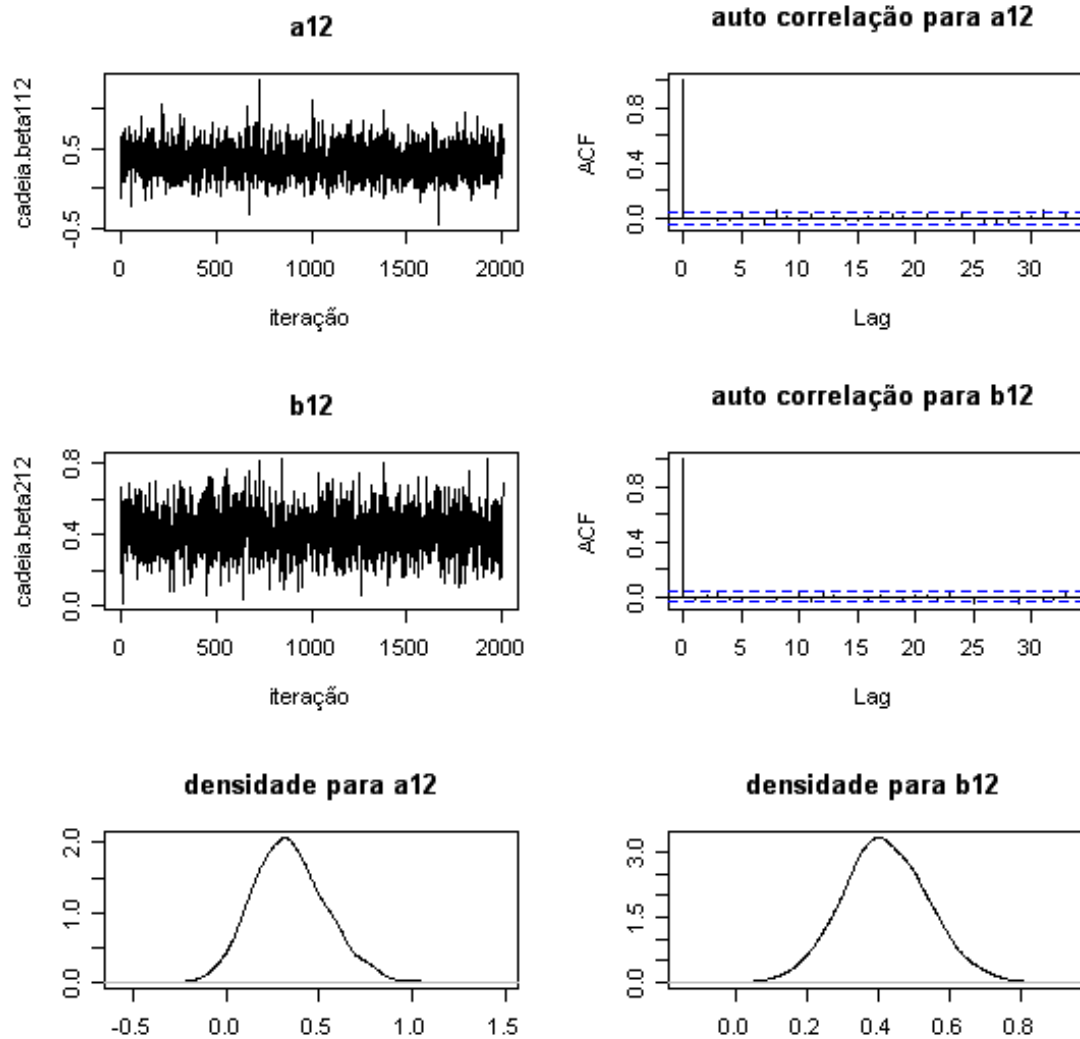


FIGURA 5.4: Densidade, histórico e autocorrelação da cadeia para os parâmetros do item 12.

De acordo com as estimativas dos parâmetros dos itens apresentadas na Tabela 5.3, podemos concluir que o item 12 é o item mais difícil da prova, pois $b[12]$ apresenta o maior valor e que o item 11 é o mais fácil. Em relação ao parâmetro de discriminação podemos dizer que o item 9 é o menos discriminatório enquanto que o item 11 é o mais discriminatório. Gráficos para os itens 9,11,12 com suas respectivas CCI são apresentados na Figura 5.5.

TABELA 5.3: *Estatísticas do modelo Probitto Normal*

node	mean	sd	MC error	2.5%	median	97.5%
a[1]	0.562	0.242	0.003	0.13	0.542	1.086
a[2]	0.245	0.215	0.003	-0.154	0.232	0.682
a[3]	0.519	0.225	0.003	0.12	0.503	1.01
a[4]	0.903	0.359	0.005	0.318	0.895	1.717
a[5]	0.501	0.252	0.003	0.054	0.482	1.044
a[6]	0.305	0.186	0.02	-0.043	0.299	0.703
a[7]	0.713	0.319	0.005	0.168	0.681	1.421
a[8]	0.985	0.354	0.005	0.401	0.949	1.76
a[9]	0.093	0.196	0.003	-0.274	0.08	0.508
a[10]	0.483	0.245	0.003	0.036	0.467	1.008
a[11]	1.285	0.401	0.006	0.580	1.25	2.15
a[12]	0.341	0.198	0.003	-0.015	0.328	0.761
a[13]	0.509	0.242	0.003	0.088	0.490	1.03
a[14]	0.435	0.310	0.004	-0.1070	0.409	1.1063
b[1]	-0.906	0.156	0.002	-1.223	-0.899	-0.622
b[2]	-1.079	0.143	0.002	-1.368	-1.075	-0.814
b[3]	-0.031	0.123	0.001	-0.268	-0.032	0.210
b[4]	-1.920	0.348	0.005	-2.752	-1.874	-1.363
b[5]	-1.228	0.181	0.002	-1.618	-1.215	-0.914
b[6]	0.346	0.119	0.001	0.113	0.348	0.585
b[7]	-1.631	0.269	0.004	-2.262	-1.6	-1.19
b[8]	-1.541	0.296	0.004	-2.214	-1.505	-1.065
b[9]	-0.799	0.126	0.001	-1.046	-0.797	-0.559
b[10]	-1.183	0.175	0.0027	-1.556	-1.17	-0.867
b[11]	-2.249	0.437	0.006	-3.24	-2.191	-1.559
b[12]	0.418	0.119	0.0018	0.187	0.4161	0.662
b[13]	-1.018	0.1604	0.003344	-1.365	-1.006	-0.727
b[14]	-1.737	0.245	0.0038	-2.290	-1.716	-1.328
média u	0.041	0.084	0.002	-0.123	0.0415	0.207
dp u	0.932	0.061	0.001	0.8143	0.931	1.054

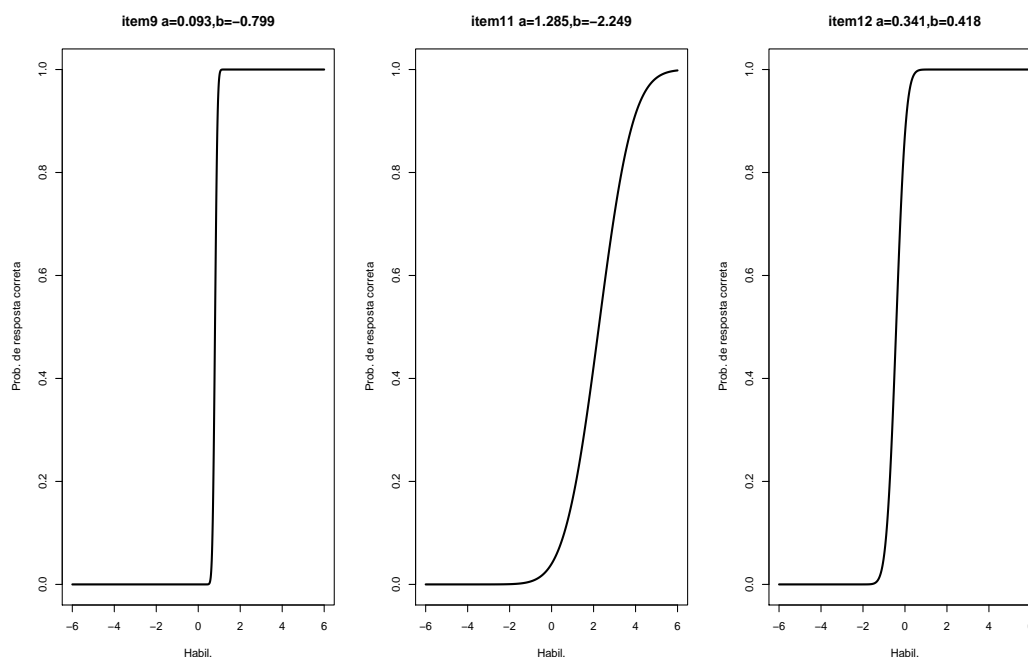


FIGURA 5.5: CCI para os itens 9,11 e 12.

É interessante sabermos o que diz cada um dos três itens identificados no teste como sendo o mais difícil, o mais fácil e o mais discriminatório, a fim de verificarmos se faz sentido estes resultados.

Item 9: Resolva as seguintes operações em decimais $0,75 - 0,2 + 1,2 - 0,3$.

Item 11: Luíza, Dora e Maria compram certa quantidade de tela, Luíza comprou a metade de um metro, Dora comprou 75 cm e Maria comprou 50 cm. Qual delas comprou a mesma quantidade de tela?

Item 12: Um recipiente recebe 4,5l de água por minuto. Quantos litros de água pode ter um recipiente depois de 1 hora e meia?

Podemos perceber que o item 9 não exige muito do aluno, além disso, este não parece discriminar entre alunos que possuem um conhecimento específico ou não. Já o item 11, consegue identificar entre alunos que possuem um conhecimento específico, neste caso se o aluno sabe que um metro possui 100 centímetros, certamente ele irá acertar a questão. Olhando para o item 12, percebemos que este exige mais do aluno, ele precisa interpretar bem a questão para conseguir respondê-la.

Em relação à média das habilidades dos alunos, notamos que esta é pouco maior que o valor médio sugerido a priori ($U_i \sim N(0, 1)$), mas comentamos anteriormente que de acordo com os escores, os alunos possuem alta habilidade, porém este modelo parece não estar identificando isto. Assim seria interessante testarmos outros modelos, visando melhorar a análise da média das habilidades (Aplicação 3).

5.3 Aplicação 2

Nesta seção trabalhamos com um conjunto de dados obtido de um teste de Ciências aplicado a 18 alunos de sexta série de uma escola da rede Municipal de Descalvado. O teste contém 10 itens e nosso interesse é estimar os parâmetros dos itens e a habilidade média destes alunos, a fim de verificar se esta sala de aula possui alunos em diferentes níveis de aprendizagem.

Uma análise descritiva é feita inicialmente no conjunto de dados em questão, como pode ser visto na Tabela 5.4.

TABELA 5.4: *Estatísticas dos escores*

Media	6	Mediana	6
Variância	7.5294	Desvio Padrão	2.7439
Mín	0	Máx	10
Assimetria	-0.5957	Curtose	0.35599

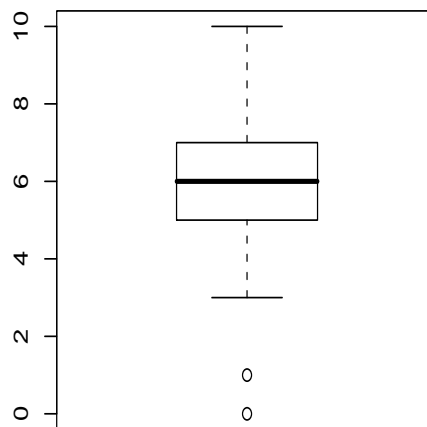


FIGURA 5.6: Boxplot dos escores.

Ao analisar a Tabela 5.4 nota-se que os dados apresentam assimetria à esquerda (parâmetro de assimetria negativo) e uma pequena curtose. Além disso verificamos a presença de alunos que erraram todas as questões e alunos que acertaram todas as questões. Ao analisar a Figura 5.13 percebemos a presença de pontos discrepantes. Afim de obter as estimativas de interesse, como estes dados apresentam uma certa assimetria e presença de *outliers*, propomos o uso de modelos assimétricos dentro da TRI-NA e o uso dos modelos t-assimétricos.

Para a escolha do melhor modelo, utilizamos o DIC e o Pseudo fator de Bayes (PFB). No software Winbugs consideramos um descarte de 4100 iterações necessários para o aquecimento da cadeia, realizamos em seguida 150000 iterações. Utilizando um salto “thin” de 50 para diminuir a autocorrelação presente entre os valores amostrados, obtemos uma amostra de tamanho 3000. Os resultados para a Priori 1 considerando os modelos com habilidade normal, isto é, $U \sim N(0, 1)$ podem ser vistos na Tabela 5.5. Considere o modelo PAN como modelo 1, e o modelo PtA-N (x) o modelo x , vamos obter o PFB para estes modelos, onde $PFB_{1,x}$ está testando o modelo 1 contra o modelo x .

TABELA 5.5: *Comparando modelos usando Priori 1 e $U \sim N(0, 1)$*

Modelo	Dbar	Dhat	ρD	DIC	$PFB_{1,x}$
PAN	143.871	135.473	8.399	152.270	*
PtA-N (6)	159.743	133.199	26.544	186.287	2.38×10^{57}
PtA-N (10)	159.865	133.324	26.541	186.406	4.24×10^{42}
PtA-N (15)	160.265	134.020	26.245	186.510	6.33×10^{36}
PtA-N (20)	160.506	134.432	26.074	186.580	4.34×10^{32}

Analisando os resultados da Tabela 5.5, notamos que de acordo com o DIC o melhor modelo é o PAN. Percebemos ao analisar o PFB que o modelo t-assimétrico foi rejeitado para todos os graus de liberdade aqui explorados. Logo, o modelo selecionado segundo os dois critérios analisados é o modelo PAN. Os resultados obtidos para os parâmetros de interesse podem ser vistos na Tabela 5.6. Além disso, mostramos na Figura 5.7 a densidade empírica, o histórico das cadeias e o Gráfico de Gelman e Rubin para o item 5 (escolha arbitrária).

Notamos pela densidade apresentada na Figura 5.7 que a e b apresentam uma única moda. Quanto aos históricos e ao gráfico de Gelman e Rubin, os parâmetros a e b parecem convergir a um valor estacionário.

De acordo com a Tabela 5.6, notamos que os itens que mais se destacam neste teste são os itens 4, 6 e 9. O item 9 apresenta o maior valor de a , o que indica que este item é o mais discriminatório do teste, ou seja é aquele que consegue identificar se o aluno possui o conhecimento necessário para responder ao item, diferentemente do item 4 que apresenta o menor valor de a , ou seja, é o item menos discriminatório do teste. Ao analisarmos o parâmetro de dificuldade b , notamos que o item 9 também representa o item mais difícil do teste, pois apresenta maior valor de b e que o item 6 é o item mais fácil do teste. Em relação à média das habilidades dos alunos envolvidos no teste, percebemos que esta é um pouco maior do que supomos a priori ($u \sim N(0, 1)$).

TABELA 5.6: *Estatísticas do modelo PAN com priori 1*

node	mean	sd	MC error	2.5%	median	97.5%
a[1]	1.018	0.4741	0.003967	0.2209	0.972	2.064
a[2]	1.229	0.5248	0.0049	0.3352	1.19	2.359
a[3]	1.194	0.5121	0.0051	0.3249	1.147	2.309
a[4]	0.658	0.3902	0.0035	0.0650	0.6075	1.557
a[5]	1.384	0.5373	0.0051	0.4561	1.347	2.524
a[6]	1.13	0.5001	0.0044	0.2509	1.096	2.204
a[7]	1.325	0.5341	0.0048	0.3871	1.284	2.473
a[8]	0.7408	0.4097	0.0033	0.09464	0.6903	1.692
a[9]	1.535	0.532	0.0051	0.5813	1.504	2.643
a[10]	1.226	0.5233	0.0049	0.3283	1.183	2.382
b[1]	0.2232	0.5322	0.0046	-0.777	0.2234	1.231
b[2]	-0.6355	0.5862	0.0059	-1.763	-0.6459	0.4725
b[3]	-0.4496	0.5755	0.0054	-1.542	-0.459	0.6393
b[4]	-0.5524	0.5084	0.0051	-1.491	-0.5641	0.4311
b[5]	-0.2745	0.5979	0.00558	-1.431	-0.2679	0.8399
b[6]	-1.805	0.7856	0.0086	-3.433	-1.775	-0.3895
b[7]	0.4481	0.5883	0.0053	-0.6471	0.4463	1.591
b[8]	-0.2365	0.5024	0.0042	-1.17	-0.2433	0.7211
b[9]	0.9033	0.6572	0.0066	-0.3038	0.8976	2.205
b[10]	-0.2649	0.5721	0.0051	-1.342	-0.2714	0.8027
média u	0.06967	0.2035	0.0018	-0.3311	0.0702	0.4675
dp u	0.9817	0.1476	0.0013	0.7104	0.9754	1.29

TABELA 5.7: *Estatísticas do parâmetro de assimetria δ no modelo PAN*

$\delta[j]$	Priori 1					Priori 2				
	média	dp	2.5%	mediana	97.5	média	dp	2.5%	mediana	97.5%
$\delta[1]$	-0.017	0.513	-0.9254	-0.013	0.875	-0.033	0.008	-0.930	-0.040	0.907
$\delta[2]$	-0.049	0.535	-0.9266	-0.077	0.924	0.033	0.552	-0.914	0.039	0.942
$\delta[3]$	-0.014	0.526	-0.9076	-0.039	0.927	0.025	0.545	-0.913	0.019	0.949
$\delta[4]$	0.015	0.499	-0.8611	0.002	0.920	0.037	0.525	-0.868	0.028	0.945
$\delta[5]$	-0.042	0.554	-0.9487	-0.063	0.927	-0.010	0.574	-0.954	-0.019	0.951
$\delta[6]$	-0.273	0.549	-0.9789	-0.386	0.878	0.010	0.578	-0.949	-4.7E-4	0.954
$\delta[7]$	0.037	0.549	-0.9417	0.060	0.926	-0.010	0.566	-0.952	-0.021	0.937
$\delta[8]$	0.019	0.501	-0.8614	0.006	0.923	0.048	0.520	-0.879	0.050	0.942
$\delta[9]$	0.115	0.585	-0.9476	0.190	0.962	0.002	0.012	-0.959	-0.008	0.958
$\delta[10]$	-0.019	0.533	-0.9263	-0.027	0.934	0.023	0.549	-0.930	0.018	0.947

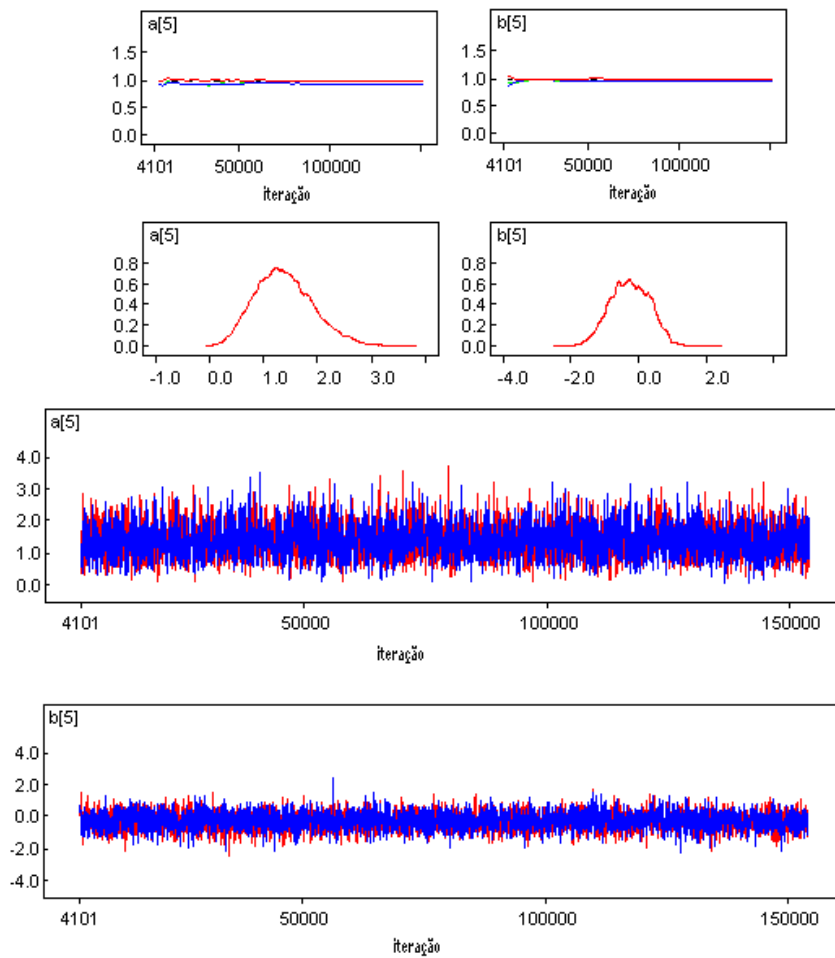
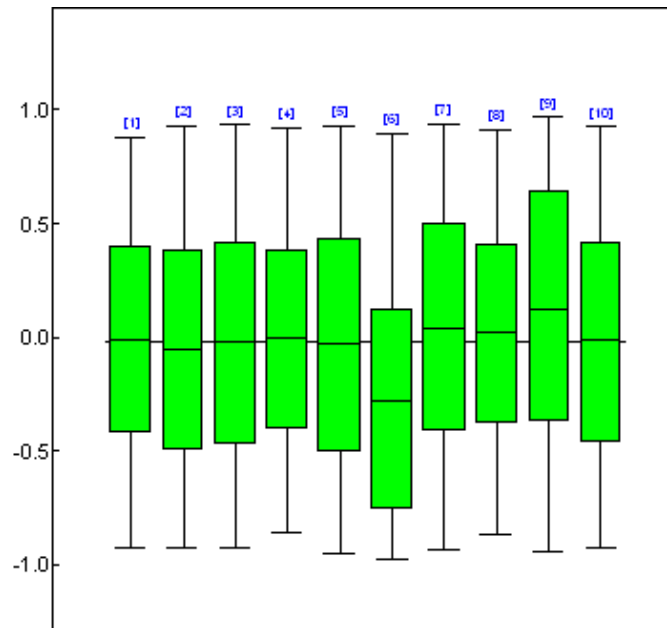


FIGURA 5.7: Gráfico de Gelman e Rubin, densidade e histórico para o item 5.

FIGURA 5.8: Boxplot para o parâmetro de assimetria δ .

Na Tabela 5.7, estão presentes os valores estimados para o parâmetro de assimetria δ e na Figura 5.8 apresentamos o seu boxplot. Notamos que apenas os itens 6 e 9 apresentam uma significativa assimetria, pelo boxplot pode se notar que o item 6 apresenta menor valor de δ (assimetria negativa) e que o item 9 apresenta o maior valor de δ (assimetria positiva). Como foi dito na Subseção 3.7.2, o parâmetro de assimetria negativo acaba bonificando aqueles alunos que possuem baixa habilidade, pois num pequeno acréscimo da habilidade, aumenta-se consideravelmente a probabilidade de acerto do aluno. Já no caso do item 9, que possui parâmetro de assimetria positivo, a questão acaba se tornando mais difícil, pois para acertá-la é preciso ter habilidades com valores altos. Para ilustrar este fato, fazemos uma figura que considera os itens 6 e 9 com as estimativas do modelo PN e do modelo PAN.

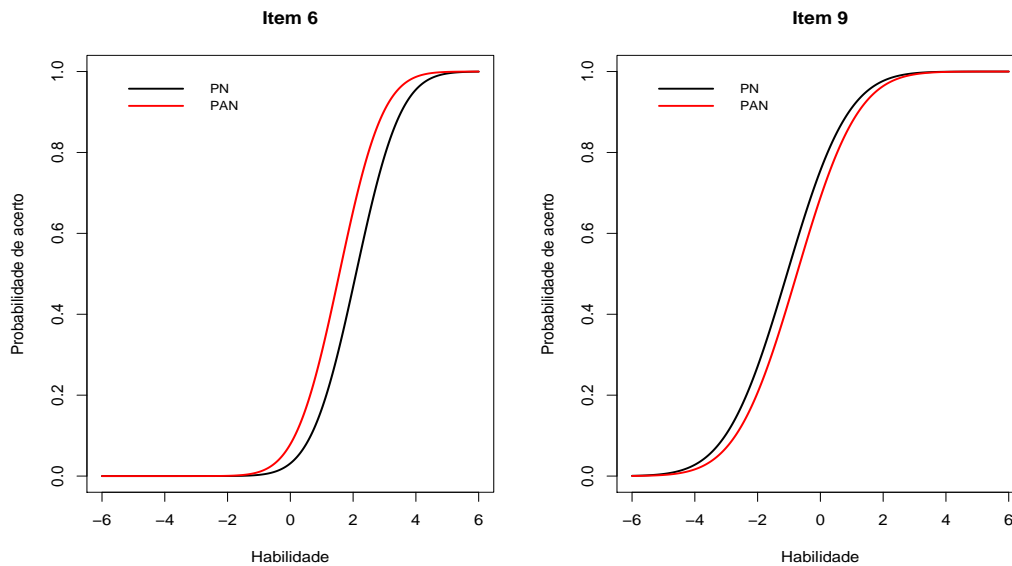


FIGURA 5.9: CCI para os modelos PN e PAN nos itens mais significativos do teste.

5.4 Aplicação 3

Como o resultado obtido na Seção 5.2 não foi muito plausível com a análise descritiva dos dados, propomos uma nova análise ao conjunto de dados referentes aos alunos das escolas peruanas, levando em consideração inicialmente, os modelos da TRI-NA propostos no Capítulo 3 e em seguida usamos os modelos probito t-assimétricos propostos no Capítulo 4.

Trabalhamos com a priori 1 utilizada anteriormente no modelo PN e com uma priori difusa para b_j considerada por Spiegelhalter *et al* (1996), como pode ser visto na Tabela 5.8.

TABELA 5.8: *Prioris utilizadas*

Priori 1	$a_j \sim N(1, 0.5)I(a_j > 0)$	$b_j \sim N(0, 2)$	$\delta \sim U(-1, 1)$
Priori 2	$a_j \sim N(0, 1)I(a_j > 0)$	$b_j \sim N(0, 10000)$	$\delta \sim U(-1, 1)$

Para a escolha do melhor modelo utilizamos o DIC (*Deviance Information Criterion*) e a SQRL.

Para obter os DIC's e a SQRL relacionados aos modelos propostos na TRI-NA, utilizamos o software Winbugs, para o procedimento geramos 4100 iterações, necessárias para o aquecimento da cadeia (*burn in*) e em seguida 200000 iterações, com um salto "thin" igual a 100, para diminuir a autocorrelação presente entre os valores amostrados. Os resultados obtidos para as prioris 1 e 2 podem ser vistos nas Tabelas 5.9 e 5.10, respectivamente.

TABELA 5.9: *Comparando modelos usando a priori 1*

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PN	1455.770	1367.780	87.994	1543.770	1852.0
PAN	1336.510	1374.800	-38.294	1298.210	1366.0
PANA	1331.250	1379.530	-48.280	1282.970	1344.0

TABELA 5.10: *Comparando modelos usando a priori 2*

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PN	1456.030	1368.100	87.931	1543.960	1856.0
PAN	1320.430	1382.800	-62.372	1258.060	1343.0
PANA	1321.640	1395.120	-73.487	1248.150	1313.0

Analisando a Tabela 5.10, notamos que o melhor modelo, segundo o DIC e a SQRL é o PANA, que envolve assimetria na curva característica dos itens e nas habilidades dos indivíduos. O mesmo modelo é escolhido ao analisarmos a Tabela 5.9. Comparando as duas Tabelas, notamos que para os modelos com CCI assimétrica os valores dos DIC's e da SQRL obtidos foram menores quando usamos a priori 2. Assim podemos concluir que os modelos da TRI-NA são sensíveis à especificação de prioris. Segundo Bazán (2005), com o uso da priori 2 as estimativas obtidas para o coeficiente de assimetria δ não identifica a assimetria presente na distribuição dos dados. Deste modo, resolvemos trabalhar com as duas prioris para obter as estimativas de δ e comparar os resultados.

Na Tabela 5.11 confirmamos os resultados de Bazán (2005). Percebemos que os valores estimados do parâmetro δ , ficam todos em torno de zero quando é utilizada a priori 2. No parâmetro $\delta[11]$ fica evidente essa diferença, assim consideramos as estimativas para os parâmetros de interesse utilizando a priori 1.

TABELA 5.11: Estatísticas para o parâmetro δ no modelo PANA

$\delta[j]$	Priori 1					Priori 2				
	média	dp	2.5%	mediana	97.5	média	dp	2.5%	mediana	97.5%
$\delta[1]$	0.011	0.501	-0.869	-0.003	0.924	0.022	0.526	-0.904	0.031	0.933
$\delta[2]$	-0.013	0.486	-0.844	-0.036	0.915	0.053	0.52	-0.855	0.047	0.943
$\delta[3]$	0.060	0.497	-0.871	0.069	0.918	-0.011	0.528	-0.93	-0.013	0.912
$\delta[4]$	-0.052	0.556	-0.934	-0.092	0.931	0.003	0.564	-0.932	6.7E-4	0.947
$\delta[5]$	0.009	0.509	-0.877	-0.025	0.932	0.057	0.516	-0.864	0.061	0.950
$\delta[6]$	0.030	0.484	-0.871	0.051	0.878	-0.036	0.515	-0.932	-0.047	0.905
$\delta[7]$	-0.012	0.530	-0.896	-0.040	0.935	0.047	0.540	-0.897	0.057	0.948
$\delta[8]$	-0.019	0.554	-0.928	-0.034	0.944	0.014	0.569	-0.935	0.002	0.958
$\delta[9]$	0.026	0.498	-0.840	0.004	0.928	0.046	0.517	-0.885	0.046	0.937
$\delta[10]$	0.007	0.512	-0.851	-0.028	0.939	0.065	0.527	-0.865	0.063	0.954
$\delta[11]$	-0.167	0.614	-0.983	-0.270	0.947	-0.063	0.616	-0.976	-0.084	0.953
$\delta[12]$	0.031	0.501	-0.889	0.047	0.904	-0.007	0.515	-0.923	-0.004	0.893
$\delta[13]$	0.012	0.503	-0.87	-0.009	0.935	0.049	0.525	-0.865	0.056	0.954
$\delta[14]$	-0.017	0.513	0.861	-0.061	0.928	0.053	0.528	-0.880	0.059	0.949

O boxplot de δ usando a priori 1 é dado na Figura 5.10.

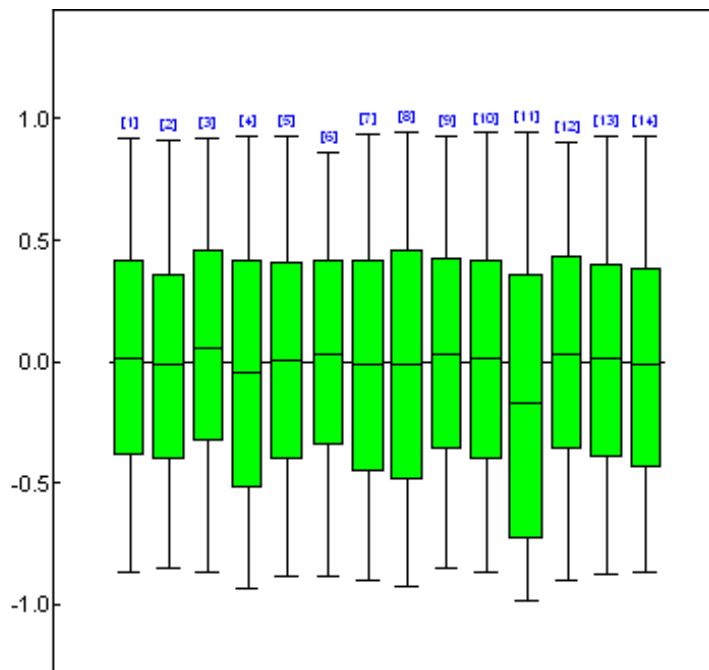
FIGURA 5.10: Boxplot para o parâmetro de assimetria δ .

TABELA 5.12: *Estatísticas do modelo PANA com priori 1*

	média	dp	erro	2.5%	mediana	97.5%
a[1]	0.951	0.403	0.015	0.261	0.918	1.817
a[2]	0.528	0.325	0.010	0.044	0.484	1.288
a[3]	0.678	0.316	0.013	0.132	0.656	1.378
a[4]	1.434	0.533	0.019	0.455	1.398	2.564
a[5]	0.904	0.428	0.014	0.201	0.848	1.84
a[6]	0.421	0.2378	0.006	0.043	0.398	0.969
a[7]	1.132	0.494	0.016	0.272	1.087	2.204
a[8]	1.449	0.509	0.016	0.538	1.41	2.518
a[9]	0.292	0.212	0.006	0.015	0.249	0.799
a[10]	0.760	0.399	0.013	0.122	0.715	1.703
a[11]	1.799	0.540	0.014	0.839	1.778	2.946
a[12]	0.435	0.246	0.009	0.048	0.408	0.978
a[13]	0.923	0.439	0.022	0.195	0.871	1.95
a[14]	0.875	0.493	0.018	0.086	0.814	1.94
b[1]	-0.200	0.441	0.012	-0.983	-0.221	0.669
b[2]	-0.630	0.439	0.015	-1.371	-0.649	0.238
b[3]	0.418	0.448	0.014	-0.411	0.436	1.265
b[4]	-0.655	0.536	0.024	-1.512	-0.726	0.446
b[5]	-0.518	0.465	0.013	-1.324	-0.543	0.399
b[6]	0.607	0.424	0.011	-0.245	0.637	1.349
b[7]	-0.697	0.494	0.015	-1.493	-0.771	0.305
b[8]	-0.300	0.526	0.026	-1.192	-0.350	0.738
b[9]	-0.523	0.410	0.012	-1.205	-0.559	0.297
b[10]	-0.555	0.447	0.011	-1.325	-0.588	0.334
b[11]	-0.436	0.586	0.024	-1.392	-0.460	0.617
b[12]	0.635	0.446	0.015	-0.2242	0.6711	1.411
b[13]	-0.299	0.457	0.020	-1.099	-0.3242	0.601
b[14]	-0.995	0.490	0.014	-1.8	-1.05	0.018
média u	0.762	0.060	0.002	0.639	0.764	0.874
dp u	0.602	0.053	0.002	0.507	0.6	0.722

Ao compararmos a Tabela 5.12 com a Tabela 5.3, notamos que embora os valores para os parâmetros a_j e b_j sejam um pouco distintos, continua-se tendo que o item 11 é o mais discriminatório e o item 9 o menos discriminatório. Além disso, o item 12 ainda é considerado o item mais difícil do teste, mas neste caso o item mais fácil é o item 14 e

não o item 11, como sugeriu as estimativas obtidas para o modelo PN.

Em relação ao valor estimado da média das habilidades, notamos que este parece explicar um pouco melhor a assimetria presente na habilidade dos alunos.

Como podemos ver nos resultados obtidos, o modelo PANA parece ser o modelo mais apropriado a este conjunto de dados, pois explica melhor a assimetria presente na distribuição dos dados, mas será que um modelo que considera a presença de pontos atípicos não se adequaria melhor a este conjunto de dados? Tentando responder a esta pergunta nos baseamos no estudo de um novo modelo dentro da TRI, o modelo probito t-assimétrico. Este modelo foi mostrado no Capítulo 4 e envolve a distribuição t-Student assimétrica.

5.4.1 Aplicação aos modelos probito t-assimétricos

Os modelos probito t-assimétricos possuem uma vantagem em relação aos modelos da família TRI-NA. Além deles detectarem a assimetria presente nos dados eles são capazes de identificar a presença de pontos discrepantes o que tornam estes modelos mais flexíveis. Assim, se estivermos trabalhando com um conjunto de dados que apresente pontos discrepantes necessitamos de uma CCI t-assimétrica com os g.l assumindo valores baixos, mas caso isso não ocorra basta utilizar a CCI t-assimétrica com g.l maiores, pois assim como a distribuição t-Student se aproxima de uma distribuição normal quando os g.l tendem ao infinito, a distribuição t-assimétrica se aproxima de uma normal assimétrica conforme aumentamos os g.l. Para exemplificar este fato, temos a Figura 5.11, que mostra a densidade de uma t-assimétrica comparada com a densidade uma normal assimétrica.

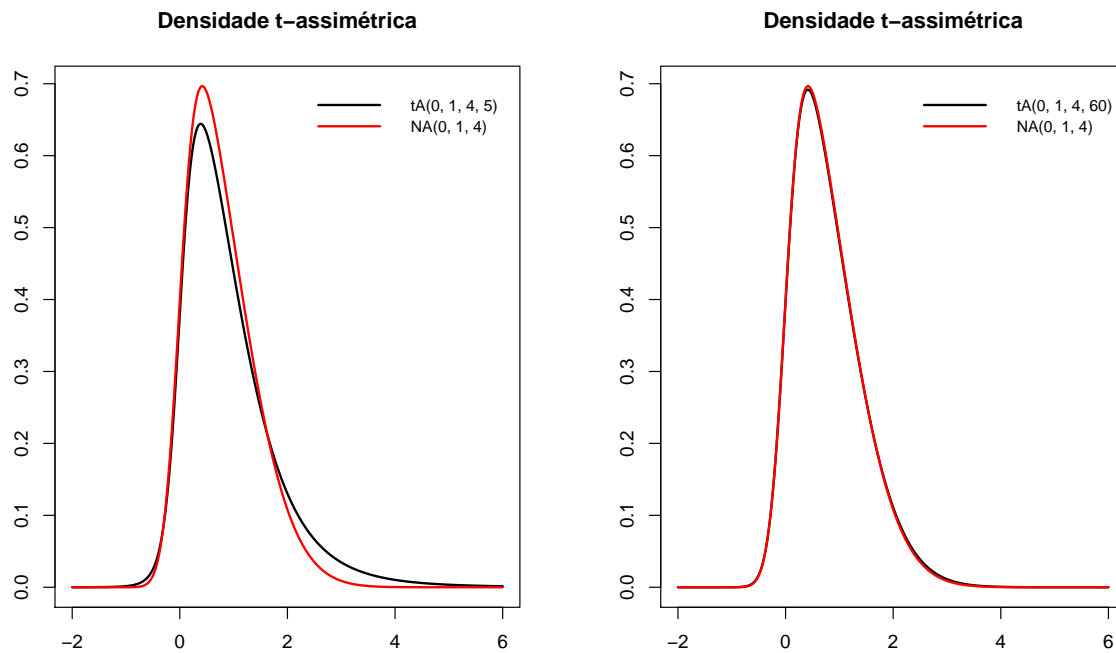


FIGURA 5.11: Densidade tA comparada com a densidade NA.

Observando a Figura 5.11, notamos que a densidade t-assimétrica com graus de liberdade igual a 60 é praticamente similar à densidade da distribuição normal assimétrica.

Comparamos os DIC's e a soma dos quadrados residuais latentes (SQRL) para os modelos probito t-assimétricos e analisamos conjuntamente com os resultados já obtidos para a família TRI-NA. Para fazer essa análise dividimos os modelos, primeiro obtemos os DIC's e a SQRL para os modelos que possuem habilidade normal, usando as prioris 1 e 2. Foi necessário um descarte (*burn in*) de 4100 iterações para o aquecimento das cadeias, para em seguida gerar 200000 iterações, com um salto "thin" igual a 100, para diminuir a autocorrelação presente entre os valores amostrados obtendo assim uma amostra de tamanho 2000. Os resultados obtidos para a priori 1 podem ser visto nas Tabelas 5.15 e 5.16 e os obtidos para a priori 2 podem ser vistos nas Tabelas 5.13 e 5.14.

TABELA 5.13: Comparando modelos com a priori 2 e $U \sim N(0, 1)$

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PAN	1320.430	1382.800	-62.372	1258.060	1313
PtA-N (4)	1345.370	1664.070	-318.697	1026.670	16980
PtA-N (8)	1389.340	1857.810	-468.475	920.863	15580
PtA-N (10)	1398.670	1916.700	-518.028	880.644	15690
PtA-N (12)	1404.100	1949.470	-545.370	858.729	15960
PtA-N (14)	1408.420	1953.710	-545.286	863.134	15600
PtA-N (15)	1410.510	1965.470	-554.954	855.557	15580
PtA-N (20)	1415.670	1958.150	-542.476	873.193	14810

Um gráfico para analisar a tendência do DIC conforme mudamos os g.l dos modelos PtA-N com o uso da priori 2 é dado na Figura 5.12. Observamos que o valor do DIC vai diminuindo conforme aumentamos os g.l até 15 e depois começa a oscilar.

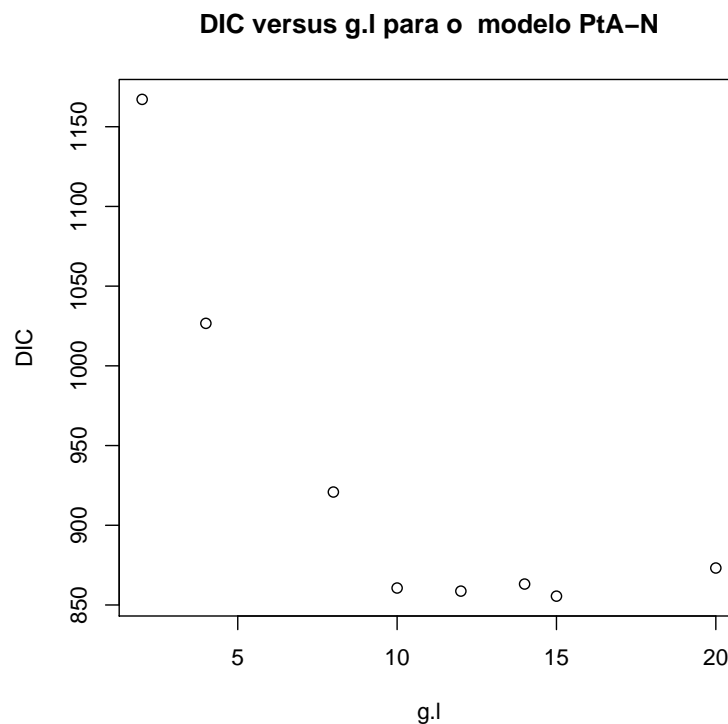


FIGURA 5.12: Tendência do DIC.

TABELA 5.14: *Comparando modelos com a priori 2 e $U \sim NA(0, 1, k)$*

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PANA	1321.640	1395.120	-73.487	1248.150	1343
PtA-NA (4)	1351.060	1640.930	-289.864	1061.200	15820
PtA-NA (8)	1395.410	1807.470	-412.060	983.351	14400
PtA-NA (10)	1405.980	1833.480	-427.505	978.473	13480
PtA-NA (12)	1409.230	1852.370	-443.139	966.093	3530

Como podemos ver nas Tabelas 5.13 e 5.14, os modelos PtA-N e PtA-NA apresentaram o valor do DIC menor do que os modelos da família TRI-NA. O mesmo não ocorreu para a SQRL, mas como foi dito na Subseção 5.1.3 este método não é adequado quando há presença de pontos atípicos no conjunto de dados. Assim consideramos apenas os resultados do DIC para selecionar o melhor modelo. Comparando conjuntamente as Tabelas 5.13 e 5.14, temos que o melhor modelo de acordo com o DIC é o PtA-N com 15 g.l, o que indica que este conjunto de dados não apresenta caudas muito pesadas, mas ainda assim há pontos discrepantes que foram detectados com o modelo PtA-N.

TABELA 5.15: *Modelos com o uso da priori 1 e $U \sim N(0, 1)$*

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PAN	1336.510	1374.800	-38.294	1298.210	1366
PtA-N (2)	1304.820	1223.700	81.117	1385.940	19540
PtA-N (4)	1358.100	1323.710	34.383	1392.480	9264
PtA-N (8)	1391.220	1380.350	1 0.877	1402.100	7186
PtA-N (12)	1412.930	1423.620	-10.697	1402.230	5822

TABELA 5.16: *Modelos com o uso da priori 1 e $U \sim NA(0, 1, k)$*

Modelo	Dbar	Dhat	ρD	DIC	SQRL
PANA	1331.250	1379.5307	-48.280	1282.970	1344
PtA-NA (4)	1364.450	1346.860	17.597	1382.050	8162
PtA-NA (8)	1398.710	1399.830	-1.117	1397.590	6334
PtA-NA (12)	1402.630	1399.880	2.747	1405.370	6724

Em relação a priori 1, temos as Tabelas 5.15 e 5.16. De acordo com os resultados obtidos, podemos perceber a influência que a seleção das prioris tem sobre a escolha dos modelos. Isto indica uma falta de informação no conjunto de dados em questão, uma maneira de evitar que isto ocorra é utilizar conjuntos de dados maiores.

As estimativas para o modelo PtA-N com 15 g.l e priori 2 são dadas na Tabela 5.17. Notemos que embora com resultados distintos os itens que se destacam no teste são os mesmos que no modelo PANA, ou seja, o item 9 como sendo o menos discriminatório, o item 11 o mais discriminatório, o item 12 o mais difícil do teste e o item 14 o mais fácil.

Em relação à média das habilidades esta é praticamente zero e não difere muito da média proposta à priori.

O boxplot para o parâmetro de assimetria δ é dado na Figura 5.13.

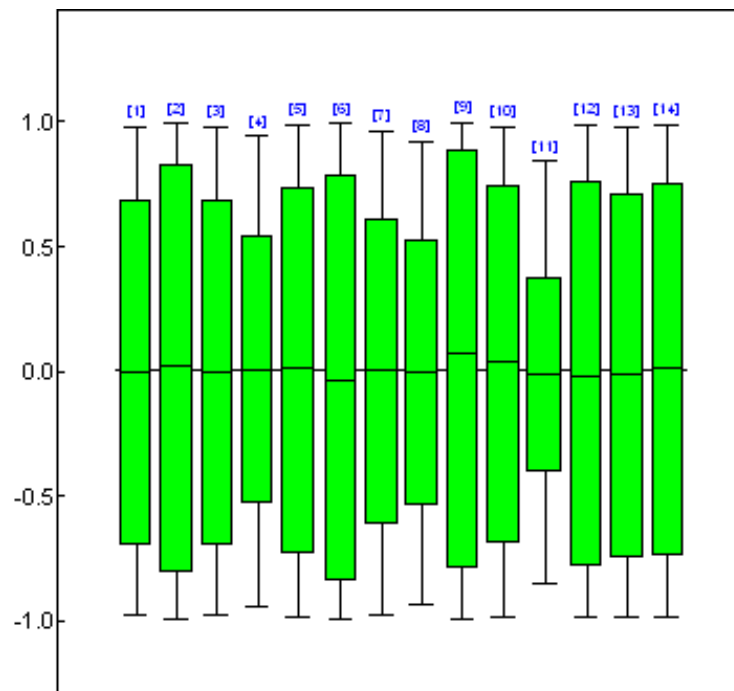


FIGURA 5.13: Boxplot para o parâmetro de assimetria δ .

TABELA 5.17: *Estatísticas do modelo PtA-N com priori 2*

	média	dp	erro	2.5%	mediana	97.5%
a[1]	0.737	0.437	0.008	0.110	0.657	1.858
a[2]	0.475	0.405	0.008	0.020	0.372	1.536
a[3]	0.759	0.432	0.007	0.141	0.680	1.882
a[4]	1.112	0.514	0.011	0.276	1.045	2.361
a[5]	0.714	0.460	0.009	0.082	0.621	1.845
a[6]	0.555	0.428	0.008	0.036	0.445	1.633
a[7]	0.916	0.479	0.009	0.172	0.854	2.069
a[8]	1.156	0.499	0.008	0.371	1.085	2.298
a[9]	0.370	0.360	0.007	0.011	0.263	1.383
a[10]	0.693	0.432	0.008	0.085	0.616	1.785
a[11]	1.549	0.568	0.012	0.607	1.501	2.809
a[12]	0.602	0.426	0.008	0.060	0.509	1.748
a[13]	0.677	0.434	0.009	0.068	0.591	1.773
a[14]	0.638	0.471	0.010	0.034	0.537	1.824
b[1]	-1.82	1.437	0.046	-5.766	-1.357	-0.505
b[2]	-3.112	2.643	0.094	-10.6	-2.169	-0.781
b[3]	-0.105	0.646	0.018	-1.305	-0.097	1.002
b[4]	-3.002	1.52	0.046	-6.919	-2.587	-1.473
b[5]	-2.756	2.141	0.066	-9.078	-2.033	-0.888
b[6]	0.967	1.106	0.034	-0.278	0.71	4.018
b[7]	-2.894	1.747	0.051	-7.669	-2.36	-1.282
b[8]	-2.301	1.081	0.024	-5.015	-2.027	-1.093
b[9]	-2.536	2.279	0.084	-9.057	-1.715	-0.444
b[10]	-2.571	1.965	0.072	-8.137	-1.953	-0.840
b[11]	-3.117	1.013	0.024	-5.57	-2.939	-1.679
b[12]	1.018	1.086	0.033	-0.169	0.752	3.968
b[13]	-2.174	1.731	0.051	-6.96	-1.60	-0.631
b[14]	-4.186	3.218	0.126	-14.07	-3.00	-1.464
média u	0.001	0.087	0.001	-0.170	0.001	0.1746
dp u	0.995	0.060	0.001	0.876	0.995	1.115

Capítulo 6

Considerações Finais

Nesta dissertação descrevemos os modelos para respostas binárias desenvolvidos na TRI com o uso da ligação probito. Para as estimativas de interesse neste modelo desenvolvemos um programa que foi implementado no software livre R. Este programa foi implementado com a finalidade de comparar os resultados obtidos por Bazán (2005) no Winbugs com os resultados obtidos no R. Notamos que os resultados obtidos nos 2 programas foram semelhantes. Apresentamos também motivados pelo trabalho de Chen *et al* (1999) e Bazán (2005), um estudo dos modelos probito assimétricos e da família TRI-NA, respectivamente. Com o conhecimento destes modelos desenvolvemos modelos mais flexíveis. No Capítulo 4 o modelo probito t-assimétrico foi desenvolvido e através da utilização do Teorema da representação estocástica da t-assimétrica e do Lema 4.1 dado por Azzalini e Capitanio (2003) conseguimos expressar a probabilidade do indivíduo acertar uma questão através da f.d.a da t-Student bivariada. E esta pode ser facilmente implementada através do algoritmo proposto por Genz e Bretz (1999). O software R também possui uma rotina disponível para este cálculo. Além disso, encontramos disponível no R um “pacote” chamado `sn` (Azzalini, 2006) que efetua estes cálculos diretamente. Com a obtenção destes resultados e da Proposição 4.1 conseguimos resultados similares aos obtidos por Bazán (2005).

Métodos de estimação usando Inferência Bayesiana foram propostos para todos os modelos apresentados, usando a abordagem de dados aumentados, o que facilitou a implementação dos programas.

Para a obtenção das estimativas dos modelos da família TRI-NA usamos os programas desenvolvidos por Bazán (2005) para o software Winbugs. E para os modelos t-assimétricos utilizamos novamente este software devido a facilidade de adaptar o modelo probito t-assimétrico através dos modelos da TRI-NA.

No Capítulo 5 ilustramos uma aplicação da teoria discutida para dois conjuntos de dados reais. Usamos na Aplicação 2 o PFB como um método alternativo aos métodos de seleção de modelos propostos na TRI-NA, ver Sung e Kang (2006). Este não foi utilizado nas Aplicações 1 e 3 devido à sobrecarga computacional. Outro método de seleção utilizado foi a SQRL. Notamos que o uso deste método não é interessante quando temos pontos atípicos presente nos dados, como pode ser visto na Seção 5.3. Nestes casos outros métodos devem ser usados.

Ainda no Capítulo 5 verificamos que a seleção da priori interfere na escolha de modelos. Na Subseção 5.4.1 ao utilizarmos a priori 1 o modelo escolhido foi o PANA e ao utilizar a priori 2 o modelo escolhido foi o PtA-N com 15 g.l. Essa sensibilidade dos modelos à especificação de prioris indica uma possível falta de informação no conjunto de dados. Assim, seria interessante realizar outras análises com os modelos probito t-assimétricos usando conjuntos de dados que tenham tamanho amostral maiores do que os utilizados nesta dissertação.

Algumas propostas para futuros trabalhos são listadas abaixo:

- Aplicar o modelo probito t-assimétrico em outras áreas do conhecimento.
- Usar a priori de referência de Bernardo (ver Bernardo e Smith 1994), para o modelo probito t-assimétrico, pois neste caso somente as informações dos dados são consideradas.
- Coletar conjuntos de dados maiores, e fazer novas análises usando o modelo probito t-assimétrico.

Apêndice A

A distribuição Normal Assimétrica

A primeira abordagem da distribuição Normal Assimétrica (NA) no caso uniparamétrico foi dada por Azzalini (1985), subsequentemente Azzalini e Dalla Valle (1996) introduziram uma versão multivariada da densidade NA, enquanto Azzalini e Capitanio (1999) examinaram as propriedades probabilísticas da distribuição, e também investigaram os aspectos estatísticos mais relevantes.

O desenvolvimento dos artigos citados acima tem animado algumas pessoas a fazer aplicações deste tema. Segundo Azzalini e Dalla Valle (1996), Arnold *et al* (1993) usaram uma distribuição NA para uma aplicação em psicometria com dados reais, Copas e Li (1997) mostraram a conexão com o problema de seleção de uma amostra. Chen *et al* (1999) partiram de uma distribuição uniparamétrica NA e definiram uma função envolvendo assimetria generalizando modelos lineares para uma variável aleatória binomial. Sobre generalizações computacionais da classe de distribuição NA citaremos o artigo de Arnold e Beaver (2000b), no qual há também uma generalização para outras distribuições, como por exemplo, Cauchy-assimétrica multivariada.

A.1 Propriedades Matemáticas

Apresentamos nesta seção as principais características da distribuição Normal Assimétrica (NA), além de alguns exemplos gráficos para melhor compreensão de como se comporta esta distribuição.

Definição A.1 Se uma variável aleatória Z tem função de densidade

$$\phi_{NA}(z; \lambda) = 2\phi(z)\Phi(\lambda z) \quad (-\infty < z < \infty), \quad (\text{A.1})$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ representam a densidade normal padrão $N(0, 1)$ e sua função de distribuição acumulada, respectivamente, então dizemos que Z é uma variável aleatória normal assimétrica com parâmetro de assimetria λ , o qual varia $(-\infty < \lambda < \infty)$, ou seja, $Z \sim NA(\lambda)$.

Para a função densidade de probabilidade de uma normal assimétrica padrão usamos a notação $\phi_{NA}(z; \lambda)$ e para a sua distribuição acumulada usaremos a notação $F_\lambda(z)$.

Em seguida apresentamos algumas propriedades que seguem desta definição. Para maiores detalhes consultar Azzalini (1985).

Propriedade A: A densidade $NA(0)$ é a densidade $N(0, 1)$.

Prova: A demonstração desta propriedade é feita ao substituírmos o valor de λ por zero.

$$f(z) = 2\phi(z)\Phi(\lambda z)$$

$$f(z) = 2\phi(z)\Phi(0)$$

$$f(z) = 2\phi(z)1/2$$

$$f(z) = \phi(z).$$

■

Propriedade B: Quando $\lambda \rightarrow \infty$, $\phi(z, \lambda)$ converge para a densidade de uma normal truncada.

Prova:

$$f(z) = 2\phi(z)\Phi(\lambda z)$$

$$f(z) = 2\phi(z) \lim_{\lambda \rightarrow \infty} \Phi(\lambda z)$$

$$f(z) = 2\phi(z) \times 1$$

$$f(z) = 2\phi(z).$$

■

Propriedade C: Se Z é uma variável aleatória $NA(\lambda)$ então $-Z \sim NA(-\lambda)$.

Prova:

$$\begin{aligned} P(Y \leq y) &= P(-Z \leq y) = P(-Z \geq -y) = 1 - P(Z \leq -y) \\ &= 1 - F_z(-y). \end{aligned}$$

Assim para obter a densidade de Y , basta derivarmos $1 - F_z(-y)$, ou seja,

$$\begin{aligned} f(y) &= \frac{d}{dy}(1 - F_z(-y)) \\ &= -f_z(-y) \times (-1) \\ &= f_z(-y), \quad \text{onde} \end{aligned}$$

$$\begin{aligned} f_z(-y) &= 2\phi(-y)\Phi(-\lambda y) \\ &= 2\phi(y)\Phi(-\lambda y) \\ &\Rightarrow Y \sim NA(-\lambda). \end{aligned}$$

■

Como exemplo da densidade Normal Assimétrica temos as Figuras A.1 e A.2.

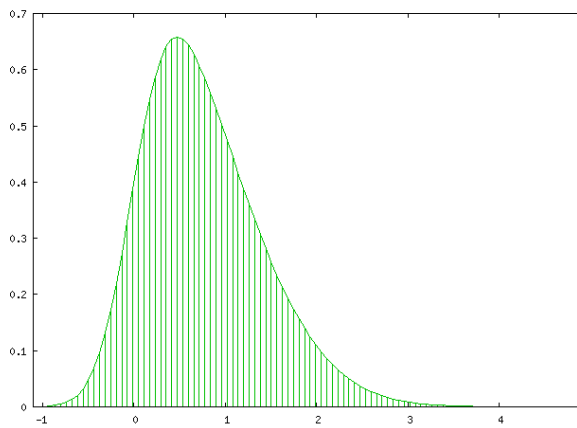


FIGURA A.1: $NA(3)$

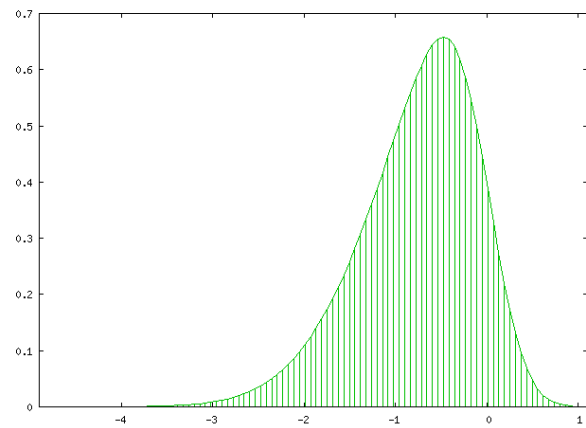


FIGURA A.2: $NA(-3)$

A Figura A.1 tem parâmetro de assimetria $\lambda = 3$ e possui uma distribuição assimétrica positiva ou à direita, pois a cauda desta distribuição está mais concentrada à direita do parâmetro de locação, que neste caso é zero. Diferentemente da Figura A.2 que possui parâmetro de assimetria $\lambda = -3$ e apresenta uma assimetria à esquerda.

Proposição A.1 A função de distribuição associada à densidade (A.1) é denotada por

$$F_\lambda(z) = 2\Phi_2(z, 0|\Omega), \quad \text{com } \Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}, \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}, \quad (\text{A.2})$$

$$(-\infty < z < \infty),$$

onde $\Phi_2(\cdot|\Omega)$ representa a função de distribuição de uma normal bivariada com vetor de médias nulo e matriz de covariância Ω .

Prova:

$$F_\lambda(z) = 2 \int_{-\infty}^{\infty} \phi(t)\Phi(\lambda t)dt = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\phi(u)dudt$$

Fazendo a mudança de variável $v = \frac{u-\lambda t}{\sqrt{1+\lambda^2}}$, temos que:

$$\begin{aligned} F_\lambda(z) &= 2\sqrt{1+\lambda^2} \int_{-\infty}^z \int_{-\infty}^0 \phi(t)\phi(v\sqrt{1+\lambda^2} + \lambda t)dvd t \\ &= 2 \int_{-\infty}^z \int_{-\infty}^0 \frac{\sqrt{1+\lambda^2}}{2\pi} \left\{ \exp \left[-\frac{1}{2} \left(t^2 + (v\sqrt{1+\lambda^2} + \lambda t)^2 \right) \right] \right\} dvd t \\ &= 2 \int_{-\infty}^z \int_{-\infty}^0 \frac{\sqrt{1+\lambda^2}}{2\pi} \\ &\quad \times \left\{ \exp -\frac{1}{2} \left(t^2(1+\lambda^2) + 2\lambda\sqrt{1+\lambda^2}vt + (1+\lambda^2)v^2 \right) \right\} dvd t \\ &= 2 \int_{-\infty}^z \int_{-\infty}^0 \frac{\sqrt{1+\lambda^2}}{2\pi} \\ &\quad \times \left\{ \exp \left[-\frac{1}{2} \left(\begin{bmatrix} t \\ v \end{bmatrix}^T \begin{bmatrix} 1+\lambda^2 & \lambda\sqrt{1+\lambda^2} \\ \lambda\sqrt{1+\lambda^2} & 1+\lambda^2 \end{bmatrix} \begin{bmatrix} t \\ v \end{bmatrix} \right) \right] \right\} dvd t \\ &= 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \frac{\sqrt{1+\lambda^2}}{2\pi} \\ &\quad \times \left\{ \exp \left[-\frac{1}{2} \left(\begin{bmatrix} t \\ v \end{bmatrix}^T \begin{bmatrix} 1 & \frac{-\lambda}{\sqrt{1+\lambda^2}} \\ \frac{-\lambda}{\sqrt{1+\lambda^2}} & 1 \end{bmatrix}^{-1} \begin{bmatrix} t \\ v \end{bmatrix} \right) \right] \right\} dvd t \end{aligned}$$

$$\begin{aligned}
&= 2 \int_{-\infty}^z \int_{-\infty}^0 \frac{\sqrt{1+\lambda^2}}{2\pi} \\
&\quad \times \left\{ \exp \left[-\frac{1}{2} \left(\begin{bmatrix} t \\ v \end{bmatrix}^T \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}^{-1} \begin{bmatrix} t \\ v \end{bmatrix} \right) \right] \right\} dv dt \\
&= 2 \int_{-\infty}^z \int_{-\infty}^0 \phi_2(t, v|\Omega) dv dt \\
&= 2\Phi_2(z, 0|\Omega).
\end{aligned}$$

■

Azzalini utilizou as propriedades da função $T(h, a)$ estudada por Owen (1956), para demonstrar algumas das propriedades da distribuição normal assimétrica. Para $h > 0$ e $a > 0$, a função $T(h, a)$ é dada por:

$$T(h, a) = \int_h^\infty \int_0^{ax} \phi(x)\phi(y)dydx.$$

Note que

$$\begin{aligned}
F_\lambda(z) &= 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\phi(u)dudt \\
&= 2 \int_{-\infty}^z \int_{-\infty}^0 \phi(t)\phi(u)dudt - 2 \int_z^\infty \int_0^{\lambda t} \phi(t)\phi(u)dudt \\
&= \Phi(z) - 2T(z, \lambda).
\end{aligned}$$

Logo,

$$F_\lambda(z) = \Phi(z) - 2T(z, \lambda). \quad (\text{A.3})$$

$T(h, a)$ é uma função decrescente de h , pois se $h_2 > h_1 > 0$,

$$T(h_2, a) = \int_{h_2}^\infty \int_0^{ax} \phi(x)\phi(y)dydx < \int_{h_1}^\infty \int_0^{ax} \phi(x)\phi(y)dydx = T(h_1, a). \quad (\text{A.4})$$

Temos ainda,

$$-T(h, a) = - \int_h^\infty \int_0^{ax} \phi(x)\phi(y)dydx = \int_h^\infty \int_0^{-ax} \phi(x)\phi(y)dydx = T(h, -a). \quad (\text{A.5})$$

$$T(-h, a) = \int_{-\infty}^{-h} \int_0^{ax} \phi(x)\phi(y)dydx = \int_h^\infty \int_0^{ax} \phi(x)\phi(y)dydx = T(h, a), \quad (\text{A.6})$$

$$2T(h, 1) = \Phi(h)\Phi(-h). \quad (\text{A.7})$$

Da propriedade C e das propriedades da função $T(h, a)$, nós obtemos as seguintes propriedades:

Propriedade D: $1 - F_\lambda(-z) = F_\lambda(-z)$.

Prova: Por (A.3), (A.5) e (A.6) temos que,

$$\begin{aligned} 1 - F_\lambda(-z) &= 1 - [\Phi(-z) - 2T(-z, \lambda)] \\ &= 1 - [\Phi(-z) - 2T(z, \lambda)] \\ &= 1 - \Phi(-z) + 2T(z, \lambda) \\ &= 1 - \Phi(-z) - 2[-T(z, \lambda)] \\ &= \Phi(z) - 2T(z, -\lambda) \\ &= F_\lambda(-z). \end{aligned} \quad \blacksquare$$

Propriedade E: $F_1(z) = \{\Phi(z)\}^2$.

Prova: Por (A.3) e (A.7) temos:

$$\begin{aligned} F_1(z) &= \Phi(z) - 2T(z, 1) \\ &= \Phi(z) - \Phi(z)\Phi(-z) \\ &= \Phi(z) [1 - \Phi(-z)] \\ &= \Phi(z)\Phi(z) \\ &= [\Phi(z)]^2. \end{aligned} \quad \blacksquare$$

Propriedade F: Se $Z \sim NA(\lambda)$ então $Z^2 \sim \chi_1^2$.

Prova: Seja $Y = Z^2$, onde $Z \sim NA(\lambda)$

$$\begin{aligned}
P(Y \leq y) &= P(Z^2 \leq y) = P(-y^{\frac{1}{2}} \leq Z \leq y^{\frac{1}{2}}) \\
&= F_z(y^{\frac{1}{2}}) - F_z(-y^{\frac{1}{2}}) \\
\frac{\partial z}{\partial y} &= f_z(y^{\frac{1}{2}}) \frac{1}{2\sqrt{y}} - f_z(-y^{\frac{1}{2}}) (-1) \frac{1}{2\sqrt{y}} \\
&= \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{y}{2}} \Phi(\lambda y^{\frac{1}{2}}) + \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{y}{2}} \Phi(-\lambda y^{\frac{1}{2}}) \\
&= \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} \left[\Phi(\lambda y^{\frac{1}{2}}) + \Phi(-\lambda y^{\frac{1}{2}})\right] \\
&= \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} \left[\Phi(\lambda y^{\frac{1}{2}}) + 1 - \Phi(\lambda y^{\frac{1}{2}})\right] \\
&= \frac{1}{\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} \\
&\Rightarrow Z \sim X_1^2.
\end{aligned}$$

■

Vamos estender o modelo (A.1) introduzindo o parâmetro de posição $\mu \in R$ e o parâmetro de escala $\sigma^2 > 0$. Neste caso diremos que $Y \sim NA(\mu, \sigma^2, \lambda)$. Para a função densidade e de distribuição de Y usamos a notação $\phi_{NA}(y; \mu, \sigma^2, \lambda)$, $F_\theta(y)$ em que θ representa os parâmetros da distribuição normal assimétrica, $\theta = (\mu, \sigma^2, \lambda)$.

Definição A.2 *Uma variável aleatória Y tem distribuição assimétrica com parâmetros de localização μ e de escala σ^2 se sua função densidade de probabilidade é da forma:*

$$\phi_{NA}(y; \mu, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad (-\infty < y < \infty). \quad (\text{A.8})$$

onde $\theta = (\mu, \sigma^2, \lambda)$.

Para provar o resultado dado pela equação (A.8) basta fazer a transformação de variáveis $Y = \mu + \sigma Z$ onde $Z \sim NA(\lambda)$.

Mostramos na Figura A.3 o gráfico da densidade de uma distribuição Normal Assimétrica com parâmetros $(\mu = 5, \sigma^2 = 4, \lambda = 3)$.

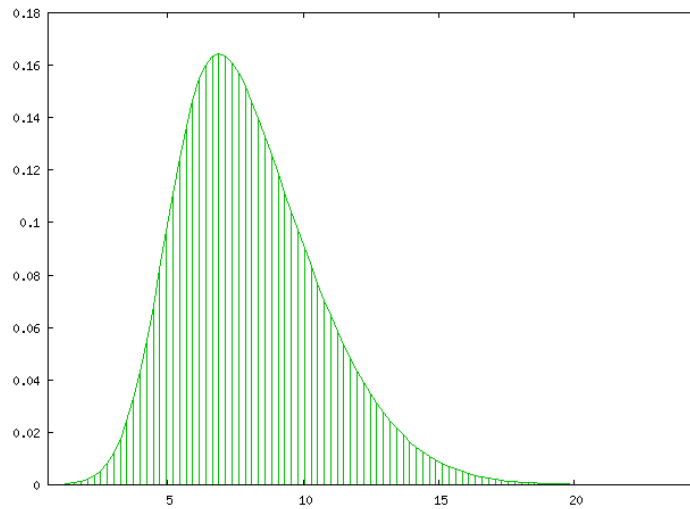


FIGURA A.3: NA(5,4,3).

Proposição A.2 A função de distribuição de (A.8) é dada por:

$$F_{\theta}(y) = 2\Phi_2\left(\frac{y-\mu}{\sigma}, 0 \mid \Omega\right), \text{ com } \Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}, \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}},$$

$$(-\infty < z < \infty), (\mu, \sigma^2, \lambda).$$

sendo $\Phi_2(\cdot \mid \Omega)$ função de distribuição da normal bivariada com média zero e matriz de covariância Ω .

Prova: Consideremos $Y = \mu + \sigma Z$ e $Z \sim NA(\lambda)$, assim

$$\begin{aligned} P(Y \leq y) &= P(\mu + \sigma Z \leq y) = P\left(Z \leq \frac{y-\mu}{\sigma}\right) \\ &= F_z\left(\frac{y-\mu}{\sigma}, \lambda\right) \\ &\Rightarrow Y \sim NA(\mu, \sigma^2, \lambda). \end{aligned}$$

■

A.2 Obtendo a Distribuição Normal Assimétrica

Existem algumas maneiras de se obter a distribuição Normal Assimétrica utilizando outras distribuições. Descrevemos nesta seção os resultados mais relevantes para este trabalho.

Uma forma de apresentar a distribuição normal assimétrica é através da representação estocástica, à qual foi obtida por Henze (1986) e cujo resultado é dado abaixo:

Proposição A.3 *Sejam Y_1 uma variável aleatória com distribuição normal padrão positiva e Y_2 variável aleatória com distribuição normal padrão. Considerando Y_1 e Y_2 variáveis aleatórias independentes temos:*

$$Z = \delta Y_1 + \sqrt{1 - \delta^2} Y_2 \sim NA(\lambda), \quad \text{onde } \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Prova: Definindo a função geradora de momentos da variável aleatória Z como $M_Z(t)$, temos:

$$\begin{aligned} M_z(t) &= E(e^{t(\delta Y_1 + \sqrt{1 - \delta^2} Y_2)}) = E(e^{t\delta Y_1}) E(e^{(\sqrt{1 - \delta^2} Y_2)t}) \\ &= M_{Y_2}(\sqrt{1 - \delta^2} t) E(e^{t\delta Y_1}) \\ &= e^{(1 - \delta^2)t^2/2} \int_0^\infty \frac{2}{\sqrt{2\pi}} e^{\delta t y_1} e^{-\frac{1}{2} y_1^2} dy_1 \\ &= 2e^{t^2/2} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} [y_1^2 - 2\delta t y_1 + \delta^2 t^2]} dy_1 \\ &= 2e^{t^2/2} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} [(y_1 - \delta t)^2]} dy_1. \end{aligned}$$

Fazendo a mudança de variável $x = y_1 - \delta t$, obtemos:

$$\begin{aligned} E(e^{tZ}) &= 2e^{t^2/2} \int_{-\delta t}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} dx \\ &= 2e^{t^2/2} (1 - \Phi(-\delta t)) \\ &= 2e^{t^2/2} \Phi(\delta t), \end{aligned}$$

De acordo com a Proposição A.7 podemos concluir que $Z \sim NA(\lambda)$. ■

Proposição A.4 *Considerando $Z \sim NA(\theta)$, e $V \sim N(0, 1)I(v > 0)$, com $\theta = (\mu, 1, \lambda)$, a f.d.a de Z pode ser escrita como:*

$$F_\theta = \int_0^\infty \Phi\left(\frac{z - \mu - \delta v}{\sqrt{1 - \delta^2}}\right) f_V(v) dv \quad (\text{A.9})$$

onde $f_V(\cdot)$ representa a f.d.p de V .

Prova:

Por condicionamento, temos:

$Z|(V = v) \sim N(\mu + \delta v, 1 - \delta^2)$, assim a densidade de Z pode ser escrita como

$$\begin{aligned} f_\theta(z) &= \int_0^\infty \phi\left(\frac{z - \mu - \delta v}{\sqrt{1 - \delta^2}}\right) f_V(v) dv \\ F_\theta(z) &= \int_{-\infty}^z \int_0^\infty \phi\left(\frac{u - \mu - \delta v}{\sqrt{1 - \delta^2}}\right) f_V(v) dv du \\ &= \int_0^\infty \left[\int_{-\infty}^z \phi\left(\frac{u - \mu - \delta v}{\sqrt{1 - \delta^2}}\right) du \right] f_V(v) dv. \\ \Rightarrow F_\theta(z) &= \int_0^\infty \Phi\left(\frac{z - \mu - \delta v}{\sqrt{1 - \delta^2}}\right) f_V(v) dv. \end{aligned}$$

■

Proposição A.5 Se $(Y_1, Y_2) \sim N_2$ com densidades marginais padronizadas e correlação δ , então a distribuição condicional de $Y|(X > 0)$ é $NA(\lambda)$.

Prova: Devemos mostrar que $Z = Y_2|(Y_1 > 0) \sim NA(\lambda)$

$$P(Y_2|Y_1 > 0) = \frac{P(Y_1 > 0|Y_2 = z)f_{Y_2}(z)}{P(Y_1 > 0)}.$$

Utilizando as propriedades da distribuição normal, temos que:

$$\begin{aligned} P(Y_1 > 0) &= \frac{1}{2}, \\ Y_1|(Y_2 = z) &\sim N(\delta z, 1 - \delta^2), \\ P(Y_1 > 0|Y_2 = z) &= \Phi\left(\frac{\delta z}{\sqrt{1 - \delta^2}}\right), \end{aligned}$$

$$\text{logo } P(Y_2|Y_1 > 0) = 2\phi_{Y_2}(z)\Phi\left(\frac{\delta z}{\sqrt{1 - \delta^2}}\right), \quad \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Logo, $Z \sim NA(\lambda)$.

■

Proposição A.6 Se Y e W são variáveis aleatórias $N(0,1)$ e independentes. Z é igual a Y desde que $\lambda y > W$, para algum real λ , então $Z \sim NA(\lambda)$,

$$Z = \begin{cases} Y & \text{se } \lambda y > W, \\ -Y & \text{se } \lambda y \leq W. \end{cases}$$

Prova: Queremos mostrar que $Z = [Y|\lambda Y > W] \sim NA(\lambda)$

$$\begin{aligned}
 P(Z \leq z) &= P(Y \leq z | \lambda y > W) = \frac{P(Y \leq z, \lambda y > W)}{P(\lambda y > W)} \\
 &= \frac{1}{P(\lambda y > W)} \int_{-\infty}^z \int_{-\infty}^{\lambda y} \phi(y)\phi(w)dw dy \\
 &= 2 \int_{-\infty}^z \phi(y)\Phi(\lambda y)dy \\
 \Rightarrow f_z(z) &= 2\phi(z)\Phi(\lambda z), \quad \text{logo } Z \sim NA(\lambda).
 \end{aligned}$$

■

A.3 Momentos

Nesta seção discutimos como encontrar os momentos da distribuição normal assimétrica.

Segundo Azzalini (1985), devido à propriedade F, dada na Seção A.1 temos que os momentos pares de Z são iguais aos momentos pares de uma distribuição normal padrão. Para maiores detalhes sobre os momentos desta distribuição consultar Rodríguez (2005). Para calcular os momentos ímpares usaremos o próximo resultado:

Lema A.1 *Se U é uma variável aleatória onde $U \sim N(0, 1)$, então*

$$E \{ \Phi(hU + k) \} = \Phi \left\{ \frac{k}{\sqrt{1 + h^2}} \right\}, \quad \text{para qualquer real } h, k.$$

A função geradora de momentos de uma variável aleatória com distribuição normal assimétrica é dada pela proposição abaixo.

Proposição A.7 *A função que gera os momentos de Z , em que $Z \sim NA(\lambda)$ é dada por:*

$$M_z(t) = 2e^{(t^2/2)}\Phi(\delta t), \quad \text{onde } \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Prova:

$$\begin{aligned}
 M_z(t) &= E(e^{tZ}) = 2 \int_{-\infty}^{\infty} e^{tZ} \phi(z) \Phi(\lambda z) dz \\
 &= 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(t^2/2) \exp -\frac{1}{2}(z^2 - 2tz + t^2) \Phi(\lambda z) dz \\
 &= 2 \exp(t^2/2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(z - t)^2 \Phi(\lambda z) dz.
 \end{aligned}$$

Fazendo $x = z - t$, temos:

$$\begin{aligned}
 M_z(t) &= 2 \exp(t^2/2) \int_{-\infty}^{\infty} \phi(x) \Phi(\lambda t + \lambda x) dx \\
 &= 2 \exp(t^2/2) E \{ \Phi(\lambda t + \lambda x) \} \\
 &= 2 \exp(t^2/2) \Phi \left(\frac{\lambda}{\sqrt{1 + \lambda^2}} \right) \\
 &= 2 \exp(t^2/2) \Phi(\delta t).
 \end{aligned}$$

■

Como conhecemos a função geradora de momentos podemos calcular $E(Z)$ e $Var(Z)$ onde $Z \sim NA(\lambda)$.

$$\begin{aligned}
 E(Z) &= M_z(t)|_{t=0} = 2\delta\phi(0) = 2 \frac{\lambda}{\sqrt{1 + \lambda^2}} \frac{1}{\sqrt{2\pi}} e^0 \\
 E(Z) &= \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}}.
 \end{aligned}$$

Fazendo os cálculos, obtemos que a variância da distribuição normal assimétrica padrão é dada por:

$$Var(Z) = 1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2}.$$

Proposição A.8 A função geradora de momentos de $Y \sim NA(\mu, \sigma^2, \lambda)$ é dada por:

$$M_Y(t) = 2e^{t\mu + \frac{t^2\sigma^2}{2}} \Phi(\delta\sigma t), \quad \text{com } \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Prova:

$$M_Y(t) = E(e^{tY}) = \int_{-\infty}^{\infty} e^{ty} \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right) dy.$$

Fazendo a mudança de variável $z = \frac{y-\mu}{\sigma}$, obtemos:

$$\begin{aligned} M_Y(t) &= \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)} \frac{2}{\sigma} \phi(z) \Phi(\lambda z) \sigma dz \\ &= e^{t\mu} \int_{-\infty}^{\infty} 2e^{(t\sigma)z} \phi(z) \Phi(\lambda z) dz = e^{t\mu} M_z(\sigma t) \\ &= e^{t\mu} 2e^{(\sigma t)^2/2} \Phi(\lambda \sigma t). \end{aligned}$$

■

Podemos também calcular a média e a variância de uma variável aleatória $Y \sim NA(\mu, \sigma^2, \lambda)$, e são dadas por:

$$E(Y) = \mu + \sigma \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}}, \quad \text{Var}(Y) = \sigma^2 \left(1 - \frac{2}{\pi} \frac{\lambda^2}{1+\lambda^2}\right).$$

Neste Apêndice comenta-se apenas o que é necessário para o entendimento do texto em questão a respeito de distribuições normais assimétricas. Para um estudos mais profundo sobre esta distribuição, inclusive no caso multivariado, consultar Azzalini (1985), Azzalini e Dalla Valle (1996), disponível em <http://azzalini.stat.unipd.it/SN/>.

Apêndice B

Distribuições esféricas e elípticas

Segundo Valle (1994), a partir da década de 70, as distribuições esféricas e elípticas vêm ocupando um importante lugar na teoria estatística. As distribuições esféricas são caracterizadas por sua invariância com respeito a transformações ortogonais e formam uma classe geral de distribuições com a mesma simetria esférica da distribuição normal padrão. As distribuições elípticas podem ser geradas através de transformações lineares de locação e escala das distribuições esféricas, o que implica que as distribuições elípticas formam uma classe generalizada de famílias paramétricas de distribuições de probabilidade que preservam a estrutura simétrica das distribuições normais.

Neste capítulo, mostramos alguns resultados importantes para a realização deste trabalho.

B.1 As distribuições Esféricas Multivariadas

Definição B.1 *Sejam Z_1, \dots, Z_d , variáveis aleatórias independentes e identicamente distribuídas, considera-se o vetor d -dimensional $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_d)'$, que tem sua função distribuição de probabilidade dada por $f_z(z)$. Se \mathbf{Z} satisfizer:*

$$\mathbf{Z} \stackrel{d}{=} \Gamma \mathbf{Z}, \quad \forall \Gamma : \Gamma \Gamma' = I_d \quad (\text{B.1})$$

onde $X \stackrel{d}{=} Y$ denota que X e Y tem as mesmas distribuições de probabilidade e Γ pertence ao conjunto de matrizes que são ortogonais, isto é, $\Gamma \Gamma' = I_d$.

A propriedade dada por (B.1) impõe a simetria esférica e está relacionada com a forma da função distribuição de probabilidade de z , isto é,

$$f_d(z) = g(z'z), \quad z \in \mathbb{R}^d,$$

indicando que esta densidade depende de z através do quadrado de seu cumulante, isto é, $\|z\|^2 = z'z$.

Podemos equivalentemente definir as distribuições esféricas pelo resultado a seguir:

Definição B.2 *O vetor aleatório Z , como é dado na definição anterior, tem distribuição esférica, se e somente se, podemos representá-lo como*

$$\mathbf{Z} = \mathbf{R}\mathbf{U}, \tag{B.2}$$

onde \mathbf{U} é um vetor uniforme em $S_d = \{z \in \mathbb{R}^d : \|z\| = 1\}$ independente da variável não negativa R , denominada radial.

A densidade de Z , se existir é dada por:

$$f_Z(z) = g^{(d)}(z'z), \quad z \in \mathbb{R}^d,$$

em que $g^{(d)}$ é chamada a função geradora e d é a dimensão desta.

Mostramos em seguida um exemplo de uma distribuição esférica.

Exemplo B.1 *Se considerarmos Z_1, \dots, Z_d , variáveis aleatórias independentes e identicamente distribuídas $N(0, 1)$, então o vetor d -dimensional $Z = (Z_1, \dots, Z_d)'$ tem função distribuição de probabilidades dada por:*

$$\begin{aligned} f(Z_1, \dots, Z_d) &= (2\pi)^{-(d/2)} e^{-\frac{1}{2} \sum_{i=1}^d z_i^2} \\ &= (2\pi)^{-(d/2)} e^{-\frac{1}{2} z'z}, \quad z \in \mathbb{R}^d, \end{aligned}$$

note que a distribuição normal é invariante por transformações ortogonais, logo a distribuição normal pertence à classe das distribuições esféricas.

B.2 As distribuições Elípticas Multivariadas

As distribuições elípticas multivariadas são obtidas através de uma transformação linear das distribuições esféricas multivariadas. Deste fato segue a próxima definição:

Definição B.3 Um vetor aleatório Y , k -dimensional, tem distribuição elíptica com vetor de posição k -dimensional μ e matriz de escala σ , de dimensão $k \times k$, $k \leq d$, se $Y = \mu + AZ$, onde Z tem distribuição esférica d -variada e Ω é uma matriz $k \times k$ de posto k tal que $AA' = \Omega$.

Quando $k = d$ e se Z tem densidade $g^{(d)}(Z'Z)$ então a função distribuição de probabilidade de Y é dada por

$$f_Y(y) = |\Omega|^{1/2} g^{(d)}[(y - \mu)' \Omega^{-1} (y - \mu)], \quad y \in \mathbb{R}^d.$$

E diremos que $Y \sim El_d(\mu, \sigma; g^{(d)})$, onde a função geradora satisfaz

$$\int_0^\infty u^{d-1} g^{(d)}(u) du = \frac{\Gamma(d/2)}{\pi^{d-2}}.$$

Exemplo B.2 Considere $Y \sim El_d(\mu, \Omega; g^{(d)})$, a distribuição t -multivariada pertence à classe de distribuições elípticas, pois

$$f_Y(y) = \frac{\Gamma((v+d)/2)}{\Gamma(v/2)(\pi v)^{d/2}} \Omega^{-1/2} [v + (y - \mu)' \Omega^{-1} (y - \mu)]^{-\frac{v+d}{2}}, \quad y \in \mathbb{R}^d,$$

onde μ, Ω e v representam respectivamente os parâmetros de posição, escala e graus de liberdade.

Apresentamos a seguir algumas propriedades das distribuições elípticas.

1. Se existem os momentos da distribuição, então $E(Y) = \mu$ e $Var(Y) = \alpha_g \Omega$, onde $\alpha_g = E(R^2)$, em que R é a variável radial e Ω é o parâmetro de escala.
2. Se $Y \sim El_d(\mu, \Omega; g^{(d)})$, A é uma matriz $k \times d$ e b um vetor k -dimensional, então $(AY + b) \sim El_k(A\mu + b, A\Omega A^T; g^{(k)})$.
3. Ao particionarmos $Y = (Y_1^t, Y_2^t)^t$, onde Y_1 um vetor de dimensão d_2 com $d = d_1 + d_2$, segue que:

$$Y_i \sim El_{d_i}(\mu_i, \Omega_{ii}; g^{(d_i)}), \quad i = 1, 2 \text{ onde}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

E os geradores marginais $g^{(d_i)}$ $i = 1, 2$, podem ser obtidos a partir da expressão:

$$g^{(k)}(u) = \frac{\pi^{\frac{d-k}{2}}}{\Gamma(\frac{d-k}{2})} \int_0^\infty v^{\frac{d-k}{2}} g^{(d)}(u+v) dv, \quad k = 1, \dots, d-1.$$

4. $Y_1 | (Y_2 = y_2) \sim El_{d_2}(\mu_{1.2}, \Omega_{11.2}; g_q(y))$, onde

$$\begin{aligned}\mu_{1.2} &= \mu + \Omega_{12}\Omega_{22}^{-1}(y_2 - \mu_2) \\ \Omega_{11.2} &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \\ q(y_2) &= (y_2 - \mu_2)^t \Omega_{22}^{-1} (y_2 - \mu_2)\end{aligned}$$

e a função geradora condicional é:

$$g_a^{(d_1)}(u) = \frac{g^{(d)}(u + a)}{g^{(d_2)}(a)}.$$

Com estes resultados podemos dizer que a classe de distribuições elípticas é fechada por marginalização e condicionamento.

As demonstrações destas propriedades e maiores detalhes podem ser vistos no texto de Branco e Valle (2004).

B.3 A distribuição Elíptica Assimétrica

Para Branco e Valle (2004) o fato das distribuições elípticas manterem as propriedades associadas à família de distribuições normais, como por exemplo, ser simétrica e fechada por marginalização e condicionamento, levou-os a pensar numa extensão natural de substituir a suposição de normalidade pela de elipicidade na construção da classe assimétrica. Sabemos de acordo com o que foi escrito no Apêndice A, que a distribuição normal Assimétrica pode ser obtida pelo método do condicionamento. A distribuição elíptica assimétrica será definida de maneira similar.

Considere $(X_1, X_2) \sim El_2(0, \Omega; g^{(2)})$ onde $\Omega = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$ com $|\delta| < 1$ e $g^{(2)}$ é a função geradora de uma distribuição elíptica bivariada. Utilizando as propriedades das distribuições elípticas podemos dizer que

$X_2 \sim El(0, 1; g^{(1)})$ e $X_1 | (X_2 = z) \sim El(\delta z, 1 - \delta^2; g_{q(z)}^{(1)})$, com $q(z) = z^2$ e

$$g_{q(z)}^{(1)}(u) = \frac{g^{(2)}(u + z^2)}{g^{(1)}(z^2)}.$$

Definindo $Z = X_2 | X_1 > 0$, obtemos:

$$\begin{aligned}f_Z(z) &= \frac{P(X_1 > 0 | X_2 = z) f_{X_2}(z)}{P(X_1 > 0)} \\ \text{isto é, } f_Z(z) &= 2f_{g^{(1)}}(z) F_{g_{q(z)}^{(1)}}(\lambda z)\end{aligned}\tag{B.3}$$

de acordo com o resultado dado pela equação (B.3) podemos formalizar o conceito de distribuições elípticas assimétricas.

Definição B.4 *Uma variável aleatória Y tem distribuição elíptica assimétrica com parâmetro de posição μ , escala σ^2 e assimetria λ se $Y = \mu + \sigma Z$, onde Z tem função densidade dada por (B.3). Utilizamos a notação $Y \sim SE(\mu, \sigma^2, \lambda; g^{(2)})$.*

B.3.1 Propriedades

Considere $Y \sim SE(\mu, \sigma^2, \lambda; g^{(2)})$, mostramos a seguir algumas propriedades da distribuição elíptica, segundo Branco e Valle (2004).

1. $Y_2 = a + bY_1 \sim ES(a + b\mu, b^2\sigma^2; g^{(2)})$;
2. Ao considerar a representação da variável elíptica simétrica $X_1 = RU$, como em B.2, $\frac{(Y - \mu)^2}{\sigma^2}$ tem a mesma distribuição de R^2 , ou seja, a distribuição dessa forma quadrática coincide com a forma quadrática da distribuição elíptica simétrica.

B.3.2 Obtenção da distribuição t-Assimétrica

Podemos obter uma subclasse importante da família de distribuições elípticas assimétricas que é a distribuição t-assimétrica ao considerarmos a função geradora dada abaixo

$$g(u) = \frac{\Gamma(\frac{\nu+1}{2})\nu^{\nu/2}}{\Gamma(\frac{\nu}{2})\sqrt{\pi}}(\nu + u)^{-\frac{\nu+1}{2}}. \quad (\text{B.4})$$

Com o uso da função geradora (B.4), os resultados da Seção B.3 e o resultado dado pela equação (B.3), obtemos:

$$f_Z(z) = 2t_\nu(z)T_{\nu+1} \left[\lambda z \sqrt{\frac{\nu+1}{\nu+z^2}} \right], \quad (\text{B.5})$$

onde t_ν representa a f.d.p de uma distribuição t-Student com ν graus de liberdades e $T_{\nu+1}$ representa a f.d.a de uma distribuição t-Student com $\nu + 1$ graus de liberdades e Z representa uma distribuição t-assimétrica com parâmetro de assimetria λ e ν graus de liberdades.

Apêndice C

Programas

C.1 Programa em R para o modelo PN

```
library(msm)
library(mvtnorm)
y=dados # Ver dados para Aplicação 1 na Seção C.3
I=14
n=131
niter= 11 # número de iterações
# Valores iniciais
u=rnorm(n,0,1)
a=rep(1, I)
b=rep(0, I)
c=rep(-1,n)
# Matrizes
Z=matrix(NA,nr=n, nc=I)
m=matrix(NA,nr=n, nc=I)
parc1=matrix(NA,nr=n, nc=I)
media.beta=matrix(NA,nr=2, nc=I)
mu.beta=matrix(c(1, 0), nr=2)
Sigma.beta=matrix(c(0.5,0,0,2),nr=2)
beta.m=c(a,b)
```

```

beta=matrix(beta.m, nc=2)
cadeia.beta=beta
cadeia.u=u A2= solve(Sigma.beta) #calcula a inversa de Sigma.beta
for(l in 1: niter){
W=matrix(c(u,c),nc=2)
A1=t(W)%*%W
inversa.A= solve(A1+A2)
# gerando Z
for(j in 1: I){
for (i in 1: n){
m[i,j]=beta[j,1]*u[i]-beta[j,2]
B2= A2%*%mu.beta
variancia.beta= inversa.A
if(y[i,j]==0)
{
Z[i,j]= rtnorm(1,m[i,j], 1, upper=0)
}else
{
Z[i,j]= rtnorm(1,m[i,j],1, lower=0)
}
}}
# Gerando beta
for(j in 1: I){
media.beta[j]= inversa.A%*% (t(W)%*%Z[,j]+B2)
beta[j,]= rmvnorm(1, c(media.beta[j]), variancia.beta)
}
# Gerando u
for (i in 1: n){
for(j in 1:I){
parc1[i, j]= (beta[j,1]*(beta[j,2] + Z[i,j]))/(1 + sum(beta[,1]^2))
} }
var.u= 1/ (1 + sum(beta[,1]^2))

```

```

med.u= apply(parc1,1,sum)
for(i in 1:n){
u[i]=rnorm(1, med.u[i], sqrt(var.u))
}
# cadeias
cadeia.beta=rbind(cadeia.beta, beta)
cadeia.u=rbind(cadeia.u, u)
cat("\n", 1)
}

```

C.2 Programa t-assimétrico

```

model{
for(i in 1:n){
u[i]~dnorm(med[i],tau2) # Habilidade NA
med[i]<- eta*x[i]
x[i]~dnorm(0,1)I(0,)

      for(j in 1:I){
V[i,j]<- 1/iv[i,j]
m[i,j]<- a[j]*u[i]-b[j]
muz[i,j]<- m[i,j]-d[j]*R[i,j]*sqrt(V[i,j])
preczs[i,j]<- 1/V[i,j]*(1-pow(d[j],2))
zs[i,j]~dnorm(muz[i,j],preczs[i,j])I(lo[y[i,j]+1],up[y[i,j]+1])
iv[i,j]~dgamma(3,3) #nu=6
R[i,j]~dnorm(0,1)I(0,)
cpo[i,j]<- sqrt(2*3.14159265/preczs[i,j])*exp(0.5*preczs[i,j]*pow(zs[i,j]-muz[i,j], 2))
resid[i,j]<- zs[i,j]-muz[i,j]
res2[i,j]<- pow(resid[i,j],2)
}}
for(j in 1:I){

```

```

a[j]~dnorm(1,2)I(0,)
b[j]~dnorm(0,0.5)
d[j]~dunif(-1,1)
lambda[j]<-d[j]/sqrt(1-pow(d[j],2))
resmean[j]<-sum(res2[,j])
}
lo[1]<- -50
lo[2]<- 0
up[1]<- 0
up[2]<- 50
mu<- mean(u[])
du<- sd(u[])
sse<- sum(resmean[]) #SQRL
eta<- k/sqrt(1+pow(k,2))
tau<- 1/sqrt(1+pow(k,2))
tau2<- (1+pow(k,2))
k~dnorm(0,pp)
pp<- 0.4052847* p
p~dgamma(0.25,0.25)
}
# Dados para Aplicação 2
list(y=structure(.Data=c(0, 1, 0, 1, 1, 1, 0, 1, 1, 1,
1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 0, 1, 0, 1,
0, 1, 1, 1, 1, 1, 1, 0, 0, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 1, 0, 1, 0, 0, 0, 0,
1, 1, 1, 1, 1, 1, 1, 0, 1, 1,
0, 0, 0, 0, 1, 1, 1, 1, 0, 1,
1, 0, 1, 0, 1, 1, 1, 0, 0, 1,
0, 1, 1, 1, 0, 1, 0, 0, 0, 1,
1, 1, 1, 1, 1, 1, 0, 1, 0, 0,

```



```

1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 0, 1, 0, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 1, 1, 1, 0, 1, 0, 1, 0, 0,
0, 1, 0, 1, 1, 1, 0, 1, 0, 0,
0,1, 1, 1, 0, 1, 0, 1, 0, 1 ),.Dim=c(18,10)),n=18,I=10)
# Valores Iniciais
list(b=c(0,0,0,0,0,0,0,0,0,0),a=c(1,1,1,1,1,1,1,1,1,1),d=c(0,0,0,0,0,0,0,0,0,0))
list(b=c(1,1,1,1,1,1,1,1,1,1),a=c(0,0,0,0,0,0,0,0,0,0),d=c(0,0,0,0,0,0,0,0,0,0))

```

C.3 Programa para a família TRI-NA

Modelo TRI- NA (PANA)

```

model{
for(i in 1:n){
u[i]~dnorm(med[i],tau2)
med[i]<-eta*x[i]
x[i]~dnorm(0,1)I(0,)
for(j in 1:I){
m[i,j]<-a[j]*u[i]-b[j]
muz[i,j]<-m[i,j]-d[j]*V[i,j]
zs[i,j]~dnorm(muz[i,j],preczs[j])I(lo[y[i,j]+1],up[y[i,j]+1])
V[i,j]~dnorm(0,1)I(0,)
resid[i,j]<-zs[i,j]-muz[i,j]
res2[i,j]<-pow(resid[i,j],2)
cpo[i,j]<-sqrt(2*3.14159265/preczs[j])*exp(0.5*preczs[j]*pow(zs[i,j]-muz[i,j], 2))
}}
for(j in 1:I){
a[j]~dnorm(0,1)I(0,) #proiri A
b[j]~dnorm(0, 0.0001)

```

```

a[j]~dnorm(1,2)I(0,) #proiri B
b[j]~dnorm(0, 0.5)
d[j]~dunif(-1,1)
preczs[j]<-1/(1-pow(d[j],2))
lambda[j]<-d[j]*sqrt(preczs[j])
resmean[j]<-sum(res2[,j])
}
lo[1]<- -50
lo[2]<- 0
up[1]<- 0
up[2]<- 50
mu<-mean(u[])
du<-sd(u[])
sse<-sum(resmean[]) #SQRL
eta<-k/sqrt(1+pow(k,2))
tau<-1/sqrt(1+pow(k,2))
tau2<-(1+pow(k,2))
k~dnorm(0,pp)
pp<-0.4052847* p
p~dgamma(0.25,0.25)
}
#Dados para Aplicação 1
list(y=structure(.Data=c(1,1,0,1,1,0,1,1,1,1,1,0,0,1,
1,1,1,0,1,0,0,1,0,1,0,0,0,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,
0,1,1,1,1,1,1,1,0,1,1,1,1,1,1,
1,0,0,1,1,0,0,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,0,1,1,1,1,
0,1,1,1,1,0,1,0,1,1,1,1,0,1,
1,0,0,1,1,0,1,1,1,1,1,0,0,1,

```

1,1,0,1,1,0,1,1,0,1,1,1,1,1,
1,0,0,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,1,0,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,0,1,1,
1,0,0,1,1,0,1,0,1,1,0,0,1,0,
1,0,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,0,0,1,1,1,1,1,1,1,
1,0,1,1,1,0,1,0,1,0,1,0,1,1,
1,1,0,1,0,1,1,1,1,1,1,1,1,
1,0,0,1,1,1,0,1,0,1,0,0,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,1,0,1,1,1,0,1,0,1,1,
1,1,0,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,1,1,1,0,0,
0,1,1,1,1,1,1,1,1,1,1,1,1,
1,0,0,1,0,0,1,1,1,1,1,1,1,
0,1,0,1,0,0,1,1,0,1,1,0,1,1,
1,0,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,0,1,1,0,1,1,
1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,0,1,1,0,1,1,
1,1,1,1,0,1,1,1,1,1,0,1,1,
0,1,0,1,1,0,1,1,1,0,1,0,0,1,
1,1,1,1,0,1,1,0,1,1,1,1,1,
1,1,0,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,1,0,0,1,1,1,0,1,1,
0,0,0,1,1,1,1,1,1,1,0,1,1,
1,1,0,1,0,0,1,0,1,1,1,0,0,1,
0,1,0,1,1,0,1,1,1,1,0,0,1,

1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,1,1,1,0,1,
0,1,1,1,1,1,1,1,1,1,0,0,1,
1,1,0,1,1,0,1,1,0,1,1,1,1,
1,1,1,1,1,0,1,1,0,1,1,0,1,1,
0,1,0,0,0,0,1,1,1,1,1,0,0,1,
1,1,1,1,1,0,1,1,1,1,0,0,1,
0,1,1,1,1,0,1,0,1,1,1,0,1,1,
0,1,0,1,0,1,1,1,1,1,0,1,1,1,
0,1,1,1,1,1,1,1,1,1,0,1,0,
1,1,0,1,1,0,1,1,1,1,0,1,1,
1,1,0,1,1,1,1,1,1,1,1,1,1,
1,0,1,1,1,0,1,1,0,0,1,0,1,1,
1,1,1,1,1,1,1,0,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,1,0,1,1,1,1,1,1,1,
1,0,1,1,1,0,1,1,0,1,1,0,0,1,
1,1,0,0,0,0,1,1,1,1,0,0,0,1,
1,1,0,1,1,0,1,1,1,1,0,1,1,
0,1,0,1,1,0,1,1,1,1,0,1,1,
0,1,0,1,1,1,1,1,1,1,0,1,1,
1,1,0,1,0,1,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,0,0,1,1,0,1,1,0,0,1,
1,1,0,1,1,0,1,1,1,0,1,1,1,1,
1,0,1,1,1,0,1,1,1,1,1,0,1,
1,1,1,1,1,0,1,1,0,1,1,0,1,1,
1,1,0,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,0,1,1,1,1,1,1,1,0,1,1,
0,1,1,1,0,0,1,1,1,1,0,0,1,1,
1,1,0,1,1,1,1,1,1,0,1,1,

1,1,1,1,1,0,1,1,1,1,0,0,1,
0,1,1,1,0,0,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,
1,1,1,1,1,1,0,1,1,1,1,0,1,
0,1,0,1,1,0,0,0,0,0,0,1,1,
1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,0,1,1,1,1,
1,1,1,1,1,0,0,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,0,
1,1,0,1,1,0,1,1,0,1,1,0,1,1,
1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,0,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,1,0,1,1,1,
1,1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,1,1,1,1,1,1,0,1,1,0,1,1,
1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,1,1,0,1,1,1,0,1,1,0,1,1,
1,1,0,1,1,0,0,1,1,0,0,1,1,1,
0,1,0,0,1,1,1,0,1,1,1,1,1,1,
1,1,0,1,1,1,1,1,1,1,0,1,1,
0,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,0,0,1,0,1,0,1,0,0,0,0,0,
1,1,0,1,1,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,
0,1,1,1,1,0,1,1,1,0,1,0,1,1,
1,0,0,0,1,1,1,1,0,1,1,0,1,0,
1,1,1,1,1,0,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,1,0,1,1,1,1,0,1,1,
0,0,1,1,1,0,1,1,1,1,1,1,1,

Referências Bibliográficas

- [1] ALBERT, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. **Journal of Educational Statistics** 17, 251-269.
- [2] ALBERT, J. H., CHIB, S. (1993). Bayesian Analysis of binary and polichotomous response data. **JASA** 88, 669-679.
- [3] ALBERT, J. H., GHOSH, M. (2000). Item Response Modeling. **Generalized Linear Models: A bayesian Perspective**. New York.
- [4] ANDRADE, D. F. (2005). **Métodos Estatísticos para Avaliação Educacional**. 39^a Regional da ABE - Maringá.
- [5] ANDRADE, D. F., TAVARES, H. R., VALLE, R. C. (2000). **Introdução à Teoria de resposta ao item: Conceitos e Aplicações**. 14^o SINAPE: Caxambu, MG.
- [6] ARNOLD, B. C., BEAVER, R. J. (2000b). The skew Cauchy distribution. **Statistics & Probabilily Letters** 49, 285-290.
- [7] ARNOLD, B. C., BEAVER, R. J., GROENEVELD, R., MEEKER, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. **Psychometrika** 58, 471-488.
- [8] AZZALINI, A., CAPITANIO, A. (1999). Statistical applications of the multivariate skew-normal distribution. **Journal of the Royal Statistical Society**, series B vol. 61, 579-602.
- [9] AZZALINI, A., CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. **Journal of the Royal Statistical Society**, series B vol. 65, 367-389.
- [10] AZZALINI, A., DALLA VALLE, A. (1996). The multivariate skew-normal distribution. **Biometrika** 83, 715-726.
- [11] AZZALINI, A. (1985). A Class of Distributions which includes the Normal Ones. **Scandinavian Journal of Statistic** 12, 171-178.
- [12] AZZALINI, A. **The Skew-Normal Probability Distribution (and related distributions, such as the skew-t)**. Disponível em: <<http://azzalini.stat.unipd.it/SN>>. Acesso em: 10 de Ago. 2006.

- [13] AZZALINI, A. (2006). **R package sn: The skew-normal and skew-t distributions (version 0.4-1)**. Università di Padova, Italia. Disponível em <<http://azzalini.stat.unipd.it/SN>>.
- [14] BAZÁN, J. (2005). **Uma família de modelos de resposta ao item normal assimétrica**. Tese de Doutorado. IME-SP.
- [15] BERNARDO, J.M., SMITH, A.F.M. (1994). **Bayesian Theory**. Chichester, Wiley.
- [16] BRANCO, D. M., DEY, K. D. (2001). A General Class of Multivariate Skew-Elliptical Distributions. **Journal of Multivariate Analysis** 79, 99-113
- [17] BRANCO, D. M. *et al* (2003). Bayesian and Classical Solutions for Binomial Cytogenetic Dosimetry Problem. **Journal of Data Science** 1, 65-82.
- [18] BRANCO, D. M., VALLE, A. (2004). **Distribuições Elípticas Assimétricas**. 16° SINAPE: Caxambu, MG.
- [19] CHEN, M., DEY, D. K., SHAO, Q. (1999). A new Skewed Link model for Dichotomous Quantal Response Data. **JASA** 94, 1171-1186.
- [20] COPAS, B. J., LI, G. H. (1997). Inference for Non-Random Samples. **Journal of the Royal Statistical Society**, series B, Vol. 59, 55-95.
- [21] GENZ, A., BRETZ, F. (1999). Numerical Computation of Multivariate t-Probabilities with Application to Power Calculation of Multiple Contrasts. **Journal of Statistical Computation and Simulation** 63, 361-378.
- [22] HAMBLETON, R. K., SWAMINATHAN, H. (1985). **Item Response Theory: Principles and Applications**. Boston: Kluwer Academic Publishers.
- [23] HANSON, T. (2004). **Stat 579: Introduction to Bayesian Modeling**. Disponível em: <<http://www.math.unm.edu/hanson/stat579/stat579.htm>>. Acesso em: 05 de Fev. 2007.
- [24] HASTINGS, W. K. (1970). Monte Carlo Sampling methods using Markov Chains and their applications. **Biometrika** 57, 97-109.
- [25] HENZE, N. (1986). A Probabilistic Representation of the Skew-normal Distribution. **Scandinavian Journal of Statistic** 13, 271-275.
- [26] JOHNSON N. L., KOTZ S. (1970). **Distributions in statistics. Continuous univariate distributions - 2**. New York: Wiley.
- [27] NETER, J., WASSERMAN, W., KUTNER, M. H., NACHTSHEIM, C. J. (1983). **Applied Linear Regression Models**. Homewood: IRWIN.
- [28] O'HAGAN, A. (1995). Fractional Bayes Factor for Model Comparison. **Journal of the Royal Statistical Society**, series B vol. 57, 99-138.
- [29] OWEN, D. B. (1956). Tables for computing bivariate normal probabilities. **The Annals of Mathematical Statistics**, Vol. 27, 1075-1090.

- [30] PATZ, R. J., JUNKER, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. **Journal of Educational and Behavioral Statistics** 24, 241-257.
- [31] PAULINO, C. D., TURKMAN, M. A. A.; MURTEIRA, B. (2003). **Estatística Bayesiana**. Ed. Fundação Calouste Gulbenkian.
- [32] PEREIRA, R. V. (2004). **Métodos Alternativos no Critério Brasil para Construção de Indicadores Sócio-Econômico: Teoria da Resposta ao Item**. Dissertação de Mestrado. PUC-Rio.
- [33] RODRÍGUEZ C. (2005). **Inferência Bayesiana no Modelo Normal Assimétrico**. Dissertação de Mestrado. IME-USP.
- [34] SAHU, S. K. (2002). Bayesian Estimation and Model Choice in Item Response Models. **Journal of Statistical Computation and Simulation** 72, 217-232.
- [35] SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G., GILKS, W. R. (1996). **BUGS 0.5 examples (Volume 1 Version i)**. Cambridge, UK: University of Cambridge.
- [36] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society**, series B vol. 64, 583-639.
- [37] SUNG J. H., KANG T. (2006). **Choosing a Polytomous IRT Model using Bayesian Model Selection Methods**. National Council on Measurement in Education annual meeting in San Francisco.
- [38] **The BUGS project**. Disponível em: <<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>>. Acesso em: 10 Out. 2006.
- [39] **The R Project for Statistical Computing**. Disponível em: <<http://www.r-project.org/>> Acesso em: 06 Mai. 2005.
- [40] VAN DER LINDEN, W. J., HAMBLETON, R. K. (1997). **Handbook of modern Item Response Theory**. New York: Springer-Verlag.
- [41] VALLE, A. R. (1994). **Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão**. Tese de Doutorado. IME-USP.
- [42] VALLE, R. C. (1999). **Teoria de Resposta ao Item**. Dissertação de Mestrado. IME-USP.
- [43] ZELLNER, A. (1971). **An introduction to Bayesian inference in econometrics**. John Wiley & Sons, INC. New York.